

JULIANA DA LUZ

**IMAGENS ALOS PARA O MAPEAMENTO DA VEGETAÇÃO ARBÓREA E OUTROS
USOS DO SOLO EM ÁREA DE FLORESTA OMBRÓFILA MISTA**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Florestal do Setor de Ciências Agrárias da Universidade Federal do Paraná, como requisito parcial à obtenção do grau e título de mestre em Engenharia Florestal, área de Concentração Manejo Florestal.

Orientador: **Prof. Dr. Nelson Carlos Rosot**

Co-Orientador: **Dra. Maria Augusta Doetzer Rosot**

CURITIBA

2009




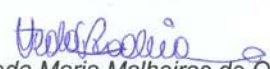
Universidade Federal do Paraná
Setor de Ciências Agrárias - Centro de Ciências Florestais e da Madeira
Programa de Pós-Graduação em Engenharia Florestal

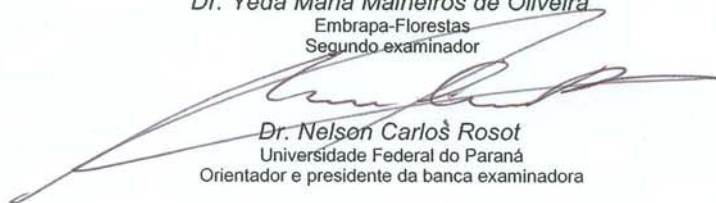
PARECER

Defesa nº. 793


A banca examinadora, instituída pelo colegiado do Programa de Pós-Graduação em Engenharia Florestal, do Setor de Ciências Agrárias, da Universidade Federal do Paraná, após arguir o(a) mestrando(a) *Juliana da Luz* em relação ao seu trabalho de dissertação intitulado "**IMAGENS ALOS PARA O MAPEAMENTO DA VEGETAÇÃO ARBÓREA E OUTROS USOS DO SOLO EM ÁREA DE FLORESTA OMBRÓFILA MISTA**", é de parecer favorável à **APROVAÇÃO** do(a) acadêmico(a), habilitando-o(a) ao título de *Mestre* em Engenharia Florestal, área de concentração em MANEJO FLORESTAL.


Dr. Paulo Costa de Oliveira Filho
Universidade Estadual do Centro Oeste - UNICENTRO
Primeiro examinador


Dr. Yeda Maria Malheiros de Oliveira
Embrapa-Florestas
Segundo examinador


Dr. Nelson Carlos Rosot
Universidade Federal do Paraná
Orientador e presidente da banca examinadora

Curitiba, 29 de maio de 2009.


Graciela Ines Boizon de Muniz
Coordenador do Curso de Pós-Graduação em Engenharia Florestal
Antonio Carlos Batista
Vice-coordenador do curso

*Com humildade, aos brasileiros,
que há anos sustentam minha formação*

DEDICO

AGRADECIMENTOS

Agradeço à Universidade Federal do Paraná pela oportunidade da realização do mestrado. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela concessão da bolsa de estudo.

À Embrapa Florestas pelo apoio financeiro e logístico na coleta e processamento dos dados e à EPAGRI por facilitar o uso de suas instalações na Reserva de Caçador.

Aos meus orientadores, professor Dr. Nelson Carlos Rosot pelas oportunidades oferecidas desde a graduação e pelo aceite e interesse em me orientar nesta empreitada, à Dra. Maria Augusta Doetzer Rosot pela co-orientação e amizade do começo ao fim do mestrado, sou muito grata a vocês dois.

À Dra. Yeda Maria Malheiros de Oliveira pela confiança, apoio e ensinamentos, todos fundamentais para o sucesso desta empreitada.

Ao professor Dr. Paulo Costa de Oliveira Filho pela participação na banca examinadora, o que muito contribuiu na melhoria do trabalho.

Aos pesquisadores Marilice Cordeiro Garrastazú e André Biscaia, da Embrapa Florestas, que não mediram esforços juntos à Augusta e Yeda em me ajudar sempre com dedicação acompanhada de alegria. Somente com a ajuda de vocês esse trabalho foi possível.

À Betina Kellermann, Arnaldo Soares e Carlos Roberto Úrio por serem pessoas especiais e divertidas, me proporcionando dias excelentes em sua companhia.

À Marlise Zonta, Patrícia Povoá e Vera Eifler pela amizade e apoio, à Simone A. Sopchaki por sempre me ajudar com as pesquisas na biblioteca. A vocês, muito obrigada.

Agradeço os meus amigos Hugo Rivera, Fernando Dlugosz, Lisâneas A. do Nascimento, Vilmar Sobrinho, Andreia Pimentel, Jorgeane Schaefer, Gustavo Thome, Lucas Gomide, Gilberto Terra, entre tantos outros do mestrado e da graduação pelos bons momentos compartilhados.

À Dayane, Ana Paula, Rafael, Márcio Cruz, Carol Jazar, Leonardo Ribeiro e meus amigos de Brusque por estarem presentes em minha vida nos momentos felizes e nos momentos não tão felizes, porque tristes jamais. Parabéns aos mestres e sorte aos quase mestres.

À minha tia Rosemeri e meu tio Luis por me acolherem em sua casa como sua própria filha e também ao meu primo Maikon por todo o carinho.

Aos meus avós maternos Saulo e Elly, meus avós paternos Augusto (*in memoriam*) e Jandira, meus tios Roseli, Geraldo, Aureo, Margit, Silvio, Noeli, Roberto e Solange e meus primos Marco Aurélio, Giorgia, Lara, Daiana, Daniel e Bruna. Obrigada por tudo.

Ao Sérgio pelo seu amor, companheirismo e por ser uma pessoa especial em minha vida. Obrigado pela sua paciência. Estendo o agradecimento a toda a sua família pelo carinho dedicado a mim.

Por fim, agradecimento especial à minha mãe Tânia Maria da Luz, ao meu pai José Ayone da Luz e a minha irmã Fernanda da Luz por todo amor e compreensão e por possibilitarem minha vinda à Curitiba.

RESUMO

O presente estudo teve como objetivo principal avaliar o uso de diferentes algoritmos classificadores em imagens do satélite ALOS para mapeamento do uso do solo e classes de vegetação em um fragmento de Floresta Ombrófila Mista. Como objetivo secundário foram aplicadas e analisadas quatro diferentes técnicas de fusão para as imagens PRISM (2,5 metros de resolução espacial) e AVNIR-2 (10 metros), utilizando três e quatro bandas. As técnicas utilizadas foram as seguintes: HSV, *Color normalized* (CN), *Gram-Schmidt Spectral Sharpening* e *Principal Components Spectral Sharpening*. Para as classificações foram utilizados os algoritmos supervisionados *Bhattacharya* e Árvore de Decisão (C4.5). A área de estudo foi a Reserva Florestal Embrapa/Epagri localizada no Município de Caçador, SC. Na classificação por árvore de decisão foi utilizado um aplicativo para a geração do conjunto de regras da árvore, o *software* WEKA e um aplicativo de processamento de imagens, o *software* ENVI, para a classificação digital (RSI, 2005). Foram utilizadas 61 variáveis (13 espectrais e 48 de textura) representadas em 317 amostras de treinamento de 3 x 3 pixels cada uma. Quando da classificação por *Bhattacharya* (algoritmo de classificação por crescimento de regiões implementado no *software* SPRING) foi necessário primeiramente realizar a segmentação da imagem para posterior uso das regiões como amostras de treinamento para a classificação. Devido à limitação do *software* em processar 61 bandas conjuntamente no processo de segmentação, a técnica Análise de *Cluster* foi empregada para a seleção de seis bandas representativas do conjunto de dados. Na análise visual e no teste de fidelidade espectral, as técnicas de fusão que apresentaram melhores resultados foram as Componentes Principais e *Gram-Schmidt*. No teste de transferência de detalhes as quatro técnicas de fusão se mostraram apropriadas. No resultado das classificações o algoritmo que apresentou melhor acurácia foi a Árvore de Decisão, apresentando valor de coeficiente Kappa de 0,966 e acurácia geral de 97% em comparação com o *Bhattacharya*, que respectivamente apresentou os seguintes valores: 0,755 e 79%. Apesar da técnica Árvore de Decisão apresentar um coeficiente Kappa superior, sua classificação se mostrou com aparência de “*salt and pepper*”, com pixels isolados, assemelhando-se a uma imagem com ruído, ao contrário da classificação por regiões.

Palavras-chave: Imagens ALOS. Fusão de imagens. Árvore de Decisão. Classificação por regiões.

ABSTRACT

This study aimed at the assessment of different classifiers for mapping land use and vegetation in Mixed Ombrophylous Forest based on ALOS satellite imagery. As a secondary objective, four different fusion techniques were applied and analyzed in PRISM (2.5 m resolution) and AVNIR (10 m resolution) images using three and four bands. The following techniques were tested: HSV, Color normalized (CN), *Gram-Schmidt* Spectral Sharpening and Principal Components Spectral Sharpening. The classification schemes were based on 2 supervised algorithms, namely *Bhattacharya* and Decision Tree (C4.5). The study area was the Embrapa/Epagri Forest Reserve located at the municipality of Caçador, in Santa Catarina State. The software WEKA and an image processing software ENVI (RSI, 2005) were used to build and apply the Decision Tree classification, which included 61 variables (13 spectral and 48 textural) represented in 317 3-by-3-meter training samples. The region growing algorithm *Bhattacharya*, implemented in the SPRING software, demanded previous image segmentation in order to use the regions as training samples for the classification. Due to the software limitations regarding the processing of 61 bands in the segmentation, a cluster analysis was performed to select the six most representative bands within the dataset. According to the visual analysis and the spectral fidelity test, the best results for image fusion were obtained by the Principal Components and *Gram-Schmidt* techniques. The detail-transference test revealed that all four techniques were appropriate. In relation to the classifiers, the most accurate algorithm was the Decision Tree, presenting a Kappa coefficient of 0.966 and an overall accuracy of 97% against the *Bhattacharya*, which presented, respectively, 0.755 and 79%. Nevertheless, despite the higher Kappa coefficient, the resulting classified image from the Decision Tree, showed a “salt-and-pepper” appearance, typical from pixel classifiers, opposite to the other method which is based on region growing algorithms.

Key-words: ALOS images. Fusion images. Decision Tree. Classification by region growing.

LISTA DE FIGURAS

FIGURA 1: Espectro eletromagnético.....	23
FIGURA 2: Propriedades radiométricas dos objetos: absorvância (A), reflectância (R) e transmitância (T)	24
FIGURA 3: Curva de reflectância de uma folha verde.....	25
FIGURA 4: Curva de reflectância da água nos três estados físicos: sólido, líquido e gasoso	27
FIGURA 5: Satélite ALOS com seus três sensores	30
FIGURA 6: (a) Sensor PRISM; (b) Visadas <i>Forward</i> , <i>Nadir</i> e <i>Backward</i> e faixa de imageamento.	31
FIGURA 7: (a) Sensor AVNIR-2; (b) Modo de observação do AVNIR-2	31
FIGURA 8: Sensor PALSAR.....	32
FIGURA 9: Roteiro para busca de imagens ALOS na internet	35
FIGURA 10: Cálculo da matriz de Co-ocorrência.	42
FIGURA 11: Padrões de texturas e suas GLCM.....	44
FIGURA 12: Localização da área de estudo.....	69
FIGURA 13: Tipologias definidas para a RFEE.....	70
FIGURA 14: (a) Imagem ALOS multiespectral com resolução de 10 metros; (b) Imagem ALOS pancromática com resolução de 2,5 metros.	73
FIGURA 15: Espacialização das amostras de treinamento e validação	77
FIGURA 16: Imagens fusionadas pelos métodos: (a) HSV; (b) <i>Color normalized</i> ; (c) <i>Gram-Schmidt</i> ; (d) Componentes Principais; (e) <i>Gram-Schmidt</i> com 3 bandas e (f) Componentes Principais com 3 bandas;	82
FIGURA 17: Histograma de comparação entre a imagem original e as fusões. As curvas em preto são da imagem original e as coloridas são das fusões.	85
FIGURA 18: Resultado da classificação por árvore de decisão	87
FIGURA 19: Instâncias e regras para a classificação dos dados na árvore de decisão	88
FIGURA 20: Imagens das classes de uso definidas para a Reserva	92
FIGURA 21: Dendrograma	94
FIGURA 22: Imagem segmentada com limiares de similaridade e área 5 e 180, respectivamente	95
FIGURA 23: Resultado da classificação por regiões pelo classificador <i>Bhattacharya</i>	97
FIGURA 24: (a) Classificação por árvore de Decisão; (b) Classificação por regiões pelo classificador <i>Bhattacharya</i>	101
FIGURA 25: (a) Detalhes da classificação por Árvore de Decisão; (b) Detalhes da classificação por regiões pelo classificador <i>Bhattacharya</i>	102

LISTA DE TABELAS

TABELA 1: Bandas espectrais do satélite ALOS e suas respectivas resoluções espaciais.....	32
TABELA 2: Valores do coeficiente Kappa	66
TABELA 3: Estrutura de arquivos no formato ARFF.....	67
TABELA 4: Bandas espectrais do satélite ALOS e suas respectivas resoluções espaciais.....	72
TABELA 5: Variáveis utilizadas na técnica de classificação por árvore de decisão.....	75
TABELA 6: Número de amostras de treinamento e validação por classe	78
TABELA 7: Coeficientes de correlação entre as imagens fusionadas filtradas e a pancromática filtrada, para o teste de transferência de detalhes	83
TABELA 8: Comparação entre as estatísticas da imagem multiespectral e das fusões	83
TABELA 9: Matriz de confusão da classificação por árvore de decisão.....	90
TABELA 10: Acurácia do produtor e do usuário para a classificação por árvore de decisão	91
TABELA 11: Limiares de similaridade e área testados na segmentação.....	95
TABELA 12: Matriz de confusão da classificação por <i>Bhattacharya</i>	96
TABELA 13: Acurácia do produtor e do usuário para a classificação por <i>Bhattacharya</i>	98

LISTA DE SIGLAS

- AD – Árvore de Decisão
- ADEOS – *Advanced Earth Observing Satellite*
- ADM – Árvore de Decisão Multivariada
- ADU – Árvore de Decisão Univariada
- ALOS – *Advanced Land Observing Satellite*
- ARFF – *Attribute-Relation File Format*
- ASF – *Alaska Satellite Facility*
- ASTER - *Advanced Spaceborne Thermal Emission and Reflection Radiometer*
- AVHRR – *Advanced Very High Resolution Radiometer*
- AVNIR-2 – *Advanced Visible and Near Infrared Radiometer Type 2*
- BSQ – *Band Sequential Format*
- CAP – Circunferência à Altura do Peito
- CEOS – *Committee on Earth Observation Satellites*
- CN – *Color Normalized*
- CROSS – *Online Satellite Images Search & Order System*
- DAP – Diâmetro à Altura do Peito
- DEM – *Digital Elevation Model*
- DN – *Digital Number*
- DSG – Diretoria de Serviços Geográficos
- EMBRAPA – *Empresa Brasileira de Pesquisa Agropecuária*
- ENVI – *Environment for Visualizing Images*
- EPAGRI – *Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina*
- ERTS – *Earth Resources Technology Satellite*
- ESA – *Europe Space Agency*
- ETM – *Enhanced Thematic Mapper*
- GLCM – *Grey-level Co-occurrence Matrix*
- HSV – *Hue, Saturation, Value*
- IBGE – Instituto Brasileiro de Geografia e Estatística
- ID – Identificador de cena

IHS – *Intensity Hue Saturation*
INPE – Instituto Nacional de Pesquisas Espaciais
IV – Infravermelho
JAXA – *Japan Aerospace Exploration Agency*
JERS – *Japanese Earth Resources Satellite*
LANDSAT – *Land Observation Satellite*
MDT – Modelo Digital do Terreno
MODIS - *Moderate Resolution Imaging Spectroradiometer*
MSS – *Multispectral Scanner*
NDVI – *Normalized Difference Vegetation Index*
PALSAR – *Phased Array Type L Band Synthetic Aperture Radar*
PCA – *Principal Component Analysis*
PRISM – *Panchromatic Remote-Sensing Instrument for Stereo Mapping*
PSR – Posição Sociológica Relativa
REM – Radiação Eletro-magnética
RFEE – Reserva Florestal Embrapa Epagri
RGB – *Red Green Blue*
RMS – *Root Mean Square* (Erro Médio Quadrático)
ROI – *Region of Interest*
RVI – *Simple Ratio*
SAD 69 – *South American Datum of 1969*
SAVI – *Soil Adjusted Vegetation Index*
SPOT – *Système Pour l'Observation de la Terre*
SPRING – Sistema de Processamento de Informações Geográficas
SRTM – *Shuttle Radar Topography Mission*
TIFF – *Tagged Image File Format*
TM - *Thematic Mapper*
URSA – *User Remote Sensing Access*
UTM – *Universal Transverse Mercator*
UV – Ultravioleta
WEKA – *Waikato Environment for Knowledge Analysis*

SUMÁRIO

1. INTRODUÇÃO	13
2. OBJETIVOS	17
2.1 OBJETIVO GERAL.....	17
2.2 OBJETIVOS ESPECÍFICOS	17
3. REVISÃO DE LITERATURA	18
3.1 FLORESTA OMBRÓFILA MISTA.....	18
3.2 O SENSORIAMENTO REMOTO E A ENERGIA ELETROMAGNÉTICA	20
3.2.1 O Sensoriamento Remoto.....	20
3.2.2 A Energia Solar e o Espectro Eletromagnético	21
3.3 COMPORTAMENTO ESPECTRAL DOS ALVOS.....	23
3.3.1 Vegetação	24
3.3.2 Solo.....	26
3.3.3 Água.....	26
3.4 SENSORIAMENTO REMOTO NA FLORESTA OMBRÓFILA MISTA	27
3.5 O SATÉLITE ALOS	29
3.5.1 Sistemas Sensores do Satélite ALOS	30
3.5.1.1 <i>Panchromatic Remote-Sensing Instrument for Stereo Mapping (PRISM)</i>	30
3.5.1.2 <i>Advanced Visible and Near Infrared Radiometer Type 2 (AVNIR-2)</i>	31
3.5.1.3 <i>Phased Array Type L-band Synthetic Aperture Radar (PALSAR)</i>	32
3.5.2 Formato e Níveis de Processamento dos Dados	33
3.5.3 Aquisição de Imagens ALOS	34
3.5.4 Conceito da Missão ALOS	35
3.6 PROCESSAMENTO DE IMAGENS DIGITAIS.....	36
3.6.1 Retificação Geométrica	36
3.6.1.1 <i>Correção Geométrica</i>	36
3.6.1.2 <i>Reamostragem de Pixels</i>	36
3.6.2 Transformação de Dados	37
3.6.2.1 <i>Componentes Principais (PC's)</i>	37
3.6.2.2 <i>Índices de Vegetação</i>	38
3.6.3 Filtragens Espaciais: Texturas	39
3.6.4 Técnicas de Fusão de Imagens	46

3.6.4.1	<i>Modelos de Domínio Espacial</i>	47
3.6.4.2	Modelos de Domínio Espectral	47
3.6.4.2.1	<i>HSV</i>	47
3.6.4.2.2	<i>Gram-Schmidt spectral sharpening</i>	48
3.6.4.2.3	<i>Principal components spectral sharpening</i>	48
3.6.4.3	<i>Modelos Algébricos</i>	49
3.6.4.3.1	<i>Color normalized (Brovey)</i>	49
3.6.5	Segmentação	50
3.6.5.1	<i>Seleção de Bandas para a Segmentação</i>	51
3.6.6	Classificação de imagens digitais	53
3.6.6.1	<i>Classificação Não-supervisionada</i>	54
3.6.6.2	<i>Classificação Supervisionada</i>	55
3.7	ÁRVORE DE DECISÃO	57
3.7.1	Árvores de Decisão Univariadas (ADU) e Multivariadas (ADM)	59
3.7.2	Algoritmo de Poda	60
3.7.3	Uso de Dados de Sensores Remotos em Árvore de Decisão	61
3.8	ACURACIDADE DOS ALGORITMOS CLASSIFICADORES.....	64
3.9	O SOFTWARE WEKA.....	66
4.	MATERIAIS E MÉTODOS	68
4.1	ÁREA DE ESTUDO	68
4.2	BASE CARTOGRÁFICA E MAPAS DA RFEE	69
4.3	IMAGENS E <i>SOFTWARE</i> UTILIZADOS	70
4.4	GEORREFERENCIAMENTO	71
4.5	FUSÃO DE IMAGENS	72
4.6	CLASSES DE USO DO SOLO NA RFEE.....	74
4.7	ÁRVORE DE DECISÃO	74
4.8	SEGMENTAÇÃO DE IMAGENS	78
4.8.1	Seleção de Bandas para a Segmentação	78
4.8.2	Algoritmo de Segmentação.....	79
4.9	CLASSIFICAÇÃO SUPERVISIONADA DA IMAGEM SEGMENTADA	79
4.10	ACURACIDADE DA CLASSIFICAÇÃO DE IMAGENS DIGITAIS	80
5.	RESULTADOS E DISCUSSÕES	81
5.1	GEORREFERENCIAMENTO	81
5.2	FUSÃO DE IMAGENS	81

5.3	ÁRVORE DE DECISÃO	85
5.4	SELEÇÃO DE BANDAS PARA A SEGMENTAÇÃO	93
5.5	SEGMENTAÇÃO DE IMAGENS	94
5.6	CLASSIFICAÇÃO DA IMAGEM SEGMENTADA.....	96
5.7	COMPARAÇÃO ENTRE OS MÉTODOS DE CLASSIFICAÇÃO.....	98
6.	CONCLUSÕES E RECOMENDAÇÕES	103
	REFERÊNCIAS BIBLIOGRÁFICAS	105
	APÊNDICES	116
	APÊNDICE A - Cabeçalho do arquivo ARFF	117
	APÊNDICE B - Árvore de decisão originada no WEKA	119
	APÊNDICE C - Imagens segmentadas e seus respectivos limiares de similaridade e área.....	121

1. INTRODUÇÃO

No Brasil o conceito de manejo florestal sustentável tem sido aplicado à Floresta Amazônica como forma de, racionalmente, explorar seus recursos madeireiros garantindo, ao mesmo tempo, a perenidade dos demais recursos naturais renováveis, a manutenção da biodiversidade e a manutenção dos processos ecológicos.

No entanto, no bioma denominado “Mata Atlântica”, que cobre 17 estados, do nordeste ao sul do Brasil, a utilização da floresta através de planos de manejo está proibida desde a década de 90 (Decreto 750/1993 e Resolução 278/2001-CONAMA) em função do alto grau de fragmentação e devastação encontrados nessa região fitoecológica. Embora bastante restritiva a legislação não foi capaz de conter o avanço da pecuária, agricultura, urbanização e outras formas de conversão de uso da terra (ROSOT, 2007).

A Floresta Ombrófila Mista, também conhecida como Floresta de Araucária por possuir a espécie *Araucaria angustifolia* como característica – presente nos três estados meridionais do Brasil – se insere como uma das regiões fitoecológicas no domínio da Floresta Atlântica ocupando, originalmente, 20 milhões de hectares, dos quais se estima que existam, hoje, apenas 400.000 ha (GUERRA et al. 2000). Na sua maioria esses remanescentes se encontram em más condições, empobrecidos devido a explorações predatórias realizadas nos últimos 60 anos (BITTENCOURT et al. 2004).

A Lei 11.428, aprovada em 22 de dezembro de 2006, também conhecida como Lei da Mata Atlântica (BRASIL, 2006) dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica, ressaltando a necessidade de se garantir a manutenção das condições necessárias para a reprodução e sobrevivência das espécies nativas, inclusive a explorada.

O Decreto 6.660, de 21 de novembro de 2008, regulamenta os dispositivos da Lei da Mata Atlântica permitindo a exploração eventual, sem propósito comercial direto ou indireto, de espécies da flora nativa provenientes de formações naturais, para consumo nas propriedades rurais, posses das populações tradicionais ou de pequenos produtores rurais. Com esta retomada do uso sustentável das florestas no sul do país, surge a necessidade de se estabelecer esquemas silviculturais condizentes. Para tanto, entende-se que a divisão da floresta a ser manejada em subtipologias homogêneas quanto à composição de espécies, estágio de desenvolvimento e

número de estratos pode ser uma estratégia efetiva e facilitadora dos tratamentos silviculturais que se pretende aplicar.

Os estudos desenvolvidos na Floresta de Araucária (LONGHI (1980); OLIVEIRA e ROTTA (1982); SILVA e MARCONI (1990); NEGRELLE e SILVA (1992), RONDON NETO et al. (2002 a) e KOZERA et al. (2006)), muitas vezes têm se restringido a uma abordagem fitossociológica dos povoamentos, analisando-se sua estrutura horizontal e vertical. Outros, como REITZ et al. (1978), procuraram caracterizar os estágios sucessionais, sem, no entanto, definir subtipologias propriamente ditas. Algumas tentativas nesse sentido foram efetuadas por KLEIN (1960) e LONGHI (1997), por meio de análises fitofisionômicas e técnicas de análise multivariada, respectivamente.

LAMPRECHT (1964) menciona que mesmo não havendo uma metodologia padrão para analisar as comunidades florestais, que são basicamente produtos da diferença de tolerância das espécies em relação ao meio (amplitude ecológica) e da heterogeneidade do meio, qualquer procedimento que se adote deve: proporcionar uma visão representativa da estrutura da floresta estudada; ser aplicável a qualquer tipo de comunidade florestal; estar livre de influências subjetivas; ser passível de comparação com outros estudos e; permitir a aplicação de métodos estatísticos na compilação e comparação dos resultados.

Embora o uso de sensores remotos seja prática comum na avaliação da cobertura florestal e discriminação de fases sucessionais da floresta no Brasil (CARREIRAS et al. 2006; VIEIRA et al. 2003), poucos trabalhos têm sido aplicados à Floresta de Araucária com o intuito de classificá-la em subtipologias. Imagens de satélite são, na maior parte das vezes, empregadas para a interpretação visual das feições de vegetação e outros usos, objetivando estudos de paisagem (LIMA et al. 2004). Mais recentemente outros pesquisadores têm se dedicado a investigar a aplicabilidade de imagens de alta resolução como as IKONOS II, por exemplo, para a definição de subtipologias na Floresta de Araucária (DLUGOSZ et al. 2004a, 2004b; DLUGOSZ et al. 2005). Nesses trabalhos têm-se abordado tanto técnicas de segmentação de imagens como de interpretação visual em tela.

A segmentação de imagens é uma técnica de processamento digital de imagens que divide a imagem em regiões homogêneas quanto às suas características espectrais e espaciais. Essas regiões podem, então, ser utilizadas como amostras de treinamento para a classificação digital, onde são geradas áreas com aspecto contínuo e de características espectrais bem distintas

em relação umas às outras. Nesse caso, o algoritmo classificador empregado é o de regiões, que possui como vantagem o fato de simular o comportamento do fotointérprete ao analisar uma cena, considerando, além da informação espectral (cores), outras variáveis de contexto como a distribuição espacial das feições, a textura ou as bordas formadas naturalmente entre regiões distintas.

Por outro lado, os classificadores “pixel a pixel”, bastante utilizados para a classificação de dados de sensoriamento remoto (XU et al. 2005), possuem a limitação referente à análise da informação espectral, somente. Entretanto, resultados promissores têm sido obtidos com este tipo de classificador ao se empregar abordagens não-tradicionais, tais como a Árvore de Decisão (AD) – uma técnica exploratória de dados. Um dos algoritmos mais utilizados em ADs é o C4.5, desenvolvido por J. Ross Quinlan, na década de 70, podendo ser descrito como um procedimento de classificação que particiona um conjunto de dados em subdivisões menores com base em um conjunto de testes definidos em cada ramo (ou nó) da árvore. O conjunto de dados é representado pelas imagens satelitárias originais ou suas derivadas, incluindo-se índices de vegetação, componentes principais e imagens-textura, dentre outras.

No presente trabalho objetivou-se avaliar a adequação de dois algoritmos classificadores em imagens do satélite ALOS – um por regiões e o outro por pixel – na discriminação e mapeamento do uso do solo e classes de vegetação (com ênfase para subtipologias florestais), usando como área-teste a Reserva Florestal Embrapa-Epagri (RFEE), em Caçador-SC, que constitui um dos últimos remanescentes de grandes dimensões de Floresta Ombrófila Mista no Brasil. A área possui 1.194 ha onde se destacam algumas das espécies que constam na lista oficial de espécies ameaçadas de extinção, como, por exemplo, *Araucaria angustifolia* (araucária), *Ocotea porosa* (imbuia) e *Ocotea catharinensis* (canela-preta) (IBAMA, 1992). Recentemente dois estudos relacionados à ordenamento territorial foram desenvolvidos na Reserva (KURASZ, 2005; RIVERA, 2007), além de uma primeira abordagem sobre a divisão da floresta em subtipologias, conforme o trabalho de Dlugosz (2005) que efetuou a classificação do uso do solo com apoio de uma imagem de 2004, proveniente do satélite Ikonos. Nessa mesma área de estudo, Rosot et al. (2007) compararam a divisão em subtipologias efetuada através de uma abordagem fitossociológica apoiada por técnicas de interpolação espacial com aquela obtida por meio de segmentação e classificação digital de imagens (Dlugosz, 2005).

Devido a alterações substanciais na fitofisionomia da Reserva ocorridas em função do fenômeno da floração e seca da taquara (*Merostachys sp.*, *Chusquea sp.*) nos anos de 2005 e 2006, novas imagens satelitárias são necessárias para um estudo atualizado da vegetação e o satélite ALOS (*Advanced Land Observing Satellite*) parece ser muito indicado para tal fim. Este satélite foi lançado no Japão em 2006 pela Agência Espacial Japonesa e carrega a bordo de sua plataforma três sensores imageadores, o PRISM (sensor com uma banda pancromática, resolução espacial de 2,5m), AVNIR-2 (sensor multispectral, possuindo 4 bandas e resolução espacial de 10m) e PALSAR (sensor radar). Entre as vantagens dos produtos ALOS encontra-se o subsídio governamental com relação ao preço de aquisição no mercado brasileiro, no caso de aplicações em pesquisas científicas.

Como objetivos secundários do estudo também foram avaliadas técnicas de fusão das imagens pancromática e multiespectrais do satélite ALOS visando a geração de produtos com informação espectral relevante associada a uma alta resolução espacial.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Avaliar a adequação de diferentes algoritmos classificadores na discriminação e mapeamento do uso do solo e classes de vegetação em um fragmento de Floresta Ombrófila Mista por meio de imagens do satélite ALOS.

2.2 OBJETIVOS ESPECÍFICOS

- Aplicar e analisar diferentes técnicas de fusão em imagens dos sensores PRISM e AVNIR-2 do satélite ALOS;
- Aplicar e analisar a utilização da técnica árvore de decisão e segmentação de imagens para a classificação do uso do solo, com ênfase em formações florestais;
- Comparar a acuracidade dos algoritmos classificadores utilizados para o mapeamento de uso do solo;
- Gerar mapas de vegetação contendo as diferentes subtipologias discrimináveis em uma área de Floresta Ombrófila Mista;

3. REVISÃO DE LITERATURA

3.1 FLORESTA OMBRÓFILA MISTA

Também conhecida como Floresta de Araucária (IBGE, 2006), a Floresta Ombrófila Mista originalmente ocorria em 40% da superfície do Paraná, 31% de Santa Catarina, 25% do Rio Grande do Sul, em manchas esparsas no sul de São Paulo (3%) e sul de Minas Gerais e Rio de Janeiro (1%) (CARVALHO, 1994).

O termo “ombrófila” indica que a Floresta de Araucária ocorre em regiões onde as chuvas são bem distribuídas ao longo do ano, e o termo “mista” refere-se ao encontro de duas floras distintas, mais antigas, a Temperada Austro Brasileira e a Tropical Afro-Brasileira (KOCH e CORRÊA, 2002).

No estado de Santa Catarina, quase todo o planalto é constituído por esse tipo de floresta, interrompidas pelos campos naturais que formam como que “ilhas cercadas por pinheiros” (KLEIN, 1960). A vegetação da região de Araucária não é uma formação homogênea e contínua. É formada por múltiplas associações e agrupamentos que se encontram nos mais variados estágios de sucessão, sendo que cada estágio possui suas espécies características (KLEIN, 1960).

Maack (1981) cita as espécies imbuia (*Ocotea porosa*) e erva-mate (*Ilex paraguariensis*) como sendo importantes em associação com a araucária. Ao lado desses agrupamentos significantes ocorre também grande variedade de pteridófitos, entre os quais existe a predominância do xaxim (*Dicksonia sellowiana*), a samambaia-açu (*Hemitelia setosa*) e mais raramente o xaxim liso (*Cyathea schanschin*). Ainda segundo este autor, são citadas outras espécies que se associam regularmente à araucária, obrigando-a a elevar-se rapidamente para em seguida perder seus galhos laterais. São elas: em primeiro lugar as espécies de canela, da família das lauráceas, *Nectandra* sp., *Persea Venosa*, *Ocotea pretiosa*. Seguem-se diversas leguminosas conhecidas por jacarandá e caviúna, *Dalbergia brasiliensis* e *Machaerium* sp., e o monjoleiro, *Acacia polyphylla*. Encontram-se também representantes da família das meliáceas, *Cedrela fissilis* e *Cedrela* sp. As mirtáceas são representadas pela gabioba legítima e gabioba da serra (*Britoa sellowiana* e *Campomanesia xanthocarpa*). Entre as coníferas encontra-se o pinheiro-bravo (*Podocarpus lambertii* e *Podocarpus sellowii*). A rutácea *Balfourodendron riedelianum*, o

pau-marfim, ocorre amplamente distribuída na floresta de araucária. Como as matas pluviais tropicais e subtropicais de folhas caducas, a mata de araucária também exibe uma série de exemplares de lianas, bignoniáceas, compostas, rosáceas, leguminosas e passifloráceas, porém em menor escala.

A Floresta de Araucária é considerada uma das regiões mais degradadas da América Latina, apresentando uma paisagem bastante alterada, ocasionando a perda de habitats, isolamento de populações, causando empobrecimento genético, extinção de espécies, perda da biodiversidade, distúrbios nos regimes dos rios e erosão de solos, entre outros (BRITZ, 2007).

Existem vários fatores de degradação da Floresta de Araucária e a grande densidade de taquaras no sub-dossel destes remanescentes é um deles. Isto dificulta os processos regenerativos, reduz o crescimento e acelera o processo de mortalidade das populações arbóreas (SANQUETTA, 2007). Os bambus e taquaras são plantas essencialmente florestais, embora algumas espécies ocorram em áreas abertas. Distribuem-se entre as latitudes 51°N e 47°S, e desde o nível do mar até 4.300m de altitude. Podem ocupar diversos estratos nas florestas temperadas e tropicais, inclusive chegando a dominá-las. Os gêneros mais populares são *Merostachys* e *Chusquea* (TERRA et al. 2006).

Rivera (2007) realizou um levantamento florístico e fitossociológico para a Reserva Florestal Embrapa Epagri (RFEE), localizada no município de Caçador-SC e que está dentro do domínio da Floresta Ombrófila Mista. O autor identificou 71 espécies, distribuídas em 54 gêneros e 33 famílias. As famílias que apresentaram maior riqueza de espécies foram Myrtaceae (14), Lauraceae e Fabaceae (5), Flacourtiaceae e Asteraceae (4), Aquifoliaceae e Sapindaceae (3), que representaram 52,8% do número total de espécies encontradas. Sete famílias tiveram duas espécies e 20 apenas uma espécie. Oliveira e Rotta (1982), Silva e Marconi (1990), Negrelle e Silva (1992) também obtiveram estas como as principais famílias encontradas nesta formação florestal. Em relação à densidade, existem 483,6 indivíduos por hectare. A espécie dominante com 75 indivíduos/ha (14,5%) foi a *Cupania vernalis*, seguida por *Capsicodendron dinisii*, *Prunus brasiliensis*, *Ocotea pulchella*, *Araucaria angustifolia* e *Matayba elaeagnoides*, que apresentaram entre 20 e 30 indivíduos/ha. As outras 51 espécies acumularam apenas 25% do total de indivíduos/ha. O valor médio de DAP (diâmetro à altura do peito) encontrado foi de 23,9 cm. As espécies *Araucaria angustifolia*, *Ocotea porosa* e *Vitex megapotamica* superaram 40 cm de DAP médio. A grande maioria das espécies apresentou a média de DAP abaixo de 30 cm e

principalmente nas faixas de 10 e 20 cm. Na distribuição dos indivíduos em classes diamétricas, a *Araucaria angustifolia* apresentou uma distribuição multimodal (mais de um ponto de maior frequência), típica de florestas que sofreram exploração. A altura média dos indivíduos amostrados foi de 11,7 m, sendo que o maior indivíduo encontrado foi de *Araucaria angustifolia*, com 33 m. Esta espécie apresentou a pior distribuição vertical, e portanto, a mais baixa posição sociológica relativa (PSR), uma vez que o número de árvores no estrato médio e inferior foi muito menor que o número de árvores em seu estrato superior. A espécie que apresentou melhor PSR foi *Cupania vernalis*, apresentando distribuição vertical regular.

Kellermann et al. (2007) também realizaram estudos de avaliação da regeneração natural na RFEE. Em duas amostragens realizadas encontraram 100 espécies, pertencentes a 35 famílias. As espécies mais importantes foram *Piptocarpha angustifolia*, *Ocotea puberula*, *Nectandra megapotamica*, *Mimosa scabrella*, *Aegiphila riedeliana*, *Prunus brasiliensis*, *Matayba elaeagnoides*, *Solanaceae* e *Cupania vernalis* com os maiores números de regeneração natural relativa. A família com maior número de espécies foi Myrtaceae (12) seguida de Lauraceae (8), Asteraceae, Solanaceae, Flacourtiaceae (6) e Sapindaceae (4), que constituem aproximadamente 55% do total de espécies. Na distribuição do número de indivíduos por família, Lauraceae é que se apresenta em primeiro lugar, com 236 indivíduos, na segunda colocação Sapindaceae com 148 indivíduos, em terceiro lugar, com apenas uma espécie, Rosaceae (*Prunus brasiliensis*) com 103 indivíduos. Foi observado apenas um indivíduo da espécie *Araucaria angustifolia*.

3.2 O SENSORIAMENTO REMOTO E A ENERGIA ELETROMAGNÉTICA

3.2.1 O Sensoriamento Remoto

O sensoriamento remoto é a ciência e arte de obter informações sobre um objeto, área, ou fenômeno através da análise de dados adquiridos a partir de um sistema que não está em contato com esse objeto, área ou fenômeno sob investigação (LILLESAND et al. 2007). Barrett e Curtis (1999) resumem sua definição como “a ciência da observação à distância”. Conforme bem relembra Eastman (1998) os olhos humanos (e de outras espécies animais) são um excelente exemplo de um dispositivo de sensoriamento remoto. Informações sobre o ambiente são reunidas julgando-se a quantidade e a natureza da reflectância da energia da luz (no caso, luz visível)

proveniente de uma fonte externa como o Sol ou uma lâmpada, conforme ela é refletida no campo visual. A primeira forma de sensoriamento remoto foi a fotografia aérea, desenvolvida em 1826, já que na conceituação básica da ciência em questão estão embutidas as etapas de observação, captação e armazenamento da informação (HOWARD, 1991; BARRETT e CURTIS, 1999). As primeiras tentativas de fotografar regiões distantes utilizavam torres ou montanhas para posicionar a câmara a uma altura considerável acima dos objetos (CENTENO, 2003). Posteriormente balões (a partir de 1859) e aviões (1909), que podiam sobrevoar regiões mais distantes foram empregados (CENTENO, 2003). As demandas militares das duas grandes guerras mundiais (especialmente a segunda) foram grandes estimuladoras da pesquisa e do uso do sensoriamento remoto (BARRETT e CURTIS, 1999), além do desenvolvimento de tecnologias, principalmente relacionadas a computadores e a tecnologias espaciais. Os mesmos autores relembram que o termo “sensoriamento remoto” foi utilizado pela primeira vez, tanto para designar uma área nova da ciência como para se referir a um conjunto de técnicas relacionadas ao ambiente humano. Não há como não mencionar a “Guerra Fria”, que levou os Estados Unidos e a então União Soviética a disputar grandes batalhas visando o domínio das tecnologias espaciais, como outro grande propulsor da ciência sensoriamento remoto como a conhecemos hoje. Foi assim que, em 1972, o ERTS (*Earth Resources Technology Satellite*) da NASA foi colocado em órbita com a finalidade específica de coletar dados a respeito dos recursos naturais da Terra, dando origem à série Landsat (CENTENO, 2003).

3.2.2 A Energia Solar e o Espectro Eletromagnético

A energia solar é a base de todos os princípios em que se fundamenta o sensoriamento remoto (MOREIRA, 2001). Estudos realizados em laboratórios no final do século XIX e início do século XX tinham a finalidade de esclarecer a natureza da radiação solar. O que se concluiu é que a radiação solar apresenta propriedades de onda e de partículas discretas (fótons), dando origem a duas teorias, a corpuscular e a ondulatória. A primeira afirma que a radiação solar é emitida através de pequenos pulsos ou feixes discretos de fótons, denominada de *quanta*, e que esta radiação se propaga como quantas individuais. A segunda diz que a radiação solar propaga-se em linha reta, através de dois campos, um elétrico e um magnético, em movimento ondulatório

(MOREIRA, 2001; CENTENO, 2003), sendo que muitas características da radiação eletromagnética são mais facilmente descritas por esta última teoria (LILLESAND et al. 2007).

A energia eletromagnética é associada a um largo escopo de comprimentos de ondas que formam o espectro eletromagnético. Comprimento de onda (λ) é a distância entre os picos sucessivos de onda e sua frequência (ν), determinando o número de picos de ondas que passam por um dado ponto em um segundo. Comprimento de onda e frequência são inversamente relacionados, ou $c = \nu\lambda$, onde c é uma constante (velocidade constante: 3×10^8 m/seg) de energia eletromagnética em um vácuo, como no espaço exterior (LILLESAND et al. 2007). A instrumentação ligada ao sensoriamento remoto detecta e armazena informações sobre a energia radiante, também chamada de radiação solar ou eletromagnética (REM). O escopo envolvendo possíveis comprimentos de ondas REM e frequências é denominado espectro electromagnético. Por razões práticas, é dividido em regiões de energia com comprimentos de onda que são semelhantes em alguns aspectos (ROBINSON et al. 1995; TSO e MATHER, 2001). Os comprimentos de onda são geralmente medidos em micrômetros (1×10^{-6} m, μm).

A luz visível é somente uma das muitas formas de expressão da energia eletromagnética. Ondas de rádio, calor, raios ultravioleta e raios-X são outras formas conhecidas. Assim, além do sol, toda a matéria com temperatura acima do zero absoluto (correspondente a -273° Celsius) emite continuamente radiação eletromagnética. Objetos terrestres são também fontes de radiação, embora a composição espectral e magnitude sejam consideravelmente diferentes das do sol (LILLESAND et al. 2007).

Quando a radiação eletromagnética recebe a denominação “óptica”, diz-se que ela se limita à região do espectro eletromagnético que considera o intervalo de comprimento de onda do visível ao infravermelho de ondas curtas (0,4 a 2,5 μm), essa faixa sendo, então, denominada de radiometria óptica espectral ou somente radiometria óptica (MENESES, 2001). A porção do visível é relativamente pequena, mas seus limites precisos não podem ser demarcados, já que variam em função da capacidade diferenciada entre olhos humanos (HOWARD, 1991). Na média se considera que o olho humano é sensível somente à faixa que vai de aproximadamente 0,4 μm a 0,7 μm . A cor azul é descrita em uma escala aproximada de 0,4 a 0,5 μm , o verde de 0,5 a 0,6 μm e o vermelho de 0,6 a 0,7 μm . A energia ultra violeta (UV) é contígua à extremidade azul da parcela do visível do espectro e a parte contígua ao vermelho na parcela do visível é o infravermelho, que se divide em três partes: o infravermelho próximo (de 0,7 a 1,3 μm), o

infravermelho médio (de 1,3 a 3 μm) e o infravermelho termal (de 3 a 14 μm). Comprimentos de onda maiores que os mencionados compreendem a porção microondas do espectro (**Figura 1**) (LILLESAND et al. 2007).

Para o sensoriamento remoto, a energia eletromagnética com comprimentos de onda menores seria muito útil, por estar associada à maior quantidade de energia e por poder ser mais facilmente medida. No entanto, esse tipo de energia é fortemente perturbado pela atmosfera. Já a região acima do vermelho (infravermelho) possui comprimento de ondas maiores e é importante para o sensoriamento remoto, pois nela é possível registrar alguns fenômenos naturais com muita clareza (CENTENO, 2003). Acima desta faixa estão localizadas as microondas, energia utilizada pelos radares.

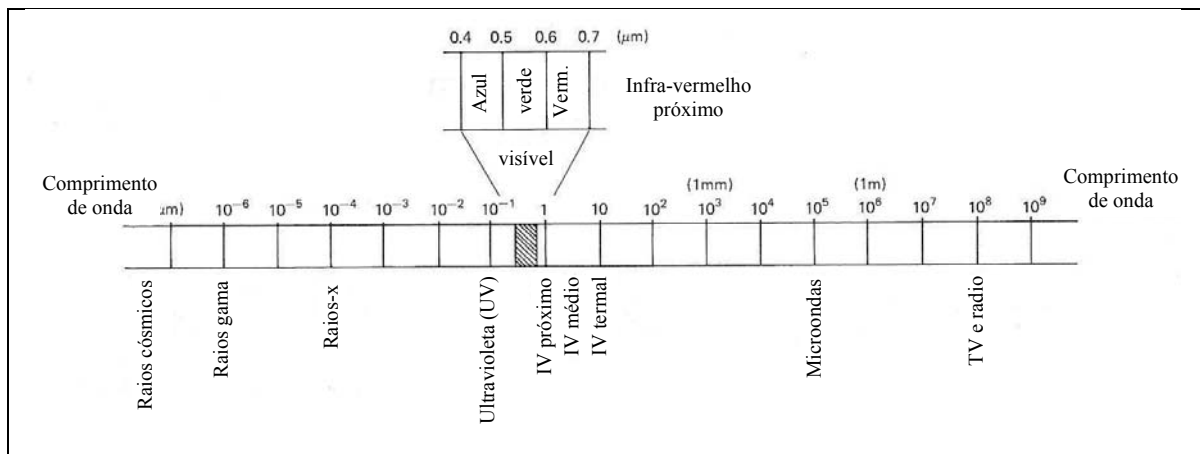


FIGURA 1: Espectro eletromagnético
Fonte: Lillesand et al. (2007)

3.3 COMPORTAMENTO ESPECTRAL DOS ALVOS

Existe interação entre a REM e os diferentes alvos (objetos) da superfície da Terra. Esta energia é emitida por ou refletida de feições no solo e transmitida sob a forma de ondas. (ROBINSON et al. 1995). A radiometria espectral é uma das mais importantes ferramentas do sensoriamento remoto, pois é através de medidas radiométricas de laboratório ou campo se detecta com qual intensidade cada material (solo, rocha, vegetação, água) reflete a radiação eletromagnética nos diferentes comprimentos de onda permitindo explicar e entender como cada um desses objetos irá aparecer em uma imagem satelitária. Portanto, é necessário entender como

interage a energia contida na radiação eletromagnética com a energia contida nos átomos e moléculas dos diferentes materiais (alvos) ao incidirem sobre eles (MENESES, 2001). Quando a energia eletromagnética incide sobre alguma feição da superfície da Terra três interações fundamentais da energia com essas feições são possíveis. Várias frações da energia incidente são refletidas, absorvidas e/ou transmitidas. As proporções dessas interações podem variar para as diferentes feições ou alvos, dependendo do tipo de material e de condições específicas, como tipo de sensor, condições atmosféricas etc.. Essas diferenças nos permitem distinguir as muitas feições presentes em uma imagem (LILLESAND et al. 2007).

3.3.1 Vegetação

O comportamento espectral da vegetação está associado às características das folhas, pois são elas os elementos mais visíveis numa imagem satelitária (CENTENO, 2003). Sua interação com características do solo podem também ser importantes para a caracterização do comportamento espectral da vegetação.

Conforme já mencionado anteriormente, quando o fluxo de energia (luz solar) incide sobre um objeto, ele interage com o seu material constituinte. Devido a essa interação, a energia pode ser absorvida, refletida ou transmitida, total ou parcialmente e, conseqüentemente, a absorção, reflexão e transmissão (**Figura 2**) são as três propriedades radiométricas dos objetos (HOWARD, 1991).



FIGURA 2: Propriedades radiométricas dos objetos: absorvância (A), reflectância (R) e transmitância (T)
Fonte: CCRS (1999)

Os estudos envolvendo a reflectância espectral de dosséis têm sido na sua maioria motivados pela necessidade de se relacionar os níveis de cinza (que definem os padrões de existentes em fotografia aérea e imagens orbitais de diferentes regiões espectrais) às diferentes tipologias vegetais (PONZONI, 2001).

Na região do visível a reflectância é muito baixa devido à absorção da energia pelos pigmentos da planta para a realização da fotossíntese (PONZONI e SHIMABUKURO, 2007). A folha absorve muita energia nos comprimentos de onda em torno de 0,45 e 0,65 μm , referentes à região do azul e vermelho respectivamente (CENTENO, 2003); (PONZONI, 2001). A reflectância em torno de 0,45 e 0,65 μm é muito baixa, originando um pico máximo aproximadamente em 0,54 μm (entre o azul e o vermelho) que corresponde à cor verde (CENTENO, 2003). Na região do infravermelho próximo, a reflectância apresenta valores elevados devido ao espalhamento interno sofrido pela radiação (PONZONI, 2001); (PONZONI E SHIMABUKURO, 2007) em função da disposição da estrutura morfológica da folha, aliado ainda ao espalhamento múltiplo entre as diferentes camadas de folhas (PONZONI E SHIMABUKURO, 2007). No infravermelho médio acontece uma nova queda nos valores de reflectância devido à presença de água no interior da folha (PONZONI, 2001; CENTENO, 2003; PONZONI e SHIMABUKURO, 2007). Os picos de absorção acontecem em torno de 1,4, 1,9 e 2,7 μm (CENTENO, 2003; PONZONI e SHIMABUKURO, 2007). (**Figura 3**).

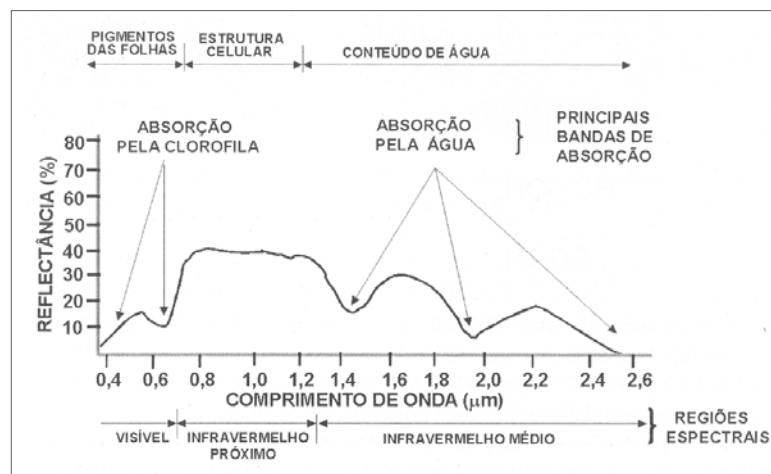


FIGURA 3: Curva de reflectância de uma folha verde
Fonte: Moreira (2001)

3.3.2 Solo

A reflectância do solo, não somente da luz visível, mas também de todos os outros comprimentos da faixa do espectro óptico, é uma propriedade cumulativa que deriva do comportamento espectral dos constituintes minerais, orgânicos e fluidos que, combinados, compõem os solos. As variações espectrais entre os diferentes tipos de solos estão normalmente associados a processos de absorção da radiação eletromagnética atribuída a esses componentes (NETTO, 2001). Os solos não são homogêneos, mas, sim, uma mistura complexa de elementos com propriedades químicas diferentes. Cada elemento participante dessa mistura contribui de maneira diferente para a resposta espectral do solo como um todo (CENTENO, 2003). Segundo este autor, alguns fatores que afetam a reflectância dos solos são:

- Conteúdo de umidade: a água no solo age de maneira similar à das plantas pois os picos de absorção acontecem nas mesmas faixas (1,4, 1,9 e 2,7 μ m).
- Matéria orgânica: em solos, o teor de matéria orgânica varia de 0,5 a 5% e sua presença diminui a reflectância, pois escurece a mistura. O efeito da matéria orgânica é sentido principalmente entre 0,48 e 0,53 μ m, no visível.
- Óxido de ferro: a presença dessa substância confere ao solo uma coloração amarelada. A reflectância do solo diminui significativamente quando o teor de ferro aumenta, principalmente entre 0,5 e 0,64 μ m. Sob a forma de óxido de ferro férrico, a absorção acontece na faixa do infravermelho, entre os comprimentos de onda 0,7 e 0,9 μ m.
- Propriedades físicas do solo: a granulometria das partículas (areia, silte, argila) condiciona a rugosidade da superfície do solo, mudando sua capacidade de refletir energia incidente.

3.3.3 Água

No geral, a água absorve grande parte da energia que nela incide, ou seja, a água possui alta absorvância, principalmente em comprimentos de ondas maiores do que 0,7 μ m, nos quais absorve toda a energia incidente. Isso se torna uma propriedade importante para o mapeamento dos corpos d'água (CENTENO, 2003). A energia solar que é absorvida pela massa de água é, em parte, devida às moléculas e aos elétrons da própria massa de água e em parte, devida às

partículas nela dispersas. O fitoplâncton e a matéria orgânica nos ecossistemas aquáticos são responsáveis por parte da absorção e espalhamento da radiação solar e determinam, até certo ponto, as propriedades ópticas da água. Na temperatura ambiente a água é encontrada em três estados: sólido, líquido e gasoso e em cada um desses estados a absorção da radiação eletromagnética ocorre de maneira diferente. Conforme comentado anteriormente, a água líquida absorve quase toda a energia que nela incide. Mesmo na faixa de 0,38 e 0,7 μm a reflectância é baixa, ultrapassando pouco mais de 5%. Na forma gasosa a água apresenta alta reflectância, em torno de 70% para todas as radiações contidas no espectro eletromagnético óptico (0,4 a 2,5 μm). A água em estado sólido, a neve, apresenta até 1,2 μm uma curva de reflectância maior do que no estado gasoso. A partir dessa faixa há um decaimento indicando uma absorção acentuada de energia nessa faixa (MOREIRA, 2001). As curvas de reflectância da água estão mostradas na **Figura 4**.

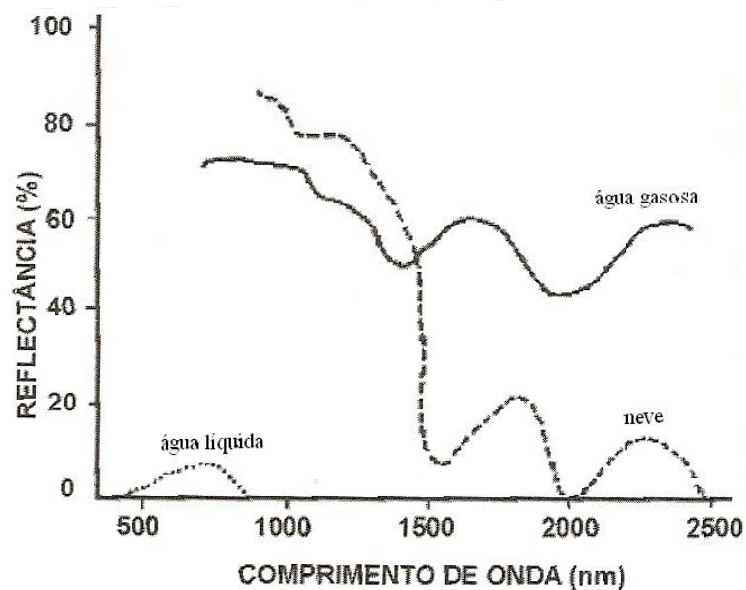


FIGURA 4: Curva de reflectância da água nos três estados físicos: sólido, líquido e gasoso
Fonte: Moreira (2001)

3.4 SENSORIAMENTO REMOTO NA FLORESTA OMBRÓFILA MISTA

Desde a descoberta de que os vegetais extraem da radiação eletromagnética emitida pelo sol parte da energia que necessitam para viver a interação entre a radiação e a vegetação passou a ser estudada detalhadamente, ainda mais com o advento do sensoriamento remoto no qual o

mapeamento da distribuição geográfica da cobertura vegetal, com base em suas características fisionômicas, ecológicas e florísticas ganhou impulso considerável, primeiramente com fotografia aérea e posteriormente com imagens orbitais (PONZONI, 2001). No Brasil, a aplicação das técnicas de sensoriamento remoto no estudo da vegetação teve início com os primeiros mapeamentos temáticos realizados na década de 40 a partir de fotografias aéreas (PONZONI e SHIMABUKURO, 2007).

Vários trabalhos que utilizam técnicas de sensoriamento remoto para o mapeamento da vegetação podem ser citados. Entre eles Dlugosz (2005), Dlugosz et al. (2005), Dlugosz et al. (2007) que utilizaram imagens de alta resolução do satélite Ikonos para o mapeamento de tipologias da Floresta Ombrófila Mista. Utilizaram técnicas de segmentação por crescimento de regiões para posterior classificação da imagem. Com isso geraram um mapa de uso do solo com treze classes, incluindo as tipologias, áreas de agricultura, lago e solo exposto.

Oliveira (1999) utilizou técnicas de sensoriamento remoto em uma área no estado do Paraná que pertencia ao domínio de Floresta Ombrófila Mista e que continha em seu interior a Floresta Nacional de Irati. O objetivo do estudo era a classificação de imagens Landsat 5 e estudo da resposta espectral de tipologias da Floresta Ombrófila Mista bem como de plantios puros de outras espécies de coníferas, notadamente do gênero *Pinus*.

Gantzel (1979), em seu trabalho, utilizou imagens monocromáticas do satélite Landsat II/MSS juntamente com diapositivos coloridos, interpretados visualmente e pôde avaliar quantitativamente as formações florestais de *Araucaria angustifolia*. Este autor estabeleceu cinco tipologias florestais sendo que três delas foram definidas conforme a densidade de Araucária em floresta natural e as outras duas estavam relacionadas aos plantios dessa espécie com outras essências florestais, como *Pinus* spp. Os resultados mostraram que foi possível identificar somente a Floresta de Araucária. Os plantios de *Araucaria angustifolia* e *Pinus* spp não foram possíveis de serem diferenciados.

Watzlawick (2003), utilizando imagens do satélite Ikonos II, realizou estimativas de biomassa arbórea e carbono orgânico arbóreo em Floresta Ombrófila Mista e plantios homogêneos de *Pinus taeda* e *Araucaria angustifolia*.

Acco (2004) realizou estimativas de micronutrientes da Floresta Ombrófila Mista Montana utilizando também imagens do satélite Ikonos II. Este estudo mostrou ser possível modelar as equações que estimaram a quantidade de micronutrientes.

Disperati e Oliveira-Filho (2005) utilizaram fotografias aéreas de alta resolução para o delineamento semi-automático de copas de árvores através de suas sombras periféricas. Este estudo foi realizado em uma área de Floresta Ombrófila Mista da Floresta Nacional de Irati-PR e os autores não obtiveram resultados satisfatórios. Segundo eles isso ocorreu devido ao tipo da floresta considerada e não à técnica utilizada e conduziu a resultados não conclusivos sobre o reconhecimento do padrão típico das copas de *Araucaria angustifolia* e *Ocotea spp.*

3.5 O SATÉLITE ALOS

O *User Handbook* (JAXA, 2007) oferece uma descrição completa das características, missão, objetivos e aplicações do satélite ALOS, tópicos esses abordados na sequência do presente trabalho.

O programa Japonês de observação da Terra por satélite consiste em duas séries: satélites usados, principalmente, para observação atmosférica e marinha e aqueles usados principalmente para observação da Terra. O *Advanced Land Observing Satellite (ALOS)* segue o *Japanese Earth Resources Satellite-1 (JERS-1)* e o *Advanced Land Observing Satellite (ADEOS)* e utiliza tecnologia avançada de observação da Terra. O ALOS é usado para cartografia, observação regional, monitoramento de desastres e levantamento de recursos (JAXA, 2007).

O satélite ALOS foi lançado em 24 de janeiro de 2006, no veículo H-IIA em *Tanegashima Space Center*, Japão, pela Agência Espacial Japonesa (*Japan Aerospace Exploration Agency - JAXA*). Está localizado em uma altitude de 691,65 km em relação à superfície terrestre com inclinação de 98,16°. Sua órbita é solar-síncrona com um ciclo completo de 46 dias e sub-ciclo de 2 dias (**Figura 5**). Está equipado com três instrumentos sensores: 1) o *Panchromatic Remote-Sensing Instrument for Stereo Mapping (PRISM)*, que é o sensor com uma banda pancromática; 2) o *Advanced Visible and Near Infrared Radiometer type 2 (AVNIR-2)*, sensor multispectral, que possui 4 bandas espectrais abrangendo a região espectral do visível e infra-vermelho próximo e que é responsável, juntamente com o PRISM pela observação da cobertura da Terra e 3) *Phased Array type L-band Synthetic Aperture Radar (PALSAR)* que é o sensor radar, para observação da Terra dia e noite e em qualquer condição de tempo (JAXA, 2007).

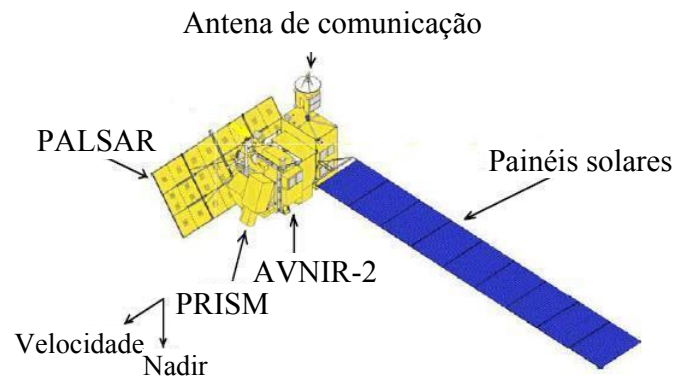


FIGURA 5: Satélite ALOS com seus três sensores
 Fonte: Adaptado de JAXA (2007)

3.5.1 Sistemas Sensores do Satélite ALOS

3.5.1.1 *Panchromatic Remote-Sensing Instrument for Stereo Mapping (PRISM)*

O *Panchromatic Remote-Sensing Instrument for Stereo Mapping (PRISM)* é um radiômetro pancromático com resolução espacial de 2,5m no nadir (**Tabela 1**). Possui três sistemas óticos independentes: nadir, *forward* e *backward*, que permitem visão estereoscópica e que podem produzir um modelo digital de superfície altamente acurado (**Figura 6a**). Cada telescópio consiste em três espelhos e diversos detectores CCD (*Charge Coupled Device*) para exploração *push-broom*. A visada do nadir possui uma largura de imageamento de 70km enquanto que as visadas *forward* e *backward* possuem uma largura de 35km (**Figura 6b**). Estes dois últimos telescópios estão inclinados a $\pm 24^\circ$ do nadir para realizar uma razão base-altura de 1.0. A largura do campo de visada (FOV) do sensor PRISM fornece três imagens *stereo (triplet)*, completamente sobrepostas, com uma largura de 35km. A resolução radiométrica desse sensor é de 8 *bits*.

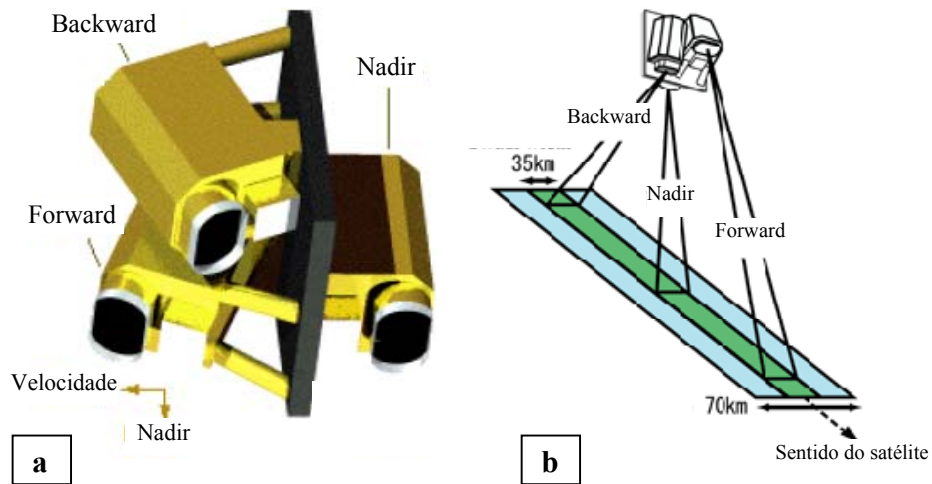


FIGURA 6: (a) Sensor PRISM; (b) Visadas *Forward*, *Nadir* e *Backward* e faixa de imageamento.
Fonte: Adaptado de JAXA (2007)

3.5.1.2 Advanced Visible and Near Infrared Radiometer Type 2 (AVNIR-2)

O *Advanced Visible and Near Infrared Radiometer type 2* (AVNIR-2), sensor criado para observação da Terra e zonas costeiras (**Figura 7a**), é um radiômetro multiespectral que possui 4 bandas com resolução espacial de 10m, abrangendo a região do visível e infra-vermelho próximo (**Tabela 1**). Sua resolução radiométrica é de 8 bits. Fornece imagens de cobertura e uso da Terra para monitoramento ambiental regional. O ângulo de apontamento é de $\pm 44^\circ$, para observação imediata de áreas afetadas por desastres naturais (**Figura 7b**). Este sensor é o sucessor do AVNIR que estava a bordo do *Advanced Earth Observing Satellite* (ADEOS), lançado em agosto 1996 (JAXA, 2007).

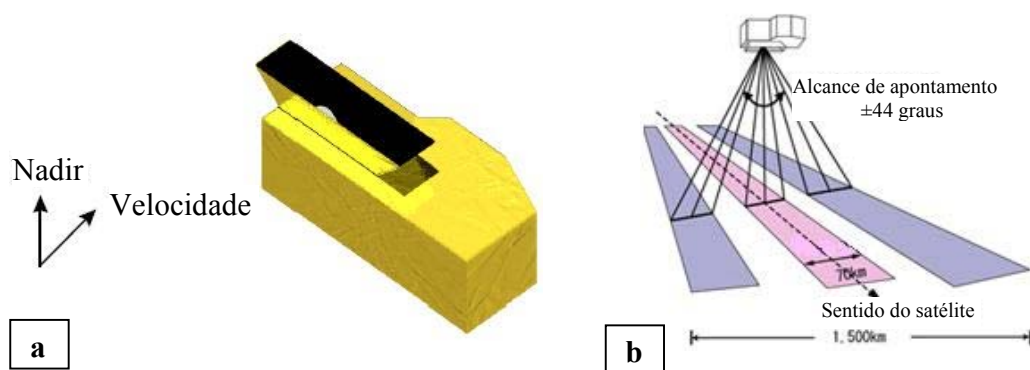


FIGURA 7: (a) Sensor AVNIR-2; (b) Modo de observação do AVNIR-2
Fonte: Adaptado de JAXA (2007)

TABELA 1: Bandas espectrais do satélite ALOS e suas respectivas resoluções espaciais

Satélite ALOS		
	Bandas espectrais (μm)	Resolução espacial (m)
Banda 1	0,42 - 0,50	10
Banda 2	0,52 - 0,60	10
Banda 3	0,61 - 0,69	10
Banda 4	0,76 - 0,89	10
Pan	0,52 - 0,77	2,5

Fonte: Adaptado de JAXA (2007)

3.5.1.3 Phased Array Type L-band Synthetic Aperture Radar (PALSAR)

O *Phased Array type L-band Synthetic Aperture Radar* (PALSAR) é um sensor micro-ondas ativo que opera na banda L (JAXA, 2007). Por não sofrer influência da atmosfera, condições climáticas e não depender da luz, o PALSAR opera dia e noite e livre de nuvens (**Figura 8**). Sua resolução espacial varia de 7 a 100m. É o segundo radar de abertura sintética (SAR) japonês. Sua alta resolução é para o modo convencional, mas o PALSAR tem outro modo de observação vantajoso, o ScanSAR, no qual permite adquirir imagens de 250 a 350 km de largura, ou seja, de três a cinco vezes maior do que as imagens SAR convencionais e por isso é considerado útil para monitoramento de florestas tropicais e da extensão do gelo oceânico (JAXA, 2006a). O PALSAR foi desenvolvido juntamente pela JAXA e *Japan Resources Observation System Organization* (JAROS).

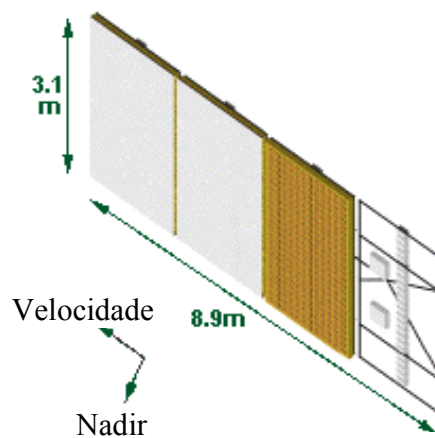


FIGURA 8: Sensor PALSAR

Fonte: Adaptado de JAXA (2007)

3.5.2 Formato e Níveis de Processamento dos Dados

As imagens ALOS (multiespectrais e pancromática) são fornecidas em diversos níveis de processamento, cabendo ao usuário definir o nível desejado para seu trabalho. A descrição desses produtos pode ser encontrada em *ALOS Product Description* (JAXA, 2006b) e também no site do Instituto Brasileiro de Geografia e Estatística (IBGE). Os níveis estão descritos a seguir:

a) nível 0: Neste nível, os dados dos sensores AVNIR-2 e PRISM são brutos, descarregados diretamente do satélite para a terra. Este produto é dividido em um tamanho equivalente a uma cena;

b) nível 1A: também dados brutos extraídos das linhas do nível 0. Informações auxiliares como coeficientes de calibração radiométricos estão anexados, mas não aplicados;

c) nível 1B1: neste nível a calibração radiométrica é efetuada nos dados do nível 1A e o coeficiente de calibração é aplicado. A imagem passa a possuir os pixels alinhados em relação à trajetória do satélite;

d) nível 1B2: neste nível a correção geométrica é efetuada nos dados do nível 1B1, e possui as seguintes opções:

- nível 1B2-R (*Geo-referenced*): as imagens neste nível possuem os pixels alinhados com a trajetória do satélite e é necessário que o usuário aplique uma rotação na imagem para torná-la orientada para o norte. Calibração radiométrica e coeficientes de correção geométrica estão aplicados. Coordenadas em projeção UTM;
- nível 1B2-G (*Geo-coded*): neste nível a imagem possui os pixels alinhados com a grade de projeção UTM. Para que a imagem geocodificada possua a rotação para o norte é necessário que os quatro cantos dessa imagem toquem os lados da imagem georreferenciada. Possui calibração radiométrica e os coeficientes de correção geométrica estão aplicados.

As imagens ALOS são adquiridas em formato CEOS (*Committee on Earth Observation Satellites*) – BSQ, banda seqüencial.

3.5.3 Aquisição de Imagens ALOS

No Brasil, as imagens ALOS podem ser adquiridas através do site do IBGE, www.ibge.gov.br/alos. Esta instituição fez uma parceria com a empresa ASF (*Alaska Satellite Facility*), que é o braço operacional responsável pelas imagens ALOS nas Américas. Foi assinado um contrato de cooperação científica entre as duas instituições sendo que o IBGE tornou-se responsável pela distribuição das imagens para órgãos do Governo Federal, instituições de pesquisa e demais usuários não comerciais do Brasil.

As imagens ALOS custam US\$ 125,00 (valores em março de 2009) e podem ser adquiridas pelo site IBGE através da ferramenta de busca no catálogo de imagens. As opções de busca incluem coordenadas geográficas (latitude e longitude) da área de interesse ou o identificador da cena (ID). É importante analisar a cobertura de nuvens e se necessário utilizar a opção “gerar Kml”, que abrirá o retângulo envolvente da cena no aplicativo *Google Earth*, sendo possível verificar se a área de interesse está dentro dos limites da cena.

Além do IBGE, existem também outros sites de busca de imagens ALOS, como o site da URSA (*User Remote Sensing Access*) pertencente ao ASF, e o CROSS (*Online satellite images search & order system*), que é o site da empresa *Remote Sensing Technology Center of JAPAN* (Restec), que disponibiliza imagens ALOS de todo o globo. Embora o acervo do CROSS seja maior do que o da URSA e do IBGE, a busca normalmente se inicia no catálogo do IBGE por disponibilizar resultados mais rápidos. A procura no CROSS é muito lenta e a página é mais sujeita a erros. Na **Figura 9** é apresentado um roteiro para a busca das imagens na internet.

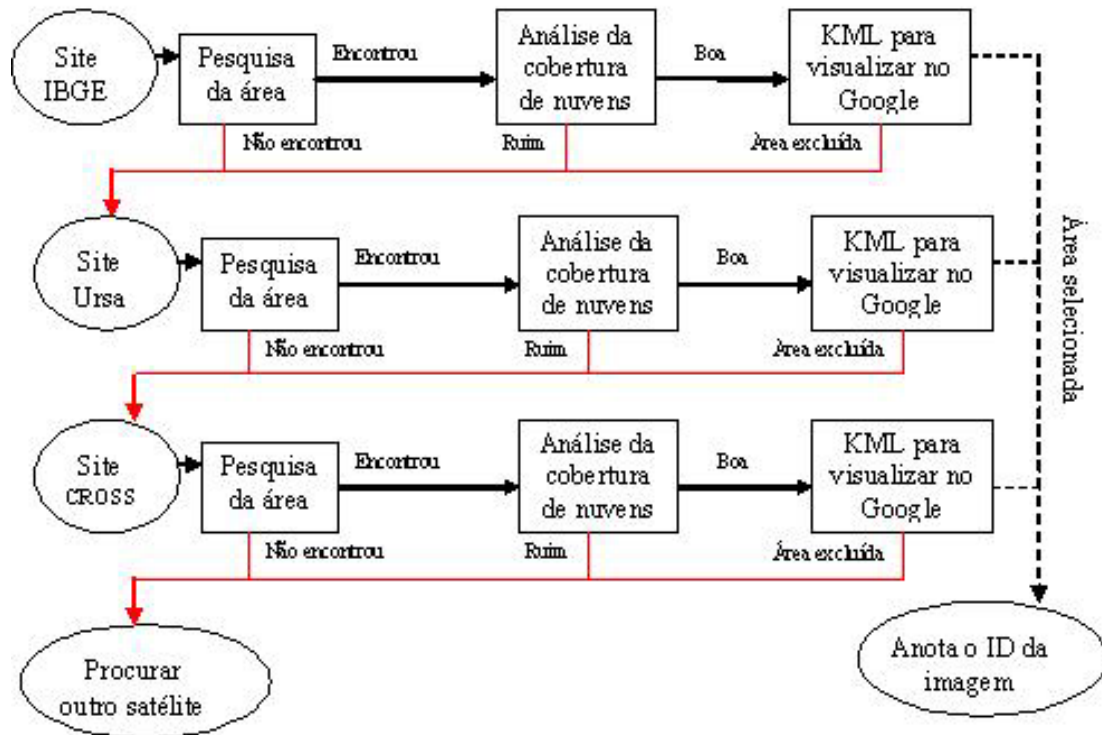


FIGURA 9: Roteiro para busca de imagens ALOS na internet

3.5.4 Conceito da Missão ALOS

O satélite ALOS foi criado com a missão de observar e obter imagens de todo o planeta para fins de monitoramento de desastres ambientais, levantamento de recursos naturais e também para dar suporte à cartografia (JAXA, 2007).

Os objetivos da missão ALOS são principalmente o de promover pesquisa científica e utilização dos dados em amplas categorias que variam desde as ciências ambientais até as ciências computacionais (JAXA, 2007). Alguns deles estão citados a seguir:

- a) pesquisas em ecossistemas terrestres, agricultura e florestas;
- b) Monitoramento e mapeamento do uso e cobertura da Terra: para este fim o principal sensor é o AVNIR-2. Ele é uma ferramenta importante para a caracterização da floresta e áreas inundáveis, pois é sensível às propriedades espectrais da vegetação, além de poder gerar um ganho em informações por ser possível o uso de composições coloridas.

Aplicando métodos de classificação supervisionada e não supervisionada nas quatro bandas do sensor AVNIR-2, é possível avaliar condições e as mudanças que ocorrem na superfície da Terra em tempo real. Este método de extrair informações sobre o uso da terra a partir dos dados AVNIR-2 pode ser aplicado também para áreas de floresta na tentativa de identificar espécies arbóreas.

3.6 PROCESSAMENTO DE IMAGENS DIGITAIS

3.6.1 Retificação Geométrica

3.6.1.1 *Correção Geométrica*

As imagens produzidas por sensores remotos apresentam uma série de distorções espaciais, resultantes dos seguintes fatores: a) rotação da Terra; b) curvatura da Terra; c) movimento no espelho de imageamento; d) variações de altitude, posição e velocidade da plataforma; e) distorções de panorama; f) distorção topográfica (KURASZ, 2005). O georreferenciamento é uma operação de transformação geométrica que relaciona coordenadas de imagem (linha, coluna) com coordenadas de um sistema de referência, corrigindo distorções espaciais e inserindo precisão cartográfica no posicionamento dos objetos, superfícies ou fenômenos representados em imagens (DLUGOSZ, 2005). Uma alternativa para a correção de erros geométricos consiste em utilizar pontos de coordenadas conhecidas, visíveis na imagem, para ajustar um modelo que permita estimar o conjunto inteiro das deformações (CENTENO, 2003). Os pontos de controle são feições possíveis de ser encontradas tanto no mapa/imagem base como na imagem a ser georreferenciada (SPRING, 2008a). Estes pontos de controle devem ter sua distribuição espacial uniforme por toda a imagem, ou seja, não devem estar aglomerados em somente uma parte da imagem e nem alinhados (CENTENO, 2003).

3.6.1.2 *Reamostragem de Pixels*

Após a correção dos erros de cada ponto de controle, aplica-se um modelo matemático de transformação na imagem original produzindo uma nova imagem corrigida de acordo com a

projeção cartográfica do mapa onde foram coletados os pontos de controle. Para se obter essa nova imagem deve-se aplicar um método de reamostragem, que pode ser o método do “vizinho mais próximo”, “bilinear” ou “convolução cúbica”. A escolha de qual método usar depende do uso que se fará da imagem. O método do “vizinho mais próximo” é o que mais preserva o valor do número digital, uma vez que não envolve nenhum cálculo de média e por isso é o mais indicado para imagens que serão submetidas à classificação (CROSTA, 1992).

3.6.2 Transformação de Dados

A maioria das imagens de sensoriamento remoto possui a vantagem de gerar novas imagens a partir de transformações radiométricas (PONZONI e SHIMABUKURO, 2007). Estas transformações podem realçar informações que não são muito visíveis nas imagens originais ou podem preservar o conteúdo das informações em um número reduzido de bandas transformadas (PONZONI e SHIMABUKURO, 2007). A transformação de dados de sensores remotos é uma maneira de reduzir o volume de dados, através de técnicas estatísticas e operações entre bandas. Muitas vezes o uso de dados espectrais transformados pode ser mais conveniente para uma classificação de imagens do que a utilização dos dados originais (MOREIRA, 2001).

3.6.2.1 Componentes Principais (PC's)

Um dos principais métodos para a redução da dimensionalidade dos dados espectrais e concentração de informação em poucas bandas não correlacionadas entre si é a técnica de Componentes Principais. As Componentes Principais são combinações lineares dos valores dos dados originais multiplicados por coeficientes de transformação apropriados, conhecidos como autovetores. Estes coeficientes são derivados da matriz de variância/covariância dos dados da imagem original (LILLESAND et al. 2007). A primeira componente irá conter a informação que é comum a todas as bandas originais, a segunda irá conter a feição espectral mais significativa do conjunto e assim por diante até a última componente que concentrará a informação restante (CROSTA, 1992).

3.6.2.2 Índices de Vegetação

Os índices de vegetação são amplamente utilizados em várias aplicações para o estudo da vegetação (PONZONI E SHIMABUKURO, 2007). As transformações realizadas para a obtenção desses índices são operações entre bandas, como adição, subtração e razão. Possuem, como principal função, realçar a contribuição espectral da vegetação, enquanto minimizam a contribuição espectral do solo, ângulo solar, vegetação senescente e da atmosfera (WATZLAWICK, 2003). Um dos índices de vegetação mais conhecidos é o *Normalized Difference Vegetation Index* (NDVI), ou como é conhecido em português, Índice de Vegetação de Diferença Normalizada. Muito explorado até os dias de hoje, ele tem sido usado em diversos estudos de culturas agrícolas e florestais (PONZONI e SHIMABUKURO, 2007). Os índices de vegetação mais utilizados atualmente utilizam as bandas do vermelho e infra-vermelho na forma de razões. A radiação refletida por um dossel na região do vermelho resulta de um processo de absorção ocasionado pelos pigmentos fotossintetizantes. O contrário acontece com o infra-vermelho, onde o processo de reflexão dessa energia é dominante (PONZONI, 2001). O NDVI é calculado da seguinte maneira:

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

Onde:

NIR = reflectância na banda referente ao infravermelho próximo.

R = reflectância na banda referente ao vermelho.

Já o *Simple Ratio* (RVI), ou como é chamado em português, Índice de Vegetação de Razão Simples é semelhante ao NDVI, porém seus valores não são normalizados. Foi o primeiro índice criado e pode ser calculado como mostrado abaixo (PONZONI e SHIMABUKURO, 2007):

$$RVI = \frac{NIR}{R}$$

Onde:

NIR = reflectância na banda referente ao infravermelho próximo.

R = reflectância na banda referente ao vermelho.

Também é bastante utilizado o *Soil Adjusted Vegetation Index* (SAVI), ou Índice de Vegetação Ajustado para o Solo. O brilho do solo (principalmente os solos escuros) tem mostrado um aumento nos valores do RVI e NDVI (PONZONI e SHIMABUKURO, 2007). Huete (1988)¹ citado por Ponzoni e Shimabukuro, 2007) desenvolveu este índice que visa amenizar o efeito do solo. Para isso foi introduzida uma constante que tem esta função. Esta constante, “L”, possui valores variando de 0 a 1, sendo que os valores ótimos são: 1 para densidades baixas de vegetação; 0,5 para densidades médias e 0,25 para densidades altas. O SAVI é calculado da seguinte maneira:

$$SAVI = \frac{(NIR - R)}{(NIR + R + L)} \times (1 + L)$$

Onde:

NIR = refletância na banda referente ao infravermelho próximo.

R = refletância na banda referente ao vermelho.

L = fator de correção para o solo, adimensional.

3.6.3 Filtragens Espaciais: Texturas

As técnicas de classificação de imagens podem incluir dois tipos de informações, adicionais àquelas obtidas de valores espectrais. O primeiro conjunto é composto por informações auxiliares, como aquelas obtidas em mapas de solos, geologia ou modelos digitais de terreno. O segundo conjunto envolve medidas de textura do pixel e de sua vizinhança.

O uso da textura como variável em procedimentos que visam à classificação de imagens tem sido relativamente limitado, principalmente em sensores passivos, devido a duas questões:

- a) Dificuldade na definição operacional de textura, em termos de sua derivação dos dados da imagem;
- b) Custo computacional envolvido nos cálculos.

Segundo Mather (1999), o fato das variáveis relacionadas à textura possuírem algum grau de dificuldade em sua definição e delimitação, não deve levar pesquisadores à decisão de ignorá-las em um processo de classificação. Na opinião do autor, a decisão deveria vir exatamente na direção contrária, já que têm sido reportadas como sendo uma importante

¹ HUETE, A.R. **A soil-adjusted vegetation index (SAVI)**. Remote Sensing of Environment, 25, p.295-309, 1988.

contribuição para a habilidade de discriminação entre alvos de interesse quando a resolução espacial da imagem é suficiente para tal.

As variáveis relacionadas à textura tem sido enquadradas como filtragens espaciais e são úteis em processamento de imagens de sensoriamento remoto para realçar bordas, feições lineares de determinadas direções e padrões de textura. Isso é feito através do realce ou da supressão de determinadas frequências espaciais (CROSTA, 1992). Eventualmente, a suavização textural das regiões de uma imagem e a precisão de suas bordas são aspectos que podem influenciar no desempenho do processo de segmentação que precede à classificação temática de uma imagem, segundo Nascimento (1997).

A textura é uma característica importante usada na identificação de objetos ou regiões de interesse em uma imagem (HARALICK et al. 1973). Para Mather (1999) a textura pode ser definida como a variabilidade em “tom” dentro de uma vizinhança ou, ainda, como o padrão de relação espacial entre níveis de cinza em pixels de uma vizinhança. Outras definições descrevem a textura como um fenômeno organizado composto de padrões que obedecem alguma regra estatística (HARALICK e SHAPIRO, 1992² citado por LOCH e KIRCHNER, 1997) e pode ser unicamente medida analisando diferentes pixels adjacentes (CENTENO, 2003). Segundo este mesmo autor, a textura em imagens de satélite é o resultado da presença de objetos menores do que a resolução espacial da imagem e que são confundidos em um único pixel, resultando em variações na tonalidade ou cor. Para Schowengerdt (1983), não existe uma caracterização definitiva de textura, mas, todas as definições numéricas são descritas para rugosidade e contraste dos detalhes da imagem. O autor define textura como uma propriedade local de uma imagem, i.e., uma medida da intensidade de cinza entre os pixels e sua vizinhança.

A textura depende de dois fatores: a) escala de variação, que pode ser local ou regional e b) escala de observação. Assim, muitas imagens contêm regiões caracterizadas pelas variações no brilho, ao invés de um único valor de brilho. A textura se refere a variações no tom da imagem em função da escala. Para serem definidos como uma área textural distinta os níveis de cinza dentro da área devem ser mais homogêneos em termos de unidade do que áreas que tem diferente textura (RSI, 2005).

² HARALICK, R. M.; SHAPIRO L. G. 1992. Texture. In: **Computer and robot vision**. v. 1. New York , USA: Addison-Wesley Publishing Company. p. 453-494.

Com relação à forma como os dados são obtidos, a textura pode ser medida por quantidades estatísticas simples, como a variância do nível de cinza local, segundo Schowengerdt (1983). Já Barber e Ledrew (1991) afirmam que várias abordagens podem ser adotadas no que se refere à medidas de textura e, entre elas, mencionam: a) estatística; b) estrutura; c) valores estocásticos; e d) análise no domínio das frequências. Para os autores em questão, as abordagens estatísticas fazem uso da função de probabilidade da densidade de nível de cinza, e tal função é computada, frequentemente, como a probabilidade condicional de articulação de pares de pixel (níveis de cinza), numa determinada região da imagem. Dentre as aplicações enquadradas no grupo “estatística”, podem ser citados os filtros baseados tanto em medidas de ocorrência como de co-ocorrência. Os filtros de ocorrência usam o número de ocorrências de cada nível de cinza dentro da janela de processamento para o cálculo da textura (SULSOFT, 2007). Algumas das medidas de ocorrência disponíveis em *softwares* de processamento de imagens digitais são: amplitude dos dados, média, variância e entropia.

As primeiras aplicações de medidas de textura em imagens digitais obtidas por sensoriamento remoto foram feitas por HARALICK et al. (1973). Esses autores propuseram o que depois ficou conhecido como “matriz de co-ocorrência de níveis de cinza – GLCM, advinda da terminologia em inglês – grey-level co-occurrence matrix”, que representa a distância e a relação angular sobre sub-regiões de uma imagem de tamanho previamente especificado. Tso e Matter (2001) citam que a matriz GLCM também é conhecida como “grey-tone spatial-dependency matrix”. Em tal matriz, cada elemento da GLCM é uma medida da probabilidade de ocorrência de dois valores na “escala de cinza” separados por uma dada distância em uma dada direção. Em outras palavras, usam uma matriz de dependência espacial de níveis de cinza para calcular valores de textura. Esta é uma matriz de frequências relativas com cada valor de pixel ocorrendo em duas janelas de processamento vizinhas separadas por uma distância e direção específica. A matriz mostra o número das ocorrências do relacionamento entre um pixel e seu vizinho específico. Estes filtros incluem média, variância, homogeneidade, contraste, dissimilaridade, entropia, segundo momento e correlação (RSI, 2005).

O processo proposto por HARALICK et al. (1973) é mais facilmente visualizado via exemplos numéricos. As explicações a seguir apresentadas estão baseadas nos exemplos apresentados por MATHER (1999), LOCK e KIRCHNER (1997) e TSO e MATTER (2001).

Considere-se a **Figura 10 (a)**, que representa uma imagem 4 x 4 com quatro níveis de cinza. Já a **Figura 10 (b)** apresenta a forma geral da GLCM correspondente. Como exemplo, pode-se demonstrar que os valores contidos na célula (2,3) representam o número de vezes que os níveis de cinza 2 e 3 ocorrem em uma específica direção e distância, situação esta assinalada com a letra **d**. Na seqüência, as **Figuras 10 (c até f)** apresentam os resultados para quatro direções (a cada 45°), com $d=1$. As iniciais “H”, “V”, “LD” e “RD” representam, cada uma delas e, respectivamente, o número de cálculos para os ângulos 0° (horizontal), 90° (vertical), 135° (diagonal esquerda) e 45° (diagonal direita) como indicado pelas setas.

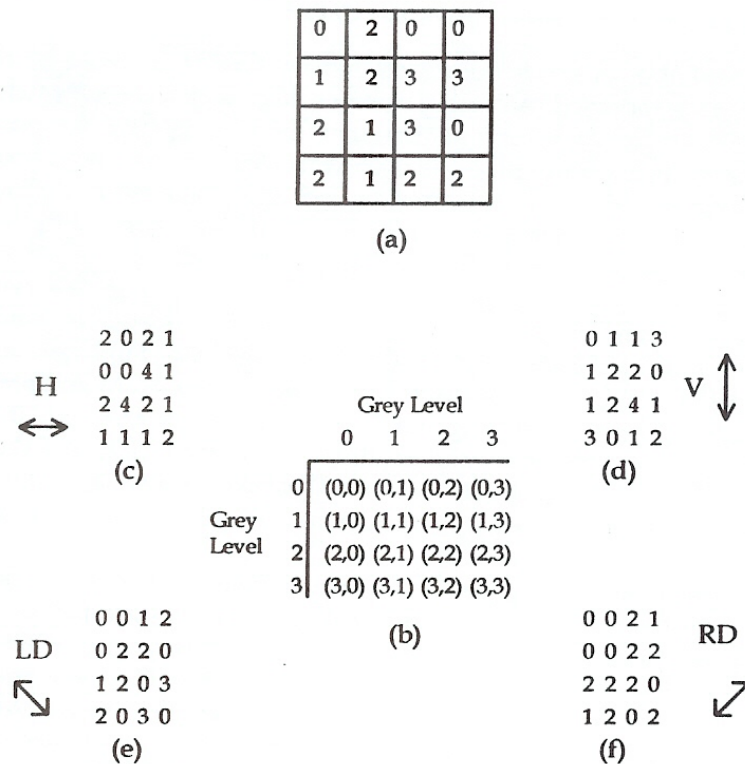


FIGURA 10: Cálculo da matriz de Co-ocorrência.

(a) Imagem 4 x 4 com quatro níveis de cinza. (b) Forma geral da GLCM com níveis de cinza de 0 a 3. O valor contido na célula (i , j) refere-se ao número de vezes que os níveis de cinza i e j ocorrem em uma específica direção e distância denominada “d”. (c até f) apresentam os resultados com respeito a cada pixel na imagem (a), para quatro direções com ângulos 0°, 90°, 135° e 45° e $d = 1$.

Fonte: TSO e MATTER (2001)

Ao invés de usar os valores de freqüência de uma GLCM diretamente, é prática comum a normalização da mesma para o escopo [0,1], para evitar efeitos de escala. O procedimento de

normalização para cada célula dentro da matriz pode ser facilmente calculado. Para a direção horizontal, com $d=1$, constarão $2 \times (\text{número de colunas menos } 1) \times \text{pares em cada linha}$. Por exemplo, no caso da **Figura 10 (a)**, o número total de pares da direção horizontal é computado por:

$2 * (4 - 1) * 4 = 24$. Assim, pode-se verificar que a soma das entradas na **Figura 10 (c)** é 24.

Para a direção vertical, a mesma regra pode ser aplicada, promovendo um total de $2 \times (\text{número de linhas menos } 1)$ multiplicado pelo número de colunas dos pares vizinhos mais próximos (**Figura 10 (d)**); assim, a soma de entradas é 24. Para o caso da diagonal esquerda, com $d = 1$, haverão $2 (\text{número de colunas menos um})$ pares para cada linha, exceto a última. Isto proporciona um total de $2 (\text{número de colunas menos um})$ multiplicado pelo $(\text{número de linhas menos um})$ pares vizinhos mais próximos da diagonal esquerda. (**Figura 10 (e)**); a soma de entradas é computada como: $2 * (4 - 1) * (4 - 1) = 18$.

Para a direção da diagonal da direita, por simetria, o número de pares próximos é o mesmo da diagonal da esquerda (**Figura 10 (f)**) e a soma das entradas é também 18. Depois que o número total de pares para cada matriz for obtido, cada uma delas é então normalizada pela divisão de cada célula pelo número total de pares.

Haralick et al. (1973) originalmente propuseram 32 feições texturais a serem derivadas de cada um dos quatro GLCM (MATTER, 1999). As quatro feições mais frequentemente usadas pelos pesquisadores envolvidos com o tema foram detalhadas por TSO e MATTER (2001). Segundo os autores, em todas, $p(i,j)$, denota a (i, j) ésima entrada em uma GLCM normalizada, e N_G denota o número de níveis de cinza em uma imagem quantizada. Algumas medidas estão mostradas a seguir (HARALICK et al. 1973).

a) Segundo Momento Angular (ASM) (HARALICK et al. 1973)

$$ASM = \sum_i \sum_j \{p(i,j)\}^2$$

A medida ASM é uma medida de homogeneidade da imagem. Em uma imagem homogênea, há poucas transições de níveis de cinza. Esta medida apresentará como saída o maior valor quando a frequência co-ocorrente $p(i,j)$ estiver concentrada em poucas áreas na GLCM, ou seja, na maior direção. Se $p(i,j)$ são próximos em valor, então ASM gerará um valor pequeno.

Por exemplo, ASM para o padrão mostrado na **Figura 11** (a), usando matriz (e), assume o valor de 0,19, enquanto para a **Figura 11** (b) (usando a matriz (f)), ASM assume o valor 0,06;

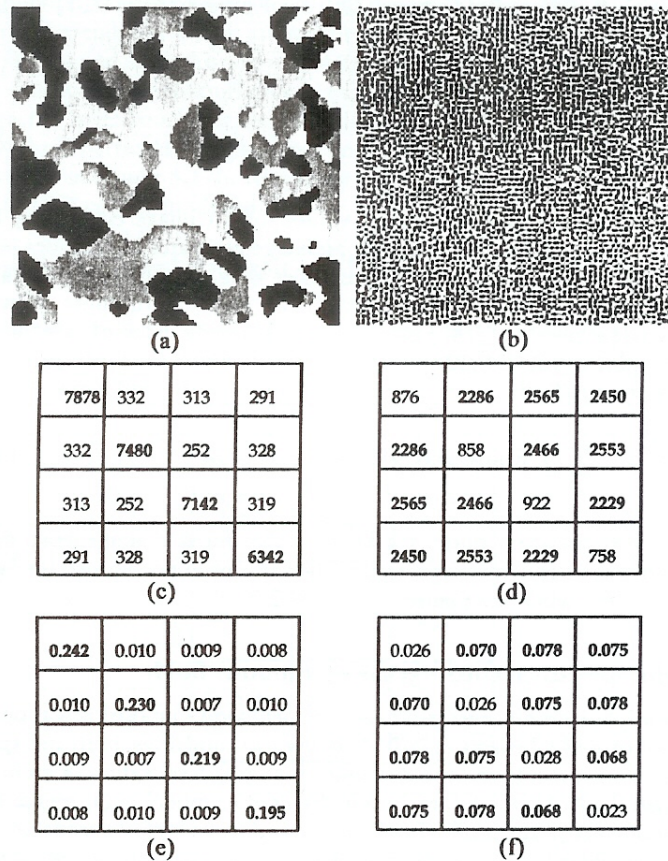


FIGURA 11: Padrões de texturas e suas GLCM

Dois padrões de textura em (a) e (b) e suas GLCM correspondentes (antes da normalização), mostradas em (c) e (d) e, depois da normalização, cada entrada tendo sido dividida por $2 * 127 * 128$ em (e) e (f)

Fonte: TSO e MATTER (2001)

b) Contraste (Con) (HARALICK et al. 1973)

$$Con = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i,j) \right\}_{|i-j|=n}$$

É uma medida do contraste ou da quantidade de variações locais presentes em uma imagem. Se em uma imagem existe maior quantidade de variações locais, os valores de contraste se apresentarão mais elevados do que uma imagem com menos variações locais;

c) Correlação (Corr) (HARALICK et al. 1973)

$$Corr = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

É uma medida das dependências lineares dos níveis de cinza em uma imagem.

Onde μ_x e μ_y são as médias e σ_x e σ_y são os desvios padrões;

d) Entropia (Ent) (HARALICK et al. 1973)

$$Ent = - \sum_i \sum_j p(i,j) \log(p(i,j))$$

A entropia representa o grau de dispersão dos níveis de cinza em uma imagem, e semelhante ao SMA pode denotar homogeneidade. Uma maior entropia corresponde a uma menor ordem de organização das estruturas, e também a uma maior probabilidade de ocorrência de um determinado nível de cinza (GUALBERTO e SOUZA FILHO, 2007);

e) Homogeneidade (Hom) (HARALICK et al. 1973)

$$Hom = \sum_i \sum_j \frac{p(i,j)}{1 + (i - j)^2}$$

Também conhecida como Momento Diferencial Inverso, esta media denota a homogeneidade da matriz de co-ocorrência (ALVES e ARAÚJO, 2006);

f) Variância (Var) (HARALICK et al. 1973)

$$Var = \sum_i \sum_j (i - \mu)^2 p(i,j)$$

Esta medida diz respeito à heterogeneidade da matriz de co-ocorrência (ALVES e ARAÚJO, 2006);

g) Dissimilaridade (Dis) (ALVES e ARAÚJO, 2006)

$$Dis = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} |(i - j)| \cdot p(i, j)$$

Esta medida mede o quanto os pares de pixel são diferentes entre si (ALVES e ARAÚJO, 2006);

h) Média (Med) (ALVES e ARAÚJO, 2006)

$$Med = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} i \cdot p(i, j)$$

Esta medida expressa a média da matriz de co-ocorrência. É uma das medidas mais utilizadas para descrever a tendência central dos valores de uma distribuição.

Dentre os autores que utilizaram variáveis relacionadas à textura, com o objetivo de aperfeiçoar a acuracidade da classificação temática pode-se citar FRANKLIN e PEDDLE (1990), que efetuaram o mapeamento da cobertura de uma determinada área utilizando imagem multiespectral SPOT HRV, empregando matrizes de co-ocorrência espacial e medidas estatísticas da textura como a entropia e o momento diferencial inverso. Em outro trabalho, os mesmos autores, Peddle e Franklin (1991) avaliaram a melhoria da acuracidade na classificação de dados da imagem Spot Multiespectral (MLA) e imagens de radar de abertura sintética (SAR), utilizando análise da textura, integrada a dados de geomorfometria e topografia.

Rennó et al. 1996 também tinham como objetivo melhorar a classificação de uma imagem SAR, bandas L e C. Os autores propuseram uma metodologia de seleção de textura que maximizasse a discriminação entre classes de uso na imagem SAR.

3.6.4 Técnicas de Fusão de Imagens

A fusão em dados satelitários é empregada para melhorar a resolução espacial das imagens, facilitando assim a identificação de alvos. Essa técnica de processamento pode ser empregada para melhorar a classificação de imagens, por exemplo em áreas urbanas, uma vez que a heterogeneidade espectral destas áreas em imagens de sensoriamento remoto pode conduzir a erros em sua classificação e que imagens de alta resolução espacial seriam imprescindíveis para

análises intra-urbanas (FUCKNER et al. 2005). Como exemplo de outro uso importante, Candeias e Silva (2004) comentam que o uso integrado de imagens de sensoriamento remoto, com boas resoluções espaciais, as técnicas de geoprocessamento, cartas topográficas convencionais e processamento digital de imagens têm auxiliado, sobremaneira, a atualização cartográfica.

Segundo Wald (1999), a fusão de dados satelitários pode ser caracterizada como uma estrutura formal na qual são definidos conceitos e ferramentas para a união de dados originários de diferentes fontes com o objetivo de obter informações de melhor qualidade.

Existem várias técnicas de fusão disponíveis e que podem ser divididas em três grupos: **a)** as que utilizam um modelo de domínio espacial; **b)** as de domínio espectral e **c)** as que trabalham com operações algébricas (PINHO et al. 2005).

Nos tópicos subsequentes são descritas algumas dessas técnicas de fusão de imagens digitais.

3.6.4.1 Modelos de Domínio Espacial

Esta técnica isola a informação espacial de alta frequência da imagem de alta resolução e a combina com a imagem multiespectral (SCHOWENGERDT, 2007), como por exemplo as transformadas wavelets.

3.6.4.2 Modelos de Domínio Espectral

As técnicas de domínio espectral utilizam transformações nos dados, gerando um novo conjunto de bandas, sendo que uma delas é altamente correlacionada com a imagem pancromática. Exemplos dessas técnicas são Componentes Principais, IHS, HSV, Gram-Schmidt (PINHO et al. 2005).

3.6.4.2.1 HSV

HSV é a sigla para: *hue*, *saturation* e *value*, que significam, respectivamente, matiz, saturação e brilho. É um espaço de cor mais uniforme que o espaço RGB e que separa a

cromaticidade da luminância. Matiz é a cor pura dominante percebida pelo observador. Saturação é o quanto a cor pura dominante (matiz) é diluída pela luz branca. Brilho (ou valor) é a quantidade de luz refletida (claro/escuro), ou a quantidade de luz branca na cor (NUÑEZ, 2006). A transformação RGB-HSV separa a informação acromática (valor) e cromática (matiz e saturação) de uma imagem RGB (SCHNEIDER et al. 2003).

No método de fusão pela transformada HSV, três bandas espectrais de baixa resolução espacial são transformadas do espaço RGB de cores para o espaço HSV. A componente V é substituída pela imagem PAN, e é efetuada a operação inversa, retornando para o espaço RGB (SCHNEIDER et al. 2003).

3.6.4.2.2 *Gram-Schmidt spectral sharpening*

A transformação *Gram-Schmidt* é baseada em rotações e translações do conjunto original de atributos. Ela é executada através da simulação de uma banda de alta resolução, obtida a partir de um conjunto de baixa resolução. Este procedimento é realizado através da obtenção de um grupo de vetores ortogonais independentes, com variância quadrática das imagens espectrais, obtidos entre duas bandas de baixa resolução, gerando uma banda de alta resolução com o menor desvio padrão da média (OLIVEIRA, 2004). Primeiramente, uma banda pancromática é simulada para a baixa resolução espacial das bandas multiespectrais. Em seguida uma transformação *Gram-Schmidt* é executada nas bandas pancromática e multiespectrais, onde a banda pancromática simulada é empregada como a primeira banda. Na seqüência a primeira banda *Gram-Schmidt* é substituída pela banda pancromática de alta resolução espacial. Finalmente, a transformação inversa *Gram-Schmidt* é aplicada para formar a banda espectral fusionada (*pan-sharpened*) (RSI, 2005).

3.6.4.2.3 *Principal components spectral sharpening*

A Análise por Componentes Principais (ACP), também chamada de Transformação por Componentes Principais ou Transformada de Karhunen-Loeve, é uma das funções de processamento de imagens mais poderosas na manipulação de dados multiespectrais. A principal tarefa da ACP é analisar as correlações entre N imagens brutas e produzir um novo conjunto de N

novas imagens com nenhuma correlação entre si. A primeira Componente Principal (PC) irá conter a informação que é comum a todas as N bandas originais. A segunda irá conter a feição espectral mais significativa do conjunto e assim sucessivamente. As últimas PCs irão conter menos informações do que as suas predecessoras, o que não significa que essa informação seja qualitativamente menos importante mas apenas que em termos quantitativos o conteúdo em informação decresce para as últimas CPs. A 1ª PC é uma imagem muito semelhante a uma vista “pancromática” da cena (CROSTA, 1992).

No processo de fusão por esse método, uma transformação de componentes principais é executada nos dados multiespectrais. A banda referente à primeira componente principal é substituída pela banda de alta resolução espacial, cuja resolução é reamostrada para ajustar-se à primeira componente principal, de forma a evitar as distorções na informação espectral. Então uma transformação inversa é executada. O dado multiespectral é automaticamente reamostrado para o tamanho de pixel de alta resolução utilizando técnicas de vizinho mais próximo, bilinear e convolução cúbica (RSI, 2005).

3.6.4.3 Modelos Algébricos

Por último as técnicas que empregam modelos algébricos operam funções aritméticas pixel-a-pixel, como por exemplo a técnica *Brovey* (PINHO et al. 2005).

3.6.4.3.1 Color normalized (Brovey)

O método “*Color Normalized (Brovey)*” é uma técnica de fusão que usa uma combinação matemática de imagens coloridas e dados de alta resolução. Cada banda da imagem colorida é multiplicada pela razão da imagem pancromática dividida pela soma das bandas da imagem colorida. A função automaticamente reamostra as três bandas coloridas para um tamanho de pixel de alta resolução utilizando técnicas de vizinho mais próximo, bilinear ou convolução cúbica. As imagens de saída RGB terão o tamanho do pixel da imagem de alta resolução (RSI, 2005).

3.6.5 Segmentação

A segmentação é a subdivisão de uma imagem em regiões (eCognition©, 2000). Este processo trata de identificar as regiões que compõem uma cena, independentemente de qual seja o significado ou a interpretação para essas regiões (USÓ, 2008). O ato de segmentar uma imagem corresponde à formação de áreas compostas por determinado número de pixels unidos por um critério de similaridade. Os resultados desse processo são áreas com aspecto contínuo, onde cada área possui características espectrais bem diferentes. O processo de segmentação consiste em uma etapa prévia à classificação temática, onde os elementos analisados e utilizados na classificação serão as regiões resultantes da aplicação do segmentador (VENTURIERI e SANTOS, 1998). A técnica tem como objetivo simplificar uma imagem em algo mais significativo e fácil de analisar (USÓ, 2008).

A segmentação subdivide a imagem em suas regiões ou objetos constituintes. O nível em que a subdivisão é executada depende do problema que está sendo resolvido. A segmentação deve parar quando os objetos de interesse forem isolados. Algoritmos de segmentação de imagens geralmente são baseados em duas propriedades básicas de valores de intensidade: descontinuidade e similaridade. A primeira categoria tem como princípio dividir uma imagem baseada em mudanças abruptas de intensidade, tais como bordas (GONZALES e WOODS, 2002). Neste método o algoritmo procura por descontinuidades espectro-texturais na imagem e efetua o cálculo da probabilidade de cada pixel pertencer a uma determinada borda; posteriormente, esses pixels são ligados para a formação de bordas contínuas (ROSOT, 2001). A segunda categoria é baseada em divisões de uma imagem em regiões similares de acordo com um conjunto de critérios pré-definidos (GONZALES e WOODS, 2002).

O algoritmo de segmentação por crescimento de regiões é um procedimento que agrupa pixels ou sub-regiões em grandes regiões baseados em um critério pré-definido. O procedimento se inicia com um conjunto de pontos denominados “sementes” a partir dos quais crescem regiões através da incorporação dos pixels vizinhos que possuem propriedades similares como nível de cinza, textura e cor de cada uma delas (GONZALES e WOODS, 2002).

Existem vários *softwares* de processamento de imagens que executam a segmentação. Entre eles está o *software* brasileiro SPRING, criado pelo Instituto Nacional de Pesquisas Espaciais (INPE). Este *software* utiliza critérios de similaridade e área para a geração das regiões.

O limiar de similaridade é o limiar mínimo, ou seja, “define a diferença mínima entre o valor de um pixel e o valor médio da região contígua a ele, para que esse pixel possa ser agrupado a essa região. Se a diferença entre o valor do pixel e a média da região vizinha for maior que o valor de similaridade definido pelo usuário, o pixel não irá pertencer àquela região. O segundo limiar, de área, significa o menor valor de área (em pixel) para que uma região seja separada de outra” (BARBOSA et al. 2000³ citados por VASCONCELOS e NOVO, 2004)

Vários estudos utilizando imagens satelitárias e segmentação foram realizados no mapeamento da cobertura da Terra. No Brasil, Dlugosz (2005) utilizou a técnica de segmentação e classificação na discriminação de tipologias florestais da Floresta Ombrófila Mista na região de Caçador-SC bem como na descrição do estado atual da floresta.

Rosot (2001) também utilizou a técnica de segmentação em imagens JERS e Landsat-5 TM para a classificação de plantios de *Eucalyptus spp.* e *Pinus spp.* no município de Sengés-PR. Concluiu que a técnica se mostrou promissora, principalmente na imagem JERS.

Em seu estudo, Nascimento (1997) também utilizou técnicas de segmentação e classificação por regiões no mapeamento de unidades da paisagem na Amazônia, utilizando para isto imagens do satélite Landsat-5 TM. O autor testou vários pares de similaridade e área para a segmentação e concluiu que, para a delimitação das classes de interesse (floresta primária, vegetação secundária, pastagem e solo exposto), o melhor par testado foi 12 e 10, respectivamente.

Schwartz e Pedrini (2005) apresentaram um novo método de segmentação baseado na extração de características de texturas e na dependência espacial entre as regiões, utilizando imagens de satélite e *laser scanners*. Com os resultados obtidos os autores concluíram que o uso em conjunto dos dois métodos apresentou resultados mais acurados.

3.6.5.1 Seleção de Bandas para a Segmentação

O *software* SPRING, no processo de segmentação, possui limitação quanto ao número de bandas participantes do processo, exigindo uma seleção prévia de quais bandas participariam da segmentação.

³ BARBOSA, C.; HESS, L.; MELACK, J.; NOVO, E. Mapping amazon wetlands through region growing segmentation and segmented-based classification JERS-1 data. **Anais do Simpósio Latino Americano de Sensoriamento Remoto**. Puerto Iguazu, Argentina. Cd Rom, p. 1065 – 1076, 2000.

Sensores multi/hiperespectrais oferecem uma infinidade de imagens (bandas) em diferentes comprimentos de ondas. Esses dados oferecem várias vantagens, mas esta informação adicional também pode trazer algumas desvantagens, como por exemplo espaço para armazenamento dos dados ou tempo de processamento (USÓ, 2008). Este mesmo autor comenta, ainda, que dados hiperespectrais sofrem importantes redundâncias devido ao fato de que as bandas são altamente correlacionadas. Ele estudou processos de extração e seleção de bandas como, por exemplo, análise de *cluster* e técnicas de transformações dos dados originais como, por exemplo, Análise por Componentes Principais. Esta última fornece uma melhor capacidade de discriminação, mas estas novas bandas podem não ter um significado físico claro. Por isso, muitas vezes é preferível a utilização de técnicas de seleção de bandas, que tem a vantagem de preservar as informações originais relevantes dos dados (USÓ, 2008).

A técnica análise de *cluster*, ou análise de agrupamento, tem como objetivo descobrir os agrupamentos naturais das variáveis (JOHNSON e WICHNER, 1998), dividindo os elementos da amostra em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis e os elementos de grupos diferentes sejam heterogêneos em relação a estas mesmas características (MINGOTI, 2005). Esta técnica não requer suposição sobre o número de grupos ou sua estrutura. O agrupamento é feito a partir de similaridades ou distâncias (JOHNSON e WICHNER, 1998).

Existem várias medidas de similaridade, que produzem agrupamentos diferentes, tais como: distância Euclidiana, distância de Minkowsky (JOHNSON e WICHNER, 1998, MINGOTI, 2005), coeficiente de concordância simples, coeficiente de concordância positiva, coeficiente de concordância de Jaccard e distância Euclidiana média (MINGOTI, 2005), entre outros.

As técnicas de construção de *clusters* se dividem em hierárquicas e não hierárquicas. As primeiras têm o intuito de identificar possíveis agrupamentos e o valor provável do número de grupos; já nas segundas é necessário que o número de grupos esteja pré-estabelecido. Nas técnicas hierárquicas, em cada passo do algoritmo, os elementos amostrais vão sendo agrupados formando novos conglomerados até o momento em que todos os elementos considerados estarão em um único grupo. Em cada fase desse procedimento, os grupos são comparados através de uma medida de similaridade (ou dissimilaridade) previamente definida (MINGOTI, 2005). Existem vários métodos de agrupamentos hierárquicos, entre eles o método de ligação simples, método de

ligação completa, método da média das distâncias e o método de Ward (JOHNSON e WICHNER, 1998, MINGOTI, 2005).

3.6.6 Classificação de imagens digitais

A classificação automática de imagens multiespectrais de sensoriamento remoto diz respeito a associar cada pixel da imagem a um rótulo descrevendo um objeto real (solo, água, vegetação, etc.). Dessa forma os valores numéricos (DNs) associados a cada pixel, definidos pela reflectância dos materiais que compõem esse pixel, são identificados em termos de um tipo de cobertura da superfície terrestre imageada (água, tipo de vegetação, de solo, rocha, etc.) chamados, então, “temas” (CROSTA, 1992). Partindo da hipótese de que pixels cobrindo um mesmo alvo devem ter características espectrais similares, algoritmos computacionais podem ser usados para dividir os pixels de uma imagem em grupos, segundo suas leituras em cada banda espectral (CENTENO, 2003). Dependendo do algoritmo utilizado nesse procedimento, a classificação é dita supervisionada ou não-supervisionada. O procedimento envolve duas fases distintas: o treinamento e classificação propriamente dita (MOREIRA, 2001).

Os classificadores podem ser divididos em dois grupos: os classificadores “pixel a pixel” e os classificadores por “regiões”. Os primeiros utilizam somente as informações espectrais da imagem, classificando cada pixel em uma determinada classe, e podem ser divididos em estatísticos e determinísticos. Já os segundos utilizam tanto informações espectrais como espaciais. A informação de borda é utilizada inicialmente para separar as regiões e as propriedades espaciais e espectrais que irão unir áreas com mesma textura (SPRING, 2008b).

Além disso, alguns algoritmos classificadores mais específicos utilizam variáveis de múltiplas fontes e formatos para construir regras de alocação de cada pixel a uma determinada classe. Um exemplo desses classificadores são as Árvores de Decisão, que utilizam a abordagem “Dividir e Conquistar” para o problema de aprendizagem (WITTEN e FRANK, 2005) e são utilizadas na mineração de dados.

A mineração de dados é definida como um processo de descoberta de padrões, automática ou semi-automaticamente, em grandes quantidades de dados, sendo que estes são armazenados eletronicamente e a pesquisa é automatizada. Esta descoberta de padrões deve ser

significativa e trazer alguma vantagem, geralmente uma vantagem econômica (WITTEN e FRANK, 2005).

Na mineração de dados é possível utilizar técnicas e algoritmos de diferentes áreas do conhecimento, principalmente inteligência artificial (especialmente aprendizagem de máquina), banco de dados (recursos para manipular grandes bases de dados) e estatística (na avaliação e validação dos resultados) (SILVA, 2004). Como o volume dos dados aumenta consideravelmente, a proporção deles que é facilmente compreendida diminui significativamente (WITTEN e FRANK, 2005), tornan-se necessárias, então, ferramentas de mineração de dados para lidar com tais bases (GUIDINI e VASCONCELOS, 2008).

As classificações podem ser divididas em classificações supervisionadas e não supervisionadas. As primeiras necessitam de um treinamento com amostras representativas das classes de interesse. Os classificadores *Bhattacharya* e Árvore de Decisão são exemplos. Na classificação não supervisionada, não é necessário esse treinamento. O algoritmo divide a imagem em classes que são definidas posteriormente. Esses dois métodos são descritos na sequência.

3.6.6.1 *Classificação Não-supervisionada*

A classificação não supervisionada é indicada para casos onde o analista não conhece a área a ser classificada a priori, não sendo este um impedimento para a utilização desta técnica (MOREIRA, 2001). Os procedimentos não-supervisionados são aplicados em duas fases diferentes. A imagem primeiramente é classificada a partir da junção de pixels semelhantes espectralmente, os agrupamentos espectrais naturais ou *clusters*, e posteriormente a esta fase, o analista identifica cada um destes grupos espectrais comparando a imagem classificada com dados de referência da terra (LILLESAND et al. 2007). A classificação não-supervisionada baseia-se no princípio de que o computador é capaz de identificar por si só as diferentes classes. Ele identifica os *clusters* formados e automaticamente eles se tornam amostras de treinamento (CROSTA, 1992). O analista somente define quantas classes deseja obter e os algoritmos se encarregam de classificar. Em muitas situações é melhor considerar um maior número de classes em virtude das eventuais diferenças espectrais dentro delas, sendo mais fácil agrupar as sub-classes posteriormente (CENTENO, 2003).

Exemplos de algoritmos não supervisionados são o “K-médias” e o “Isodata” para classificações por pixel, e o “Isoseg” para classificação por regiões (DLUGOSZ, 2005).

O *software* SPRING (INPE) tem disponível em suas ferramentas o classificador Isoseg. Esse classificador atua sobre regiões que são caracterizadas por seus atributos estatísticos de média, matriz de covariância e área. Ele utiliza uma medida de similaridade (distância de Mahalanobis) entre a classe e as regiões candidatas a pertencer a esta classe. Este classificador utiliza três etapas: 1) Definição de um limiar de aceitação em porcentagem pelo analista. Este limiar define uma distância de Mahalanobis sendo que todas as regiões pertencentes a uma mesma classe possuem distância menor que esta. 2) Detecção de classes: as regiões são ordenadas em valores decrescentes de área e o processo de classificação se inicia. Serão tomados como parâmetros estatísticos de uma classe (média e matriz de covariância), os parâmetros estatísticos da região de maior área que ainda não tenha sido associada a classe alguma. Em seguida toda região que tiver a distância de Mahalanobis inferior ao limiar estabelecido pertencerá a esta classe e assim sucessivamente. 3) Competição entre classes: as regiões são reclassificadas de acordo com as novas estatísticas estabelecidas para cada classe gerada de acordo com o item anterior (SPRING, 2008b).

3.6.6.2 *Classificação Supervisionada*

Nesse tipo de classificação, o analista da imagem “supervisiona” o processo de categorização do pixel, especificando, para o algoritmo do computador, descritores numéricos de vários tipos de cobertura da terra presentes na imagem, com base em amostras representativas de cobertura da terra, chamadas de áreas de treinamento. Estas áreas são usadas para compilar uma “chave de interpretação” numérica que descreve os atributos de cada feição de interesse. Cada pixel no conjunto de dados é então comparado numericamente com cada categoria desta “chave de interpretação” e classificada com o nome da categoria mais semelhante (LILLESAND et al. 2007). Segundo Crosta (1992), é necessário que o analista conheça a área a ser classificada, e o ideal seria que observações de campo coincidisse com a data de passagem do satélite. Centeno (2003) afirma que a fase mais importante da classificação supervisionada é a amostragem, pois as características de cada classe são definidas neste processo e permanecem até o final. Afirma ainda que a amostra deve ter um tamanho tal que represente a classe e que o ideal é escolher

várias áreas pequenas ao invés de apenas uma de maior extensão, para evitar a redundância da alta correlação dos pixels vizinhos. De acordo com Venturieri e Santos (1998) não existe um número certo de amostras a serem coletadas, mas elas devem refletir a variabilidade de cada classe dentro da área de estudo. Pela experiência dos autores na área agrícola e florestal 3N seria um bom número de amostras a serem coletadas para cada classe, onde N indica o número de bandas espectrais da imagem.

Os algoritmos “Máxima verossimilhança” (disponível no SPRING), “Paralelepípedo” e “Distância Mínima”, que operam em nível de pixel, estão entre os mais conhecidos. O *software* SPRING (INPE) disponibiliza também outro algoritmo, este baseado em regiões, o *Bhattacharya*, que mede a separabilidade estatística (distância) entre um par de classes espectrais. O princípio é semelhante ao do classificador Ioseg, porém a distância utilizada neste caso é a “distância de *Bhattacharya*”. Este classificador utiliza as amostras de treinamento para estimar a função densidade de probabilidade para as classes apontadas no treinamento, e avalia, para cada região, a distância de *Bhattacharya* entre as classes (MOREIRA, 2001). O cálculo da distância de *Bhattacharyya*, descrito por este autor é feito por meio da seguinte equação:

$$B_{(p_i, p_j)} = \frac{1}{2} (m_i - m_j)^T \Sigma (m_i - m_j) + \frac{1}{2} \ln \frac{|\Sigma(m_i - m_j)|}{|\Sigma_i|^{\frac{1}{2}} |\Sigma_j|^{\frac{1}{2}}}$$

em que:

m_i é a média da classe i;

m_j é a média da classe j;

T é a matriz transposta;

ln é o logaritmo neperiano;

i e j são as classes dentro do contexto.

A classe que apresentar a menor distância é atribuída à região avaliada.

Além desses dois classificadores por regiões, existe um terceiro, também disponível no SPRING, o ClaTex. Este algoritmo também é supervisionado e utiliza atributos texturais das regiões de uma imagem segmentada para efetuar a classificação por regiões.

3.7 ÁRVORE DE DECISÃO

Os algoritmos por árvore de decisão surgiram na década de 70 com o ID3 (*Iterative Dichotomiser*) criado por J. Ross Quinlan. Posteriormente, Quinlan apresentou o C4.5 (sucessor do ID3) que se tornou uma referência com a qual os algoritmos supervisionados mais recentes são frequentemente comparados. Em 1984, um grupo de estatísticos (L. Breiman, J. Friedman, R. Olshen, e C. Stone) publicaram o livro *Classification and Regression Trees* (CART), onde descreveram a criação de árvores de decisão binárias (HAN e KAMBER, 2006).

A Árvore de Decisão é um procedimento de classificação que particiona um conjunto de dados em subdivisões menores com base em um conjunto de testes definidos em cada ramo (ou nó) da árvore (FRIEDL e BRODLEY, 1997; PAL e MATHER, 2003). Assim, uma decisão complexa é dividida em diversas decisões mais simples, que podem conduzir a uma solução que seja mais fácil de interpretar (XU et al. 2005). As árvores de decisão são estruturas semelhantes a diagramas de fluxo em forma de árvore, onde cada nó representa uma decisão sobre um atributo, e cada folha representa uma classe de distribuição (HAN e KAMBER, 2006).

A árvore é composta por um nó raiz (formado por todos os dados), um conjunto de nós internos (divisões) e um conjunto de nós terminais (folhas) (XU et al. 2005). Cada nó em uma árvore de decisão tem somente um nó pai e dois ou mais nós descendentes, (FRIEDL e BRODLEY, 1997). Os nós em uma árvore de decisão envolvem testar um atributo particular. Geralmente este teste compara um valor de atributo com uma constante, podendo, entretanto existir algumas árvores que comparam dois atributos entre si ou utilizam alguma função com um ou mais atributos. Se um atributo testado é nominal, o número de nós filhos é geralmente o número dos valores possíveis desse atributo. Se for numérico, o teste em um nó geralmente determina se o seu valor é maior ou menor do que uma constante pré-determinada, gerando uma divisão em dois sentidos (WITTEN e FRANK, 2005).

As árvores de decisão utilizam uma abordagem ambiciosa e o princípio de “dividir para conquistar” (HAN e KAMBER, 2006). Sua estruturação é geralmente realizada movendo-se a árvore para baixo até que o nó folha seja encontrado. Esta abordagem é conhecida como *top-down* (XU et al. 2005).

Na escolha de um critério de divisão que melhor separa os dados em classes individuais é utilizada uma “medida de seleção de atributo”, que tem como princípio obter classes o mais

puras possível. Estas medidas são também conhecidas como “regras de divisão” porque determinam como os dados em um determinado nó serão particionados. Elas fornecem um *ranking* para cada atributo descrevendo os dados de treinamento. O atributo que tiver a melhor pontuação para a medição é escolhido como o “atributo de divisão” e será o nó raiz da árvore. Este nó será classificado então com o critério de divisão e os “galhos” da árvore serão produzidos para cada resultado do critério, dividindo os dados. Algumas medidas de seleção de atributos conhecidas são *information gain*, *gain ratio* e índice de Gini (HAN e KAMBER, 2006).

A medida *informatio gain* é baseada na teoria da informação de Claude Shannon na qual estuda o valor ou o conteúdo de informação das mensagens. O atributo com o maior ganho de informação é escolhido como o atributo de divisão do nó raiz. Este método minimiza a informação necessária para classificar os dados resultando nas divisões dos mesmos e reflete e “impureza” dessas divisões (HAN e KAMBER, 2006).

O *gain ratio* é a medida utilizada no algoritmo C4.5. Ele aplica um tipo de normalização no *information gain* usando um valor de “divisão da informação”. Este valor representa a informação potencial gerada a partir da divisão dos dados de treinamento. O atributo com o maior *gain ratio* é selecionado como atributo de divisão (HAN e KAMBER, 2006).

E, finalmente, o índice de Gini também pode ser utilizado como critério de divisão. É definido como uma medida da impureza dos dados. Este índice considera uma divisão binária para cada atributo. O atributo que maximiza a redução da impureza (ou, de forma equivalente, tem o mínimo Índice Gini) é selecionado como a divisão desse atributo (HAN e KAMBER, 2006).

A árvore de decisão não assume nenhum modelo estatístico *a priori*, sendo a divisão do espaço de atributos feita de acordo com as amostras provenientes do treinamento (SHIBA et al. 2005).

O método de classificação por árvore de decisão possui as mesmas vantagens de classificadores como redes neurais e lógica *fuzzy*, no entanto é mais fácil de usar e entender. Por possuírem uma estrutura hierárquica, fornecem um método mais simples, porém mais robusto de interpretar, testar e analisar os resultados (FRIEDL e BRODLEY, 1997). Algumas vantagens do métodos estão citadas a seguir

- a) As árvores de decisão são não-paramétricas;
- b) Não requerem suposições de distribuição dos dados de entrada;

- c) Tratam de relações não-lineares entre feições e classes, permitindo valores ausentes;
- d) São capazes de manusear tanto entradas numéricas quanto categóricas em um modelo;
- e) Possui um apelo intuitivo significativo porque a estrutura de classificação é explícita e, por isso, facilmente interpretável.

Para se construir uma árvore de decisão precisa, é necessário obter amostras representativas das classes desejadas (FRIEDL e BRODLEY, 1997).

Após a construção de uma árvore de decisão é importante avaliá-la, o que pode ser realizado através da utilização de dados que não tenham sido usados no treinamento. Esta estratégia permite estimar como a árvore generaliza os dados e se adapta a novas situações, podendo, também, se estimar a proporção de erros e acertos ocorridos na construção da árvore (BRAZDIL, 1999⁴ citado por SHIBA et al. 2005).

Existem três tipos de árvore de decisão: árvores homogêneas, heterogêneas e híbridas. As árvores homogêneas utilizam somente um algoritmo para cada divisão e se dividem em univariadas e multivariadas. As heterogêneas utilizam dois ou mais algoritmos e as árvores híbridas possuem uma combinação de diferentes hipóteses homogêneas (FRIEDL e BRODLEY, 1997).

3.7.1 Árvores de Decisão Univariadas (ADU) e Multivariadas (ADM)

As ADU são árvores na qual o limite de decisão é definido a partir de uma única característica dos dados de entrada (SWAIN e HAUSKA, 1969⁵ citado por FRIEDL e BRODLEY, 1997). Cada nó interno é dividido em dois ou mais subconjuntos com base em um teste de uma única característica dos dados de entrada. Os valores específicos dos limites de decisão são estimados empiricamente a partir dos dados de entrada (PAL e MATHER, 2003). As árvores de decisão multivariadas (ADM) são similares às univariadas, mas os testes de divisão podem ser baseados em mais de uma característica dos dados de entrada. Devido ao aumento na

⁴ BRAZDIL, P. **Construção de Modelos de Decisão a partir de Dados**. 1999. Disponível em: <<http://www.ncc.up.pt/~pbrazdil/Ensino/ML/DecTrees.html>> . Acesso em: abril de 2008.

⁵ SWAIN, P.H., HAUSKA, H. The decision tree classifier: design and potential. **IEEE Trans. Geosci. Remote Sensing**, v. 15, p. 142-147, 1969.

complexidade das regras de divisão dos nós internos, ADM são mais difíceis de interpretar do que as ADU, porém, frequentemente são mais compactas (FRIEDL e BRODLEY, 1997).

3.7.2 Algoritmo de Poda

Os algoritmos de poda são geralmente usados para eliminar os nós folhas que foram ajustados por erros nos dados de treinamento (FRIEDL e BRODLEY, 1997). Árvores podadas tendem a ser menores e menos complexas, e assim, mais fáceis de compreender. São geralmente mais rápidas e melhores na correta classificação dos dados de treinamento do que as árvores não podadas (HAN e KAMBER, 2006). A simplificação da árvore envolve a remoção de partes que não contribuem na acurácia da classificação (PAL e MATHER, 2003). Existem dois caminhos para efetuar a poda da árvore, a pré-poda e a pós-poda.

Na primeira opção a árvore é podada parando sua construção previamente, ou seja, não dividindo mais o conjunto de dados (HAN e KAMBER, 2006; WITTEN e FRANK, 2005). Esta parece ser uma perspectiva atrativa, pois evitaria que todo o trabalho de desenvolvimento da sub-árvore fosse descartado (WITTEN e FRANK, 2005), além de não desperdiçar tempo na montagem de uma estrutura que não será usada posteriormente. Este método procura o melhor caminho para dividir o conjunto de dados com base em critérios como *Information gain*, redução do erro (PAL e MATHER, 2003) índice de Gini entre outros. Se o particionamento dos dados em um nó resultar em uma divisão que cai abaixo de um limiar pré-especificado o processo de divisão é interrompido. Existem dificuldades, entretanto, na escolha de um limiar apropriado, uma vez que altos limiares resultam em uma excessiva simplificação da árvore enquanto que baixos limiares resultam em uma baixa simplificação (HAN e KAMBER, 2006).

A segunda opção é a abordagem mais comum, a pós-poda, que remove as sub-árvores a partir de uma árvore completamente crescida (HAN e KAMBER, 2006). Os galhos da sub-árvore selecionada são substituídos por uma folha que é nomeada de acordo com a classe mais freqüente da sub-árvore substituída (WITTEN e FRANK, 2005). O tempo requerido para realizar esta operação é maior, mas este custo é compensado com os benefícios resultantes de uma maior exploração através de possíveis partições (PAL e MATHER, 2003).

Para a realização da poda o algoritmo C4.5 utiliza um método chamado “poda pessimista” e utiliza estimativas da taxa de erro para tomar decisões relativas sobre a poda da

sub-árvore. O princípio deste método é comparar a taxa de erro calculada para a árvore podada e para a não podada. De acordo com o resultado, o que possuir menor erro é selecionado. Este método não requer o uso de um conjunto de ações para a poda, ele utiliza os dados de treinamento para estimar a taxa de erro. (HAN e KAMBER, 2006). Os autores ainda comentam que a estimativa da precisão ou erro baseada no conjunto de treinamento é excessivamente otimista, portanto distorcida. O método de poda pessimista entretanto ajusta a taxa de erro obtida pelo conjunto de treinamento adicionando uma penalidade, de forma a contrariar a tendência.

3.7.3 Uso de Dados de Sensores Remotos em Árvore de Decisão

A árvore de decisão pode ser utilizada em diversas áreas do conhecimento quando o objetivo é classificar dados. Na sequência são apresentados alguns trabalhos que utilizaram esta técnica em dados de sensores remotos.

Friedl e Brodley (1997) utilizaram três tipos de algoritmos de classificação por árvore de decisão e os avaliaram em três tipos diferentes de dados de sensoriamento remoto, para mapeamento da cobertura da terra. Os dados utilizados foram: uma série temporal mensal de imagens NDVI (*Normalized Difference Vegetation Index*) do satélite AVHRR de 1987, uma série temporal de imagens NDVI do mesmo satélite para o ano de 1990 e uma imagem *Landsat Thematic Mapper* (TM), com resolução de 30m. Os resultados dessa análise mostraram que os algoritmos árvore de decisão obtiveram um melhor desempenho em relação aos algoritmos máxima verossimilhança e função discriminante linear com relação à acuracidade.

Para a predição do risco de infestação por esquistossomose em Minas Gerais, Martins et al. (2007) também trabalharam com árvore de decisão. Utilizaram 22 variáveis de sensoriamento remoto derivadas do sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) em duas épocas, uma no verão e uma no inverno, listadas a seguir: bandas azul, vermelho, infravermelho próximo e médio, índice de vegetação melhorado (EVI), NDVI, índices derivados do modelo espectral de mistura (vegetação, solo e sombra), variáveis derivadas do SRTM (*Shuttle Radar Topography Mission*) como o DEM e a declividade e outras duas com informações hidrográficas, a média de acumulação hídrica (AH1) e a mediana de acumulação hídrica (AH2), obtidas através de um mapa de acumulação hídrica. Além dessas variáveis de sensoriamento remoto foram utilizadas mais seis variáveis climáticas e 16 sócio-econômicas, totalizando 44 variáveis. Para o processamento dos dados os autores utilizaram o *software* WEKA. Como resultado dessa

classificação, o coeficiente Kappa apresentou um valor de 51% observando que a maior confusão foi entre as classes média e baixa, alta e média e muito alta e alta.

Figueiredo e Carvalho (2006) também avaliaram o uso do algoritmo árvore de decisão para o mapeamento da cobertura da terra em Capixaba, Acre. Foram utilizadas 11 variáveis de sensoriamento remoto obtidas a partir de dados do satélite Landsat 5 sensor TM. As variáveis foram as seguintes: bandas 1, 2, 3, 4, 5 e 7, imagens fração solo, vegetação, sombra, erro e NDVI. As imagens foram corrigidas geometricamente e convertidas radiometricamente para reflectância. Como resultado, o coeficiente Kappa apresentou um valor de 0,90, considerado excelente. Em comparação com outros métodos de classificação, o de máxima verossimilhança apresentou valor semelhante (0,89). Já pelo método de classificação não supervisionada Isodata o valor foi inferior (0,79). Os maiores erros de classificação pela árvore de decisão foram nas classes pasto alto, pasto baixo e capoeira.

Carvalho et al. (2008) mapearam a Floresta Atlântica da Serra dos Órgãos, RJ com o algoritmo árvore de decisão para que pudessem integrar de dados de sensoriamento remoto e modelo digital do terreno (MDT). Utilizaram para esse estudo imagens do satélite ASTER e fotografias aéreas da região. A imagem ASTER foi primeiramente classificada e aprimorada com a ajuda da fotografia e posteriormente essa classificação foi usada como informação junto com o MDT para a árvore de decisão. As classes delimitadas foram: Agropecuária, Campos de Altitude, Floresta Ombrófila Densa Alto-Montana, Floresta Ombrófila Densa Montana, Floresta Ombrófila Densa Sub-Montana, Floresta Secundária e áreas sem vegetação. A imagem distinguiu ambientes florestais e não florestais e o MDT distinguiu áreas de campos. Os autores afirmaram que a árvore de decisão foi extremamente eficiente para a classificação de dados oriundos de diferentes sensores.

Outro trabalho que também utilizou a árvore de decisão para a classificação de imagens de sensoriamento foi realizado por Shiba et al. (2005). Os autores tinham como objetivo testar essa técnica em comparação com a Máxima Verossimilhança (MaxVer). Os resultados mostraram que a árvore de decisão teve um tempo de processamento menor (30s) em comparação com MaxVer (150s). A diferença entre os classificadores foi de 2,21% ou 9841 pixels.

Zagaglia et al. (2007) fizeram o mapeamento de manguezais utilizando árvore de decisão como método de classificação. Como variáveis utilizaram as seguintes imagens derivadas do satélite Landsat ETM+: radiância, reflectância, componentes principais da reflectância,

componentes principais da radiância, índice de vegetação ajustado ao solo (SAVI) e frações solo, vegetação e sombra do modelo de mistura para a imagem reflectância. As classes definidas foram mangue, apicum e outros. Como resultado dessa classificação o índice de exatidão global e coeficiente Kappa foram de 84,88% e 0,70 respectivamente.

Dutra et al. (2007) tinham como objetivo definir um modelo de estratos ambientais visando à recuperação da vegetação ciliar na área marginal do reservatório da Usina Hidrelétrica do Funil baseado nas características ambientais locais. Utilizaram como variáveis mapas de uso e ocupação do solo, mapa de declividade e mapa de classe de solos. Para a modelagem dos estratos de recuperação foi utilizado o algoritmo árvore de decisão. O coeficiente Kappa obtido foi de 0,94, mostrando a importância dos três atributos para a modelagem dos estratos.

Silva et al. (2008) procuraram identificar fatores ambientais que contribuíssem significativamente para a ocorrência de incêndios em áreas de risco. Além de uma imagem proveniente do satélite Landsat ETM+ os autores utilizaram dados de focos diários de calor, teores de umidade do material combustível e dados altimétricos. A árvore de decisão serviu para identificar o grau de importância de cada fator ambiental. Empregaram-se como indicadores: número e focos de calor, vias de acesso e relevo e umidade do material combustível. Os autores concluíram que o emprego da árvore de decisão na hierarquização da importância dos fatores ambientais constituiu um avanço em relação a procedimentos empíricos adotados na avaliação de áreas sob risco de incêndios.

Sampaio (2007) utilizou dados temporais de imagens MODIS e índices de vegetação NDVI e EVI para classificar a cobertura vegetal do Parque Nacional da Chapada dos Veadeiros, testando os algoritmos ISODATA e árvore de decisão. Com os resultados alcançados a autora conseguiu separar as fitofisionomias mais contrastantes, cerrado denso e campo.

As Árvores de Decisão tem mostrado fornecer melhores acurácias em relação a outros classificadores mais tradicionais. Apesar disso o uso desta técnica com dados de alta resolução espacial, como Landsat por exemplo, não tem sido explorado completamente (COLSTOUN et al. 2003).

3.8 ACURACIDADE DOS ALGORITMOS CLASSIFICADORES

Historicamente, a habilidade em produzir classificações digitais da cobertura da Terra excedeu a capacidade de quantificar significativamente sua acurácia. Este problema algumas vezes impossibilita a aplicação de técnicas automáticas de classificação da cobertura da terra, mesmo quando seus custos são mais favoráveis comparados com outras técnicas mais tradicionais (LILLESAND et al. 2007). Segundo estes mesmos autores, uma classificação não é completa até que sua acurácia seja avaliada.

Recomenda-se adquirir amostras significativas de alvos distintos por toda a imagem para diminuir a confusão entre as classes, além de avaliar a matriz de erros (SPRING, 2008a).

A matriz de erros (também chamada de matriz de confusão ou tabela de contingência) compara, classe por classe, a relação dos dados de referência com os resultados correspondentes da classificação automática (LILLESAND et al. 2007). É uma matriz quadrada, de dimensão $C \times C$, onde C representa o número de classes. Nas colunas está representado o número de pontos amostrados da amostra de referência e nas linhas está o número de pontos encontrados em cada classe pela classificação automática. A diagonal principal mostra o número de pixels classificados corretamente (ROSOT, 2001).

Existem dois tipos principais de erros encontrados em uma classificação: os erros de omissão e os erros de comissão. Os erros de omissão são pixels de uma determinada classe que foram classificados como sendo de outra, ou seja, foram omitidos da classe. Já os erros de comissão correspondem aos pixels que, sendo na realidade de outra classe, foram incluídos na classe considerada (CENTENO, 2003).

Outra medida descritiva que pode ser obtida a partir da matriz de erros é a acurácia do produtor e do usuário. Lillesand et al. (2007) descrevem essas medidas do seguinte modo:

- Acurácia do produtor: está relacionada aos erros de omissão. É a razão entre pixels classificados corretamente em cada classe (valores na diagonal principal) e o número real de pixels naquela classe (valores na coluna).
- Acurácia do usuário: está relacionada aos erros de comissão. É a razão entre os pixels classificados corretamente em cada classe (valores na diagonal principal) e o número total de pixels classificados naquela classe (os valores da linha).

A acurácia geral é calculada dividindo-se o número de pixels classificados corretamente (a soma dos valores da diagonal principal) pelo número total de pixels amostrados (LILLESAND et al. 2007).

Uma estatística muito utilizada também para avaliar a acurácia da classificação é o coeficiente Kappa. Este coeficiente mede a diferença entre o mapa temático produzido e um mapa temático completamente aleatório (CENTENO, 2003) e pode ser definido como (LILLESAND et al. 2007):

$$\hat{k} = \frac{\text{acurácia observada} - \text{acurácia esperada}}{1 - \text{acurácia esperada}}$$

Esta estatística serve como um indicador no qual a porcentagem dos valores corretos de uma matriz de erros é devida a concordâncias “reais” versus concordâncias “ao acaso” (LILLESAND et al. 2007). O coeficiente Kappa pode ser calculado da seguinte maneira:

$$\hat{k} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}$$

Onde:

r = número de linhas da matriz de erros

x_{ii} = número de observações na linha i e na coluna i (na diagonal principal)

x_{i+} = número total de observações na linha i

x_{+i} = número total de observações na coluna i

N = número total de observações incluídas na matriz

O valor do coeficiente Kappa pode variar de 0 a 1. O número 1 indica que a imagem classificada está em completa concordância com o mapa de referência, e o número zero, o inverso. De acordo com Landis e Koch (1977)⁶ citado por Figueiredo e Carvalho (2006), o valor do Kappa pode ser utilizado para avaliar a imagem classificada (**Tabela 2**):

⁶ LANDIS, J.; KOCH, G. G. The measurements of agreement for categorical data. **Biometrics**, Washington, v. 33, n. 3, p. 159-179, Mar. 1977.

TABELA 2: Valores do coeficiente Kappa

Valor do Kappa	Qualidade da classificação
<0,0	Péssima
0,0 - 0,2	Ruim
0,2 - 0,4	Razoável
0,4 - 0,6	Boa
0,6 - 0,8	Muito boa
0,8 - 1,0	Excelente

Fonte: Landis e Koch (1977)⁶, adaptada por Figueiredo e Carvalho (2006)

3.9 O SOFTWARE WEKA

O *software* WEKA é uma coleção de algoritmos de aprendizagem de máquina e de ferramentas de pré-processamento de dados. Foi desenvolvido pela Universidade de Waikato na Nova Zelândia e sua sigla em inglês significa *Waikato Environment for Knowledge Analysis*. O sistema está escrito em Java e é distribuído pela *GNU General Public License*, estando disponível no site: <http://www.cs.waikato.ac.nz/ml/weka>.

O *software* funciona em quase todas as plataformas, como os sistemas operacionais Linux, Windows e Macintosh. Fornece uma interface uniforme a muitos algoritmos diferentes de aprendizagem, juntamente com pré e pós-processamento, além da avaliação dos resultados desta aprendizagem em qualquer conjunto de dados.

O WEKA tem implementado algoritmos de regressão, classificação, *clustering*, *association rule mining* (mineração de regras de associação) e seleção de atributos. Todos os algoritmos utilizados nas análises requerem que os dados estejam em formato ARFF (*Attribute Relation Format File*), que é um arquivo ASCII e serve para definir atributos e seus valores. A **Tabela 3** ilustra a estrutura que um arquivo deve ter:

TABELA 3: Estrutura de arquivos no formato ARFF

@Relation	É o nome da relação, indica o assunto dos dados
@Attribute	São os atributos (variáveis) que esta relação pode ter, bem como os possíveis valores que podem ser aplicados em cada atributo (numéricos ou nominais)
@Data	São os dados propriamente ditos. Estão escritos em linha, separados por vírgula e seguindo a sequência dos atributos.

Fonte: modificado de Lobo e Kalil (2006).

Um exemplo de arquivo ARFF pode ser encontrado no Apêndice 1.

O *software* WEKA oferece ao usuário três interfaces gráficas e uma por comandos de programação, sinteticamente descritas a seguir com base em Witten e Frank, (2005):

a) *Explorer*: esta é a interface gráfica mais amigável para o usuário, que a partir de comandos por menus de seleção, permite realizar de forma rápida as etapas de pré-processamento do conjunto de dados a serem analisados (filtros e conversões), sua classificação (aprendizagem por árvore de decisão), análise de agrupamentos (*cluster*), associação (regras de associação e avaliação dos dados), seleção de atributos e por fim, a visualização dos resultados a partir de gráficos bi-dimensionais;

b) *Knowledge flow*: É uma interface gráfica alternativa que permite diretamente, a partir do lay-out do fluxo de dados em processamento, o planejamento e a execução das configurações desejadas simultaneamente, como por exemplo, a origem dos dados, ferramentas de pré-processamento, algoritmos de aprendizado e métodos de avaliação;

c) *Experimenter*: esta terceira interface do WEKA constitui um ambiente de experimentação prática que automatiza os processos de classificação e filtragem dos dados para diferentes parâmetros de forma a avaliar o desempenho de diferentes algoritmos de aprendizado;

d) *The command-line interface (CLI)*: Esta última não possui uma interface gráfica, seu acesso ocorre por meio das linhas de programação, permitindo acessar profundamente todas as características e comandos do sistema, restrita para usuários avançados que dominam linguagens de programação.

4. MATERIAIS E MÉTODOS

4.1 ÁREA DE ESTUDO

A área de estudo é uma reserva florestal localizada no Município de Caçador, que está inserido no Planalto Ocidental Catarinense entre os paralelos 26° e 27° Sul e os meridianos 50° e 52° Oeste (THOMÉ, 1978). Segundo a classificação de Köppen, a área de estudo possui clima tipo Cfb, isto é, temperado úmido com geadas severas (com maior intensidade entre os meses de maio a setembro), predominando os ventos de direção norte, secundados pelos de direção nordeste (CALDATO *et al.* 1999). A temperatura média anual é de 16,6 °C, com máxima e mínima média de 22,5 e 11 °C, respectivamente, e máxima e mínima absoluta de 38 e -14°C, respectivamente. A precipitação média anual é de 1.613 mm (DLUGOSZ *et al.* 2005).

A Reserva localiza-se entre as coordenadas geográficas 26°50'32,69'' e 26°52'36,73'' de latitude sul e 50°54'51,69'' e 51°58'40,36'' de longitude oeste e compreende uma área de 1194,48 ha (**Figura 12**). A cobertura vegetal existente é constituída por um dos últimos remanescentes de grande dimensão de Floresta Ombrófila Mista, onde se destacam algumas das espécies que constam na lista oficial de espécies ameaçadas de extinção, como por exemplo a *Araucaria angustifolia* (araucária), *Ocotea porosa* (imbuia) e *Ocotea catharinensis* (canela-preta), todas na categoria vulnerável (IBAMA, 1992). Segundo Rivera (2007) as espécies que obtiveram o maior Valor de Importância Ampliado foram *Araucaria angustifolia* e *Cupania vernalis* devido à alta dominância relativa e a alta densidade relativa obtidas por essas espécies. Segundo este mesmo autor, entre as 10 espécies com maiores posições sociológicas relativas, a *Araucaria angustifolia* teve a pior distribuição vertical, pois o número de árvores no estrato médio e inferior foi muito menor do que o número de árvores no seu estrato superior (oito vezes menor). Por outro lado a espécie com melhor posição sociológica relativa foi *Cupania vernalis*, que apresentou uma distribuição vertical regular.

A região na qual está inserida a Reserva Florestal Embrapa/Epagri pertence à bacia hidrográfica do Rio Uruguai, sub-bacia do Rio do Peixe, estando posicionada à esquerda de sua margem, entre os afluentes Rio Castelhana (ao norte) e Rio do Veado (ao sul). Em setembro de 2003 foi realizado um levantamento pedológico semidetalhado, em escala 1:25.000, específico para a área da reserva, cujos resultados revelaram a presença de nove classes de solo. A classe

predominante foi a do tipo “Associação de Cambissolo Háplico Tb Distrófico Léptico + Neossolo Litólico Distrófico típico, A moderado” em 466,85 ha, correspondendo a 39,08% da área de estudo (DLUGOSZ, 2005).

A Reserva Florestal Embrapa/Epagri (RFEE) está em mãos governamentais há aproximadamente 61 anos, quando, através do decreto nº 25.407, de 30 de agosto de 1948, foi declarada área de utilização pública pela presença de 10.000 pinheiros e 1500 imbuias. Pertence formalmente à Embrapa, mas está sob contrato de comodato junto à Epagri (Empresa Catarinense de Pesquisa Agropecuária). Inicialmente foi denominada “Parque Florestal do Contestado”, sendo inaugurada pela Epagri em 1997 como Reserva Florestal Embrapa-Epagri – Caçador (OLIVEIRA et al. 2006).

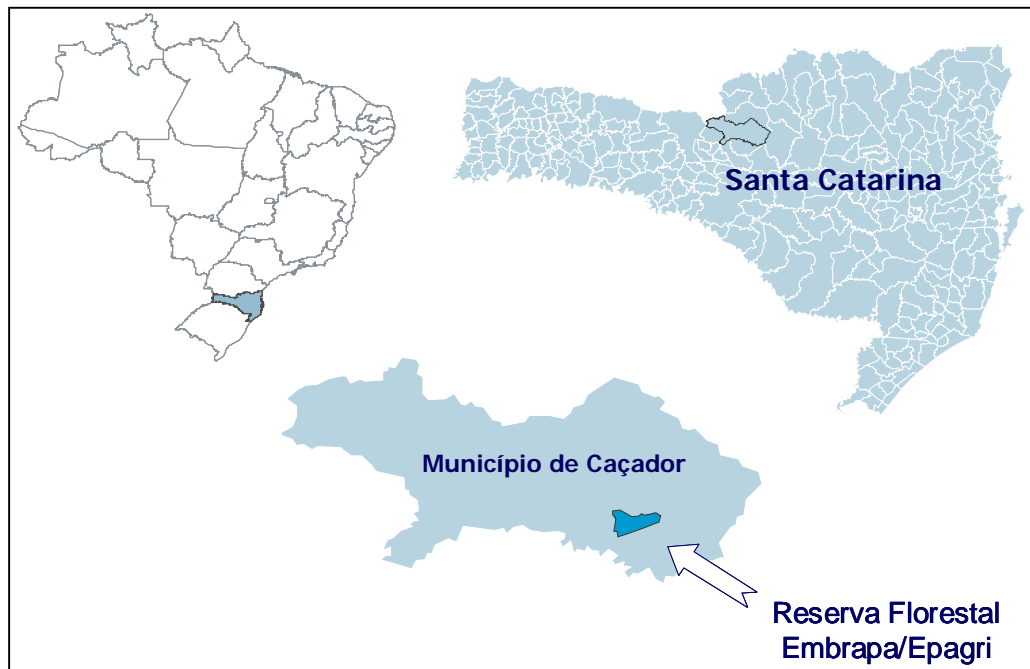


FIGURA 12: Localização da área de estudo

4.2 BASE CARTOGRÁFICA E MAPAS DA RFEE

A base cartográfica que a reserva possui foi confeccionada a partir de uma carta topográfica de Caçador (folha SG-22-Z-A-IV-3) adquirida da Diretoria de Serviços Geográficos (DSG), de onde foram vetorizadas a hidrografia, rede viária, pontos cotados e curvas de nível. A partir destes dois últimos produtos foi possível gerar o Modelo Digital de Elevação (MDE) e as

camadas derivadas deste modelo, como a altimetria, declividade e exposição (ROSOT et al. 2006). Fotografias aéreas tiradas no ano de 1979 também foram utilizadas em trabalhos anteriores (KURASZ, 2005; DLUGOSZ, 2005).

A reserva possui um mapa temático de uso do solo desenvolvido por Dlugosz (2005) através de interpretação visual de uma imagem Ikonos II de fevereiro de 2004. Para a definição das classes tipológicas de vegetação o autor utilizou dentre as quatro bandas espectrais presentes no satélite, a composição colorida 4NIR3R2G, além de parâmetros como cor, tonalidade e textura. Os resultados deste trabalho estão mostrados na **Figura 13**.

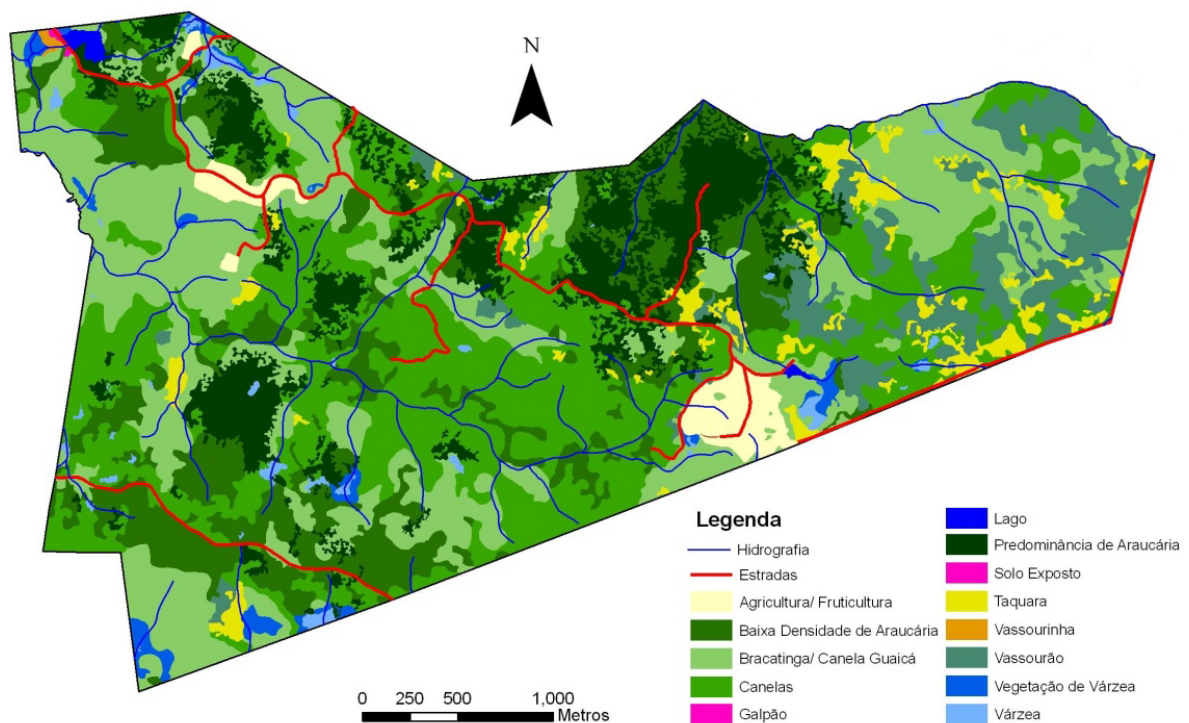


FIGURA 13: Tipologias definidas para a RFE
Fonte: Dlugosz (2005).

4.3 IMAGENS E SOFTWARE UTILIZADOS

As imagens utilizadas nesse estudo são provenientes do satélite ALOS e foram adquiridas através do site do IBGE. Uma delas é multiespectral possuindo quatro bandas que abrangem o visível e o infra-vermelho, com resolução espacial de 10 metros. Essa imagem é do

dia 25 de setembro de 2006. A segunda imagem é pancromática, com resolução espacial de 2,5 metros e é do dia 10 de novembro de 2006.

Para o processamento dos dados foram utilizados os *softwares* ENVI© (RSI) versão 4.3, ArcMap™ (ESRI®) versão 9.1, SPRING (INPE) versão 5.0.1, WEKA (*University of Waikato, New Zealand*) versão 3.4.13.

4.4 GEORREFERENCIAMENTO

As imagens ALOS chegaram previamente corrigidas geometricamente de acordo com o nível de processamento solicitado por ocasião da aquisição. Para aumentar a precisão, elas passaram por um processo de refinamento do georreferenciamento. A imagem ALOS multiespectral (10m) foi georreferenciada a partir de uma outra imagem corrigida do satélite Ikonos da área de estudo, que possui resolução espacial de 4m e que já havia sido georreferenciada por Dlugosz (2005). A técnica utilizada foi *master-slave* (mestre-escravo). Foram coletados oito pontos de controle bem distribuídos nas duas imagens. A equação de transformação utilizada foi o “polinômio do 1º grau” e o método de reamostragem foi o interpolador “vizinho mais próximo”.

O polinômio do 1º grau foi escolhido como equação de transformação porque a imagem já estava previamente corrigida. Neste caso, poucos pontos de controle são necessários para refinar o processo. Uma equação do 1º grau deve ter no mínimo 3 pontos de controle para ser calculada, no entanto, como as coordenadas medidas dos pontos de controle estão sujeitas a erros recomenda-se usar número de pontos maior que o mínimo. Aconselha-se o uso de 6 pontos de controle para o polinômio de 1º grau, 10 pontos de controle para o polinômio de 2º grau e 14 pontos para o polinômio de 3º grau (SPRING, 2008c). No método de interpolação por vizinho mais próximo o pixel da imagem corrigida terá o valor do pixel que se encontra mais próximo. Este método foi escolhido por ser o que mais preserva o valor original do número digital (DN), uma vez que não envolve nenhuma média. Esta é uma propriedade importante para dados que serão submetidos a um processo de classificação automática (CROSTA, 1992).

No georreferenciamento da imagem pancromática foram seguidos os mesmos procedimentos. A técnica utilizada também foi *master-slave*, mas a imagem base foi a ALOS multiespectral, que previamente teve seu pixel reamostrado para 2,5m (resolução da

pancromática). Foram coletados nove pontos de controle bem distribuídos nas imagens. A equação de transformação utilizada também foi o “polinômio do 1º grau” e o método de reamostragem foi o interpolador “vizinho mais próximo”.

4.5 FUSÃO DE IMAGENS

As bandas multiespectrais e a pancromática da imagem ALOS (**Tabela 4**) foram fusionadas com o objetivo de melhorar a resolução espacial das imagens. Para esse estudo foram testadas quatro técnicas de fusão disponíveis no *software* ENVI: HSV (três bandas), *Color Normalized (Brovey)*(três bandas), *Gram-Schmidt Spectral Sharpening* (com três e quatro bandas) e Componentes Principais (com três e quatro bandas). As três bandas selecionadas para participarem da fusão foram as bandas 2, 3 e 4, correspondente ao verde, vermelho e infravermelho próximo respectivamente. Essas bandas foram escolhidas pois representam melhor a vegetação, que é o objeto de estudo do trabalho.

TABELA 4: Bandas espectrais do satélite ALOS e suas respectivas resoluções espaciais

Satélite ALOS		
	Bandas espectrais (μm)	Resolução espacial (m)
Banda 1	0,42 - 0,50	10
Banda 2	0,52 - 0,60	10
Banda 3	0,61 - 0,69	10
Banda 4	0,76 - 0,89	10
Pan	0,52 - 0,77	2,5

Fonte: Adaptado de JAXA (2007)

A **Figura 14** mostra as imagens ALOS adquiridas para a RFEE, a multiespectral (10m) com composição colorida falsa cor (432) e a imagem pancromática (2,5m).

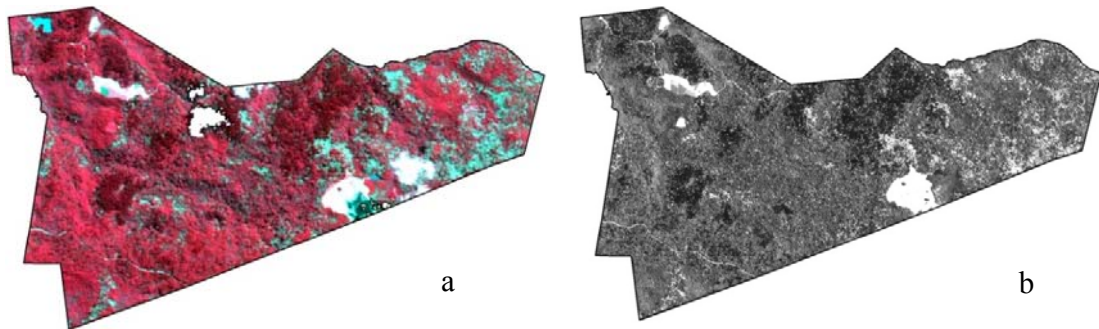


FIGURA 14: (a) Imagem ALOS multiespectral com resolução de 10 metros;
(b) Imagem ALOS pancromática com resolução de 2,5 metros.

Para analisar o resultado das fusões foram realizados três testes, sendo o primeiro uma análise visual entre o resultado das técnicas de fusões e a imagem multiespectral original.

O segundo teste procurou aferir a capacidade de transferência de detalhes de cada técnica a partir da comparação de detalhes entre a imagem pancromática e as fusões. Para isso foram extraídos os detalhes das imagens fusionadas e pancromática original através de uma operação de filtragem passa-alta com um filtro Laplaciano 3x3, com soma nula. Assim, novas imagens foram geradas contendo apenas os detalhes de cada imagem. Em seguida foi calculado o coeficiente de correlação entre a imagem pancromática filtrada e cada banda das imagens fusionadas filtradas, conforme seguido por Pinho et al. 2005. Os filtros Laplacianos são filtros do tipo passa alta, úteis na detecção de bordas. Esses filtros eliminam as feições de baixa frequência, deixando apenas as de alta frequência, normalmente expressas por bordas ou limites entre áreas de diferentes valores de número digital (DN). Essas filtragens são ricas em informação textural, porém pobres em contraste, com pouca informação tonal, sendo portanto de difícil interpretação (CROSTA, 1992).

Por último foi realizado um teste para verificar a fidelidade espectral das imagens fusionadas com relação à imagem multiespectral original. Como possuem resoluções espaciais diferentes, não é possível compará-las. Sendo assim, as imagens tiveram de ser degradadas de tal forma que as imagens fusionadas tivessem a mesma resolução espacial da imagem multiespectral original. A fusão entre as imagens degradadas é considerada uma fusão simulada somente para efeito de avaliação. Para esse estudo, a imagem pancromática original foi degradada para resolução de 10m e a multiespectral original para 40m. A fusão entre essas duas imagens gerou uma nova imagem fusionada com resolução de 10m. Para a comparação das imagens foram

utilizadas a média, desvio padrão e correlação entre bandas. Para efeito de comparação entre as bandas da imagem original e das fusionadas, os histogramas de cada imagem foram plotados para melhor visualização.

4.6 CLASSES DE USO DO SOLO NA RFEE

Com base em trabalhos de campo e em mapeamentos efetuados previamente foram definidas 10 classes de uso do solo para a RFEE com as seguintes características:

- a) Predominância de Araucária: compreende o domínio da Floresta Ombrófila Mista, com predominância de *Araucaria angustifolia* no estrato superior;
- b) Floresta com taquara: representada pela Floresta Ombrófila Mista com predominância de taquara morta e seca em seu sub-bosque;
- c) Floresta Sem araucária: representada pela Floresta Ombrófila Mista com predominância de folhosas;
- d) Taquara: compreende as áreas de taquaral morto e seco praticamente puro, com pouquíssimas ou nenhuma árvore;
- e) Agricultura: área destinada ao uso em pesquisas agrícolas, sendo ocupada alternadamente por plantios de soja, trigo e milho;
- f) Fruticultura 1: Área da Reserva destinada a pesquisas em macieiras;
- g) Fruticultura 2: Área da Reserva destinada a pesquisas em pereiras;
- h) Água: representada pelos corpos d'água presentes na Reserva, como rios e lagos;
- i) Nuvem: nuvens que aparecem na imagem;
- j) Sombra: está relacionada exclusivamente à sombra das nuvens que ocorreram na Reserva à hora da passagem do satélite.

4.7 ÁRVORE DE DECISÃO

Na classificação pelo algoritmo de árvore de decisão (C4.5) foi utilizado um aplicativo de mineração de dados, o WEKA, para geração do conjunto de regras da árvore e, posteriormente, um aplicativo de processamento de imagens, o ENVI©, para a classificação digital. Nesta técnica de classificação foram utilizadas 61 variáveis (13 espectrais e 48 de

textura), geradas a partir das quatro bandas originais do satélite ALOS. As variáveis são: 4 bandas originais, 4 componentes principais, NDVI, SAVI, RVI, medidas de ocorrência e co-ocorrência. As variáveis utilizadas no modelo estão mostradas na **Tabela 5**. Todas estas variáveis foram organizadas em um único arquivo de imagem.

TABELA 5: Variáveis utilizadas na técnica de classificação por árvore de decisão

	Simbologia utilizada	Variáveis
ESPECTRAIS	Banda 1	Banda 1 original
	Banda 2	Banda 2 original
	Banda 3	Banda 3 original
	Banda 4	Banda 4 original
	PC_1	Primeira Componente Principal
	PC_2	Segunda Componente Principal
	PC_3	Terceira Componente Principal
	PC_4	Quarta Componente Principal
	NDVI	Índice de Vegetação de Diferença Normalizada
	SAVI_1	Índice de Vegetação Ajustado para o Solo com constante 1
	SAVI_05	Índice de Vegetação Ajustado para o Solo com constante 0,5
	SAVI_025	Índice de Vegetação Ajustado para o Solo com constante 0,25
	RVI	Índice de Vegetação da Razão Simples
	TEXTURAIS	Oc_DR_B1
Oc_mean_B1		Medidas de ocorrência - média da banda 1
Oc_var_B1		Medidas de ocorrência - Variância da banda 1
Oc_entropy_B1		Medidas de ocorrência - entropia da banda 1
Oc_DR_B2		Medidas de ocorrência - Amplitude de dados da banda 2
Oc_mean_B2		Medidas de ocorrência - média da banda 2
Oc_var_B2		Medidas de ocorrência - Variância da banda 2
Oc_entropy_B2		Medidas de ocorrência - entropia da banda 2
Oc_DR_B3		Medidas de ocorrência - Amplitude de dados da banda 3
Oc_mean_B3		Medidas de ocorrência - média da banda 3
Oc_var_B3		Medidas de ocorrência - Variância da banda 3
Oc_entropy_B3		Medidas de ocorrência - entropia da banda 3
Oc_DR_B4		Medidas de ocorrência - Amplitude de dados da banda 4
Oc_mean_B4		Medidas de ocorrência - média da banda 4
Oc_var_B4		Medidas de ocorrência - Variância da banda 4
Oc_entropy_B4		Medidas de ocorrência - entropia da banda 4
Cooc_mean_B1		Medidas de co-ocorrência - média da banda 1
Cooc_var_B1		Medidas de co-ocorrência - variância da banda 1
Cooc_homog_B1		Medidas de co-ocorrência - homogeneidade da banda 1

Continua...

Simbologia utilizada	Variáveis
Cooc_contrast_B1	Medidas de co-ocorrência - contraste da banda 1
Cooc_dissim_B1	Medidas de co-ocorrência - dissimilaridade da banda 1
Cooc_entropy_B1	Medidas de co-ocorrência - entropia da banda 1
Cooc_2_moment_B1	Medidas de co-ocorrência - segundo momento da banda 1
Cooc_correl_B1	Medidas de co-ocorrência - correlação da banda 1
Cooc_mean_B2	Medidas de co-ocorrência - média da banda 2
Cooc_var_B2	Medidas de co-ocorrência - variância da banda 2
Cooc_homog_B2	Medidas de co-ocorrência - homogeneidade da banda 2
Cooc_contrast_B2	Medidas de co-ocorrência - contraste da banda 2
Cooc_dissim_B2	Medidas de co-ocorrência - dissimilaridade da banda 2
Cooc_entropy_B2	Medidas de co-ocorrência - entropia da banda 2
Cooc_2_moment_B2	Medidas de co-ocorrência - segundo momento da banda 2
Cooc_correl_B2	Medidas de co-ocorrência - correlação da banda 2
Cooc_mean_B3	Medidas de co-ocorrência - média da banda 3
Cooc_var_B3	Medidas de co-ocorrência - variância da banda 3
Cooc_homog_B3	Medidas de co-ocorrência - homogeneidade da banda 3
Cooc_contrast_B3	Medidas de co-ocorrência - contraste da banda 3
Cooc_dissim_B3	Medidas de co-ocorrência - dissimilaridade da banda 3
Cooc_entropy_B3	Medidas de co-ocorrência - entropia da banda 3
Cooc_2_moment_B3	Medidas de co-ocorrência - segundo momento da banda 3
Cooc_correl_B3	Medidas de co-ocorrência - correlação da banda 3
Cooc_mean_B4	Medidas de co-ocorrência - média da banda 4
Cooc_var_B4	Medidas de co-ocorrência - variância da banda 4
Cooc_homog_B4	Medidas de co-ocorrência - homogeneidade da banda 4
Cooc_contrast_B4	Medidas de co-ocorrência - contraste da banda 4
Cooc_dissim_B4	Medidas de co-ocorrência - dissimilaridade da banda 4
Cooc_entropy_B4	Medidas de co-ocorrência - entropia da banda 4
Cooc_2_moment_B4	Medidas de co-ocorrência - segundo momento da banda 4
Cooc_correl_B4	Medidas de co-ocorrência - correlação da banda 4

Foram utilizadas 317 amostras de treinamento de 3 x 3 pixels cada uma (30 x 30 m). Congalton e Green (2009) sugerem como uma boa regra coletar no mínimo 50 amostras por classe para áreas menores do que 1 milhão de acre (1 million acres) e com menos de 12 classes, além de ajustar a quantidade de amostras com base na importância de cada classe para o mapeamento. No caso da RFEE foi dada mais importância às classes de tipologias florestais, onde foi possível alocar um maior número de parcelas (**Tabela 6**). Congalton e Green (2009) comentam ainda que este tamanho de amostra (3 x 3 pixels) é frequentemente utilizado como

unidade amostral em imagens de resolução espacial moderada (10m, 30m). Segundo os autores, um grupo de pixels (ao invés de somente 1 pixel) minimiza problemas de registro por ser mais fácil localizar os dados sobre a referência. Como citado anteriormente, é recomendável que o erro de registro seja de aproximadamente meio pixel, sendo assim, a seleção de um grupo homogêneo de 3 x 3 pixels garante que o centro da amostra cairá dentro desse agrupamento (CONGALTON e GREEN, 2009).

Utilizou-se também 71 amostras para validação do modelo. Esses dois conjuntos de amostras foram distribuídos aleatoriamente pela imagem de acordo com as classes previamente estabelecidas (**Figura 15, Tabela 6**) e seus respectivos valores de números digitais foram coletados em todas as imagens utilizadas. Para a definição dessas amostras na imagem, foi utilizada uma composição colorida falsa cor (R4G3B2).

Gerou-se uma tabela em planilha eletrônica com todos esses valores que, posteriormente, foi exportada em formato “txt” para a montagem do arquivo final com cabeçalho em conformidade para abertura no *software* WEKA (apêndice 1). Dos recursos disponíveis no *software* Weka, foi utilizado o algoritmo de indução de árvore de decisão C4.5 implementado em sua versão para linguagem Java (no Weka) com o nome J4.8.

O conjunto de regras da árvore gerado no *software* WEKA foi inserido no *software* ENVI©. Para a geração da árvore foi utilizado o classificador por árvore de decisão e matemática de bandas. Em seguida foi efetuada uma verificação da acuracidade temática utilizando o Índice Kappa e matriz de confusão.

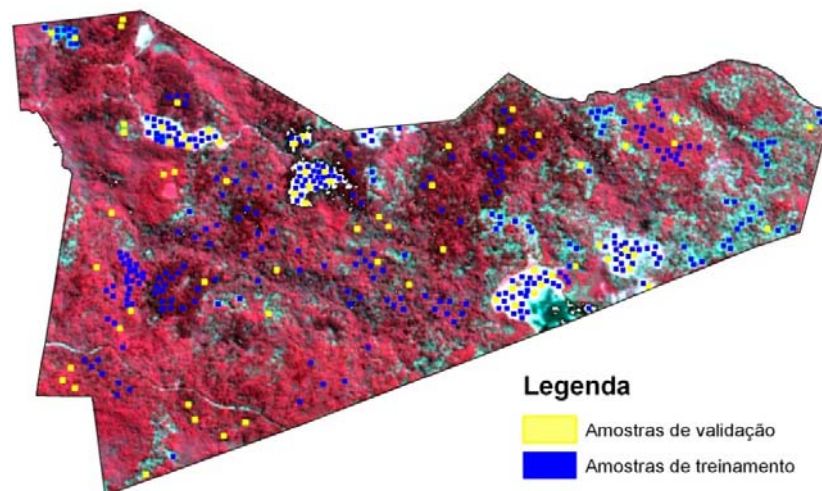


FIGURA 15: Espacialização das amostras de treinamento e validação

TABELA 6: Número de amostras de treinamento e validação por classe

Classes	Treinamento	Validação
Floresta com araucária	50	10
Floresta sem araucária	50	10
Floresta com taquara	55	11
Taquara	50	10
Agricultura	40	10
Fruticultura 1	4	1
Fruticultura 2	4	2
Água	10	3
Nuvem	26	6
Sombra de nuvem	28	8
Total	317	71

4.8 SEGMENTAÇÃO DE IMAGENS

4.8.1 Seleção de Bandas para a Segmentação

Devido à limitação do *software* SPRING em processar 61 bandas juntas, no processo de segmentação foi necessário reduzir o número de variáveis. Para isso foi utilizada a técnica de Análise de *Cluster* e uma análise de correlação entre bandas. Foi utilizada a mesma matriz de dados da árvore de decisão, com os valores de pixels das 317 amostras de treinamento em 61 imagens. O método utilizado para o agrupamento foi o de Ward empregando como critério de similaridade a distância Euclidiana.

O método de Ward (1963) tem como princípio propor um método de agrupamento que é fundamentado na mudança de variação entre os grupos e dentro dos grupos que estão sendo formados em cada passo do agrupamento. Inicialmente cada elemento é considerado como um conglomerado. Calcula-se a soma de quadrados de cada conglomerado. Os dois conglomerados que minimizam a distância são combinados, calculando-se então a soma de quadrados novamente e assim sucessivamente (MINGOTE, 2005).

4.8.2 Algoritmo de Segmentação

A maioria dos algoritmos de segmentação emprega alguma operação de gradiente local para detectar limites de regiões (detecção de bordas) ou agrega pixels vizinhos, similares, formando regiões maiores (crescimento de regiões).

No presente estudo foi utilizado o algoritmo de crescimento de regiões disponível no *software* SPRING. Este algoritmo rotula cada pixel como uma região diferente de acordo com um critério de similaridade e área. Este critério de similaridade baseia-se em um teste de hipóteses de média. Se a similaridade entre dois pixels ou região satisfaz o critério estabelecido e estas regiões são adjacentes, então elas são agregadas; caso contrário o processo se inicia novamente (SPRING, 2008a). A similaridade é o limiar referente à proximidade radiométrica entre pixels, correspondendo ao valor da distância euclidiana mínima entre as médias das regiões, abaixo da qual duas regiões são consideradas similares e, então, agrupadas. O limiar de área é o número mínimo de pixels para que uma região seja individualizada, sendo, portanto, o tamanho mínimo que cada segmento deve possuir na operação de segmentação (DLUGOSZ, 2005).

Antecedendo à fase de classificação, as seis bandas selecionadas pela Análise de *Cluster*, foram importadas para o *software* SPRING onde se realizou o processo de segmentação por “crescimento de regiões”, que posteriormente serviram de amostras para a classificação.

Foram testados vários pares de limiares de similaridade e área. A avaliação realizada foi somente visual, comparando as regiões formadas com a imagem ALOS em uma composição colorida falsa cor (R4G3B2) e verificando se essas regiões representavam bem as classes pré-definidas.

4.9 CLASSIFICAÇÃO SUPERVISIONADA DA IMAGEM SEGMENTADA

Na classificação supervisionada as regiões formadas no processo de segmentação serviram como amostras de treinamento para a classificação. A correspondência entre as regiões e suas respectivas classes de uso foram obtidas pela superposição das Regiões de Interesse (ROIs) sobre a imagem segmentada. O algoritmo utilizado neste processo foi o *Bhattacharya*, presente no *software* SPRING e o limiar de aceitação foi de 95%.

4.10 ACURACIDADE DA CLASSIFICAÇÃO DE IMAGENS DIGITAIS

A acuracidade (geral e específica de cada classe) das classificações por Árvore de decisão e *Bhattacharya* foram analisadas por matriz de erros, acurácia geral e coeficiente Kappa, gerados para cada classificação. Para esta análise foram utilizadas 71 amostras independentes distribuídas aleatoriamente de acordo com as diferentes classes pré-estabelecidas.

5. RESULTADOS E DISCUSSÕES

5.1 GEORREFERENCIAMENTO

Conforme Rosot (2001), o Erro Médio Quadrático Total (RMS) total descreve o erro de posicionamento característico de todos os pontos de controle em relação à equação, indicando a probabilidade com a qual uma posição mapeada poderá variar em relação à sua localização verdadeira. O RMS obtido no georreferenciamento da imagem ALOS multiespectral foi de 0,64 pixels, ou seja, 6,4 metros. Segundo Crosta (1992), a margem de erro recomendável é de ± 1 pixel, estando o resultado de acordo com o esperado.

Para a imagem pancromática, o RMS obtido foi 1,38 pixels. Apesar de ultrapassar o valor recomendado, que é de aproximadamente 1 pixel, esse erro, em metros (3,45m), é menor do que o obtido na imagem multiespectral.

5.2 FUSÃO DE IMAGENS

Os resultados das quatro técnicas de fusão (HSV, *Color normalized*, *Gram-Schmidt* e Componentes Principais) aplicadas em imagens ALOS são apresentados na **Figura 16**.

Na análise visual entre as fusões e a imagem multiespectral original, a fusão realizada pelo método HSV apresentou cores mais escuras, com pouco contraste, sendo que algumas regiões de vegetação aparecem quase pretas. Já as fusões pelo método *Gram-Schmidt* apresentaram as cores bem vibrantes, principalmente na operação que utilizou as quatro bandas. O método *color normalized* apresentou cores semelhantes ao método *Gram-Schmidt* com quatro bandas, apresentando tons azulados por toda a imagem. O que se pode perceber é que esses tons aparecem principalmente em regiões sombreadas. As fusões pelo método das Componentes Principais com 3 e 4 bandas não apresentaram cores tão vibrantes como os outros métodos (exceção para o HSV) e também não muito escuras como o HSV. Em todas as imagens foi possível ver o ganho em detalhes, como estradas mais visíveis e principalmente a vegetação ficou com melhor definição.

Para o segundo teste, a capacidade de transferência de detalhes foi medida pela correlação entre as imagens filtradas (pancromática e fusões). A **Tabela 7** mostra as correlações entre essas imagens.

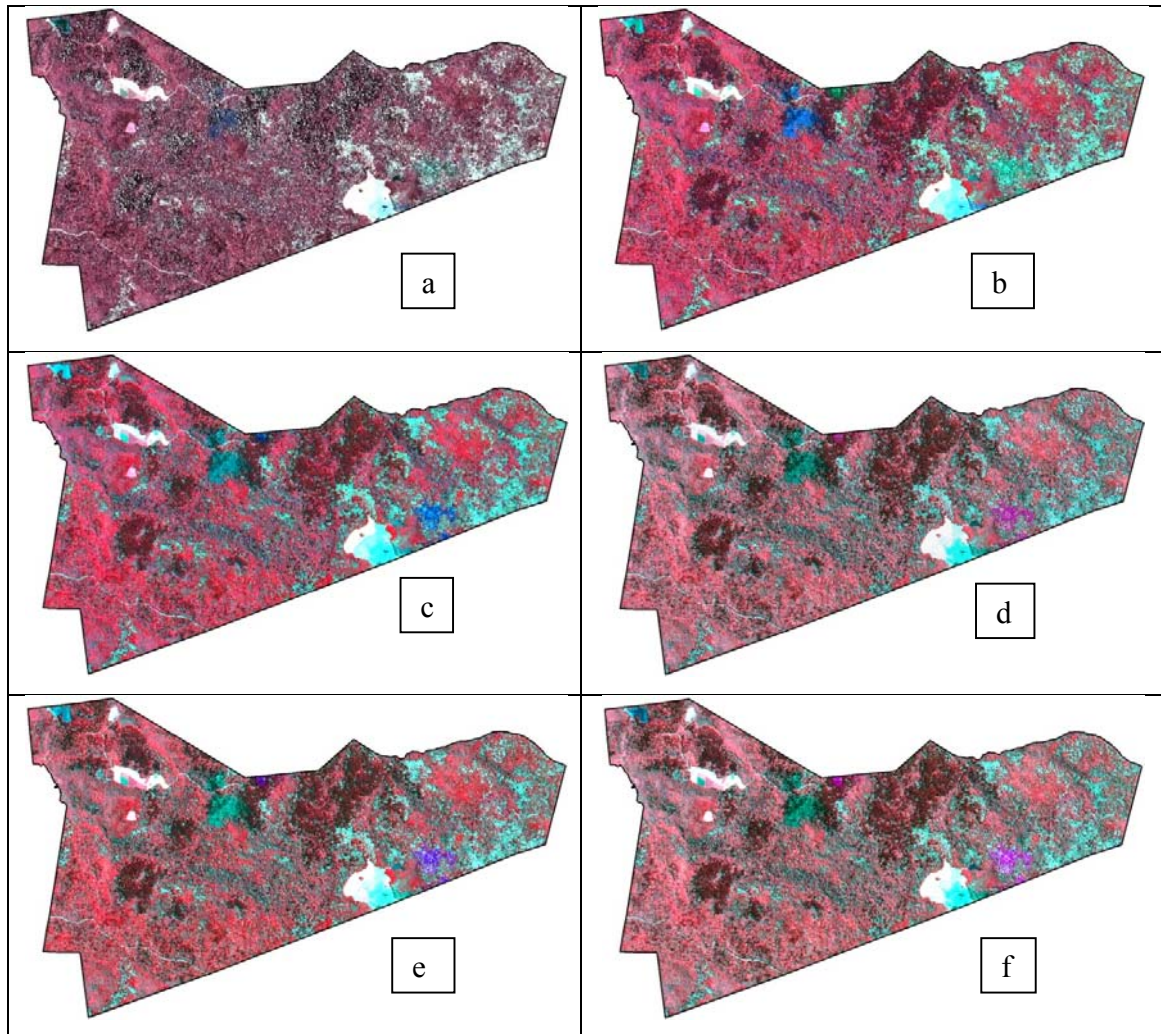


FIGURA 16: Imagens fusionadas pelos métodos: (a) HSV; (b) *Color normalized*; (c) *Gram-Schmidt*; (d) Componentes Principais; (e) *Gram-Schmidt* com 3 bandas e (f) Componentes Principais com 3 bandas;

TABELA 7: Coeficientes de correlação entre as imagens fusionadas filtradas e a pancromática filtrada, para o teste de transferência de detalhes.

Métodos de fusão	Banda 1	Banda 2	Banda 3	Banda 4
HSV	-	0.961	0.931	0.957
<i>Color normalized</i>	-	0.923	0.843	0.863
<i>Gram-Schmidt</i>	0.891	0.929	0.896	0.588
Componentes Principais	0.958	0.981	0.990	0.483
<i>Gram-Schmidt - 3 bandas</i>	-	0.968	0.965	0.674
Componentes Principais - 3 bandas	-	0.981	0.990	0.491

Nos métodos HSV e *Color Normalized* as correlações foram altas para as três bandas utilizadas. Para os outros métodos (*Color Normalized* e *Gram-Schmidt* com três e com quatro bandas) as correlações foram altas somente para as três primeiras bandas. A banda 4, nesses métodos obteve valores variando entre 0,48 a 0,67.

No terceiro teste, foi analisada a fidelidade espectral das imagens fusionadas com relação à multiespectral original. Na **Tabela 8** são apresentados os valores de média dos pixels, desvio padrão e correlação para a imagem multiespectral e para as fusões.

TABELA 8: Comparação entre as estatísticas da imagem multiespectral e das fusões

	Banda 1			Banda 2			Banda 3			Banda 4		
	Média	DP	Correl.	Média	DP	Correl.	Média	DP	Correl.	Média	DP	Correl.
Original	75,93	9,33	1,00	55,38	12,91	1,00	43,52	19,08	1,00	57,45	10,91	1,00
HSV	-	-	-	88,22	54,00	0,59	72,05	56,95	0,77	88,49	48,00	0,32
CN	-	-	-	25,11	5,32	0,56	19,42	7,43	0,80	26,03	5,21	0,42
GS	75,78	9,80	0,50	55,58	12,73	0,66	44,03	19,58	0,74	57,32	10,97	0,54
PC	76,01	8,46	0,65	55,51	12,56	0,64	43,69	19,71	0,68	57,49	10,82	0,64
GS - 3B	-	-	-	55,49	12,94	0,65	43,95	19,69	0,73	57,27	10,79	0,52
PC - 3B	-	-	-	55,51	12,46	0,66	43,69	19,44	0,68	57,49	10,75	0,37

DP = desvio padrão

Correl. = correlação

Pode-se observar que a média e o desvio padrão dos métodos HSV e *Color Normalized* se mostraram bem diferentes dos valores da imagem original. A média e o desvio padrão estão relacionados com o brilho e contraste da imagem, respectivamente. Havendo uma menor variação na média, o brilho é conservado. Uma menor variação entre os valores de desvio padrão demonstra melhor preservação do contraste da imagem (PINHO et al. 2005).

Segundo Fuckner (2007), a fusão pelo método IHS apresenta bons resultados quanto à representação de cores somente quando a imagem pancromática apresenta um intervalo espectral que engloba o intervalo das três bandas multiespectrais de entrada. Apesar de o sistema de cores HSV não ser o mesmo do IHS, o princípio da técnica de fusão é o mesmo. Como mostrado na tabela 1, a banda pancromática engloba completamente as bandas dois e três da imagem multispectral, e apenas uma pequena faixa da banda 4. A baixa correlação nessa banda pode ser observada na tabela 3 para a técnica de fusão HSV, com o valor de 0,318. A técnica *Color Normalized* também apresentou correlação baixa nessa banda, com um valor de 0,41. Esses dois métodos tiveram seus histogramas bem diferentes comparados com a imagem original, apresentando mais contraste em comparação com os outros métodos, principalmente o HSV que teve seu contraste aumentado (**Figura 17**). O método *Color Normalized* teve suas curvas deslocadas para a esquerda, indicando que a imagem ficou mais escura do que a original. Para os outros métodos (*Gram-Schmidt* e Componentes principais), os histogramas da banda 4 se mostraram muito semelhantes à curva da imagem original. Para as outras bandas a semelhança não foi tão grande. Elas coincidiram com os valores dos números digitais, porém, não nos números de pixels presentes naqueles valores.

Telles Júnior e Rosa (2005) consideram que uma imagem com erros de registro pode levar ao surgimento de falsas cores e feições artificiais após a fusão. Comentam ainda que estas técnicas são válidas para produzir imagens para a interpretação visual, pois realçam a textura. Por outro lado, esses métodos podem levar a significantes distorções na resolução radiométrica das imagens.

Comparando os três testes realizados e também a análise dos histogramas, os métodos que apresentaram os melhores resultados foram o *Gram-Schmidt* e Componentes Principais.

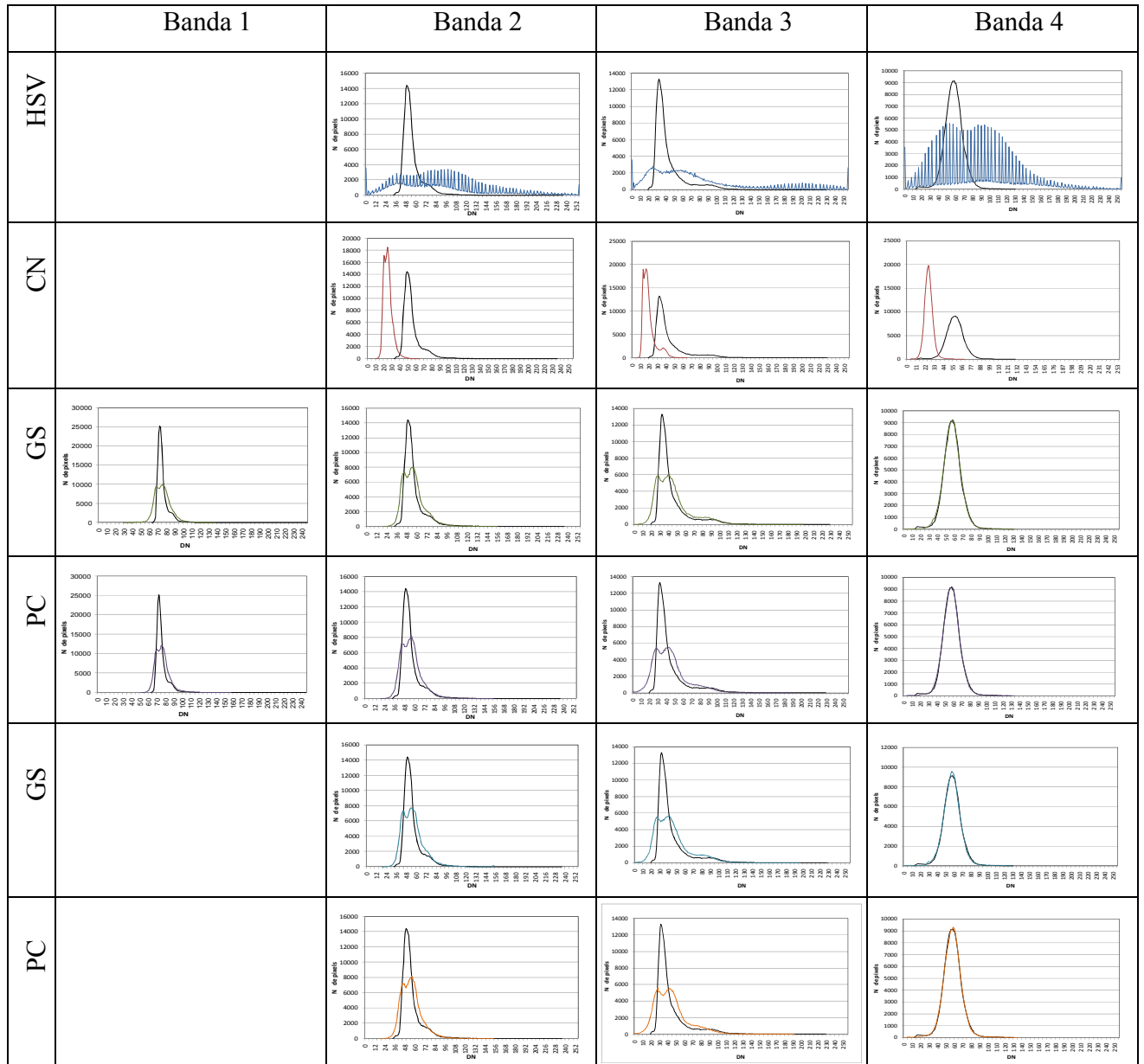


FIGURA 17: Histograma de comparação entre a imagem original e as fusões. As curvas em preto são da imagem original e as coloridas são das fusões.

5.3 ÁRVORE DE DECISÃO

A classificação dos dados realizada pelo *software* Weka apresentou resultado satisfatório em termos de correta discriminação das classes de uso do solo quando comparadas à verdade de campo. O algoritmo utilizado foi o C4.5 implementado no Weka com o nome de J4.8. O tempo

de processamento para construir o modelo foi extremamente rápido (1,09 segundos), uma vez que utilizou uma grande quantidade de dados (317 amostras x 9 pixels cada uma x 61 variáveis = 174.033 valores). Como resultado, uma árvore com 26 folhas foi gerada. O resultado da árvore já elaborada com todas as variáveis utilizadas na sua construção e conforme a saída original do *software* está mostrado no **apêndice 2**.

Após essa fase, os dados da árvore foram utilizados para fazer a classificação digital da imagem no *software* ENVI©. As variáveis foram então inseridas no *software* e a árvore executada, gerando a classificação mostrada na **Figura 18**. A estrutura da árvore formada utilizando o *software* ENVI© está mostrada na **Figura 19**.

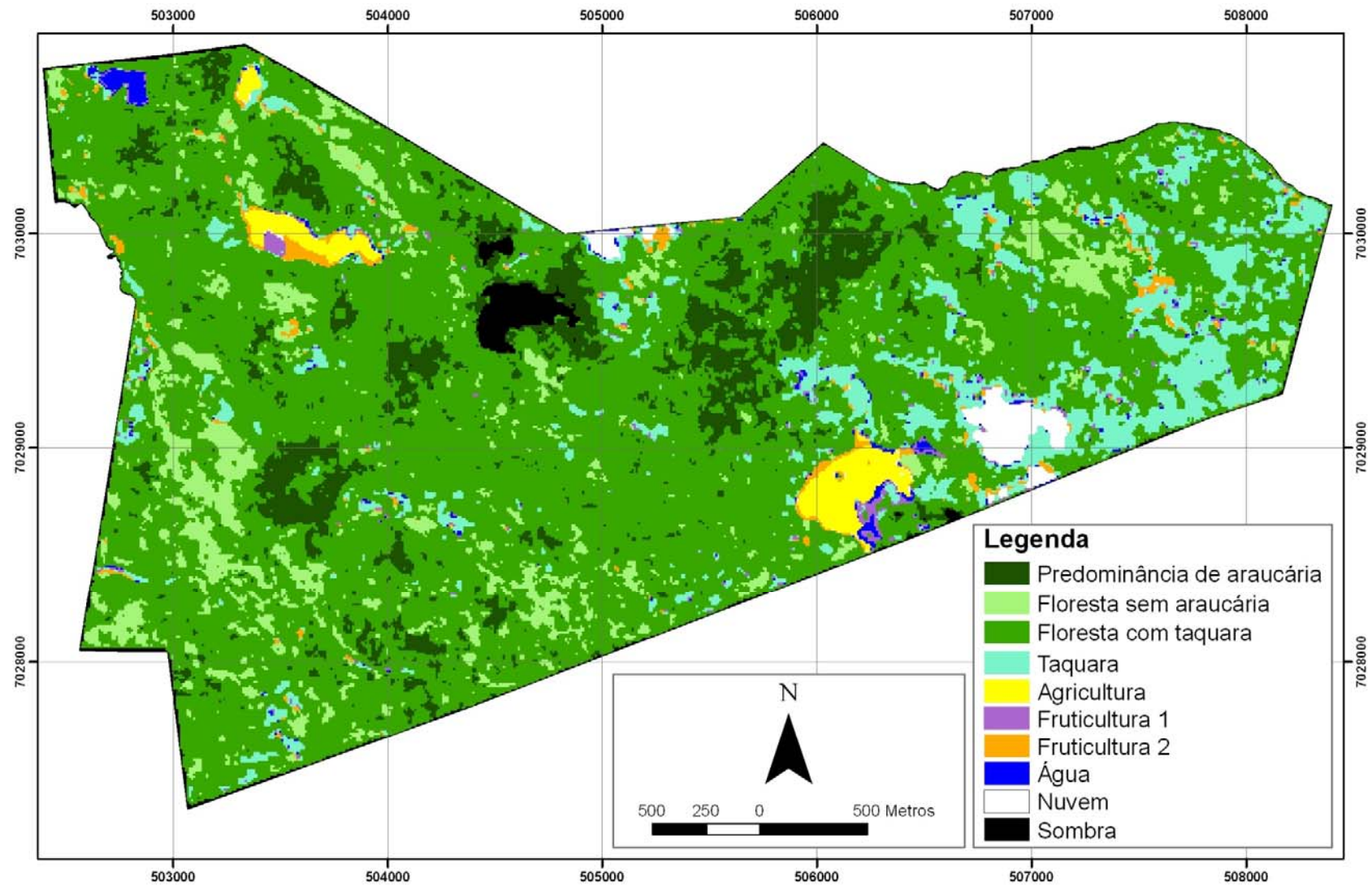


FIGURA 18: Resultado da classificação por árvore de decisão

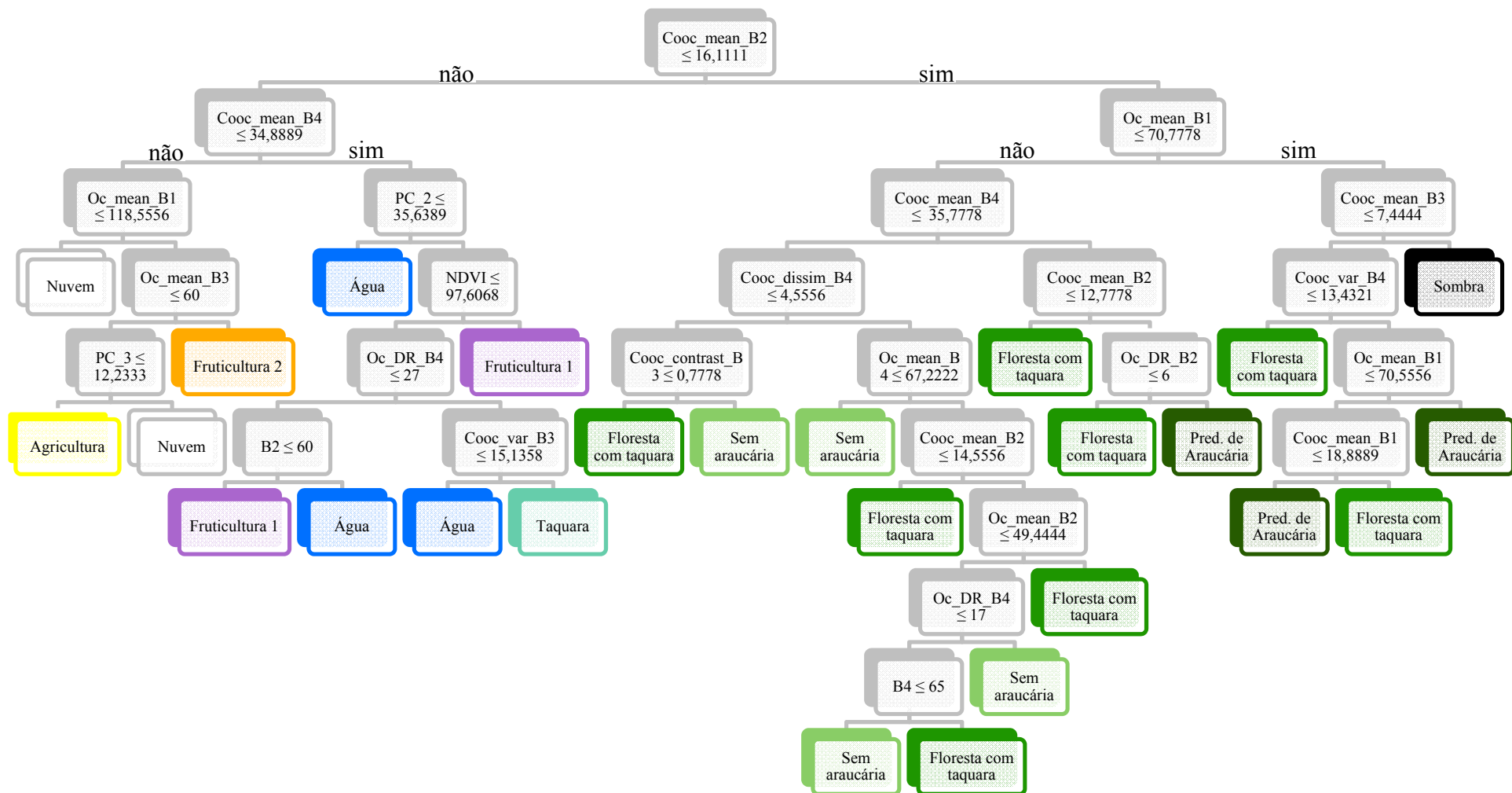


FIGURA 19: Instâncias e regras para a classificação dos dados na árvore de decisão

B1, B2, B3, B4: Bandas originais 1, 2, 3 e 4, respectivamente
 NDVI: Índice de Vegetação de Diferença Normalizada
 PC_2 e PC_3: Componentes Principais das bandas 2 e 3 respectivamente
 Oc_mean_BX: Medida de ocorrência de média da banda "X"
 Oc_DR_BX: Medida de ocorrência da amplitude dos dados da banda "X"
 Cooc_mean_BX: Medida de co-ocorrência de média da banda "X"

Cooc_var_BX: Medida de co-ocorrência de variância da banda "X"
 Cooc_contrast_BX: Medida de co-ocorrência de contraste da banda "X"
 Cooc_dissim_BX: Medida de co-ocorrência de dissimilaridade da banda "X"

Das 61 variáveis utilizadas, as 19 que entraram no modelo foram: Banda 2, banda 4, componentes principais 2 e 3, NDVI, medida de ocorrência de média das bandas 1, 2, 3 e 4, medida de ocorrência de amplitude dos dados das bandas 2 e 4, medida de co-ocorrência de médias das bandas 1, 2, 3 e 4, medida de co-ocorrência de variância das bandas 3 e 4, medida de co-ocorrência de contraste da banda 3 e medida de co-ocorrência de dissimilaridade da banda 4.

A primeira variável a entrar no modelo foi a medida de co-ocorrência da média da banda 2, sendo portanto a variável que apresentou o maior valor de ganho em comparação com as outras variáveis. A primeira divisão neste ramo se baseou na questão: o valor da amostra para a medida de co-ocorrência da média da banda 2 é menor ou igual a 16,1111 ou maior que 16,111? Esta variável conseguiu discriminar áreas de floresta (incluindo a classe sombra) e não floresta. A opção 1 determinou as três classes florestais, Floresta de Araucária, Floresta sem Araucária e Floresta com Taquara, além da classe Sombra. A opção 2 separou as classes Fruticultura 1 e 2, Agricultura, Nuvem, Água e Taquara. A classe que apresentou o menor conjunto de regras foi a classe Sombra, utilizando somente 3 variáveis, que foram as medidas de co-ocorrência das médias da banda 2, 1 e 3, respectivamente. Em contrapartida, as classes com o maior conjunto de regras foram Floresta de Araucária e Floresta sem Araucária, possuindo mais que três variáveis e seguindo instâncias diferentes para chegar à mesma classificação. Isto provavelmente se deve à variabilidade das classes.

As variáveis que mais apareceram no modelo foram as medidas de ocorrência e co-ocorrência de médias para todas as bandas, sendo que a medida de ocorrência de médias da banda 1 apareceu quatro vezes no modelo. Dos índices de vegetação utilizados somente o NDVI foi selecionado, aparecendo uma vez. Resultado semelhante foi apresentado por Rennó et al. (1996), com imagens SAR. Os resultados mostraram que a classificação baseada somente em informações tonais obtiveram um valor de Kappa ligeiramente superior a 0,50. Com o uso de 15 medidas de textura este valor aumentou para 0,90. Resultado diferente foi apresentado por Novo et al. (1998). Os autores tinham como objetivo avaliar o uso de textura em imagens RADARSAT para a discriminação de ambientes aquáticos. Utilizaram várias medidas de textura, e as mais frequentemente selecionadas foram média, desvio padrão e contraste. O resultado da classificação mostrou um mau desempenho com o valor de Kappa variando de 0,103 a 0,355. Concluíram que o uso de medidas de textura não aumenta a separabilidade entre os diferentes gêneros de plantas aquáticas, pois no caso específico dos ambientes aquáticos, as diferenças texturais entre áreas de

ocorrência de floresta e áreas de ocorrência de macrófitas se encontram relacionadas mais à textura da topografia do que a textura do dossel.

A validação do modelo foi efetuada também no *software* ENVI© através da opção de pós-classificação – matriz de confusão. Para essa validação foram utilizadas 71 amostras independentes e que não participaram do processo de criação da árvore (**Figura 15**). O resultado dessa classificação gerou uma acurácia geral de 97,0266% e um Kappa de valor 0,9659 o que significa, neste caso, que a classificação efetuada é 96,6% melhor do que seria obtido por uma classificação (atribuição de classes) aleatória. A matriz de confusão obtida é apresentada na **Tabela 9**. O trabalho de Colstoun et al. (2003) também avaliou a performance do classificador Árvore de Decisão na classificação de 11 tipos de cobertura da Terra utilizando imagens do satélite Landsat ETM+. O resultado que os autores obtiveram para acurácia global foi 82%. Porém, quando os autores consideraram somente floresta vs. não floresta, o valor da acurácia global foi aumentado para 99%.

TABELA 9: Matriz de confusão da classificação por árvore de decisão

Classe	Pred. de Araucaria	Sem araucaria	Floresta com taquara	Taquara	Agricultura	Fruticultura 1	Fruticultura 2	Água	Nuvem	Sombra	Total
Pred. de araucária	86	0	2	0	0	0	0	0	0	0	88
Floresta sem araucária	0	90	0	0	0	0	0	0	0	0	90
Floresta com taquara	4	0	97	0	0	0	0	1	0	0	102
Taquara	0	0	0	87	0	7	0	0	0	0	94
Agricultura	0	0	0	0	90	0	0	0	0	0	90
Fruticultura 1	0	0	0	0	0	1	0	1	0	0	2
Fruticultura 2	0	0	0	0	0	0	18	0	0	0	18
Água	0	0	0	3	0	1	0	25	0	0	29
Nuvem	0	0	0	0	0	0	0	0	54	0	54
Sombra	0	0	0	0	0	0	0	0	0	72	72
Total	90	90	99	90	90	9	18	27	54	72	639

Como pode ser observado na **tabela 9**, as classes que não apresentaram confusão nenhuma foram: agricultura, fruticultura 2, nuvem, sem araucária e sombra. As classes Araucária e Floresta com taquara apresentaram alguma confusão entre si. Isso pode ser devido, em parte, ao fato de a Floresta com Taquara apresentar indivíduos de araucária também, isolados.

Também é possível analisar a acurácia do produtor e do usuário, que são mostradas na **Tabela 10**. Em relação à acurácia do produtor as classes que apresentaram os maiores valores foram as que não apresentaram confusão, ou seja, agricultura, fruticultura 2, nuvem, sem araucária e sombra, apresentando valores de 100% de acurácia. A classe de menor valor foi a

fruticultura 1, com somente 11%. Na acurácia do usuário, as classes agricultura, fruticultura 2, nuvem, sem araucária e sombra também apresentaram o maior valor, 100% e a fruticultura 1 o menor, 50%. Este baixo valor da acurácia para a classe Fruticultura 1 provavelmente se deve ao fato da baixa representatividade de amostras para essa classe. Segundo Friedl e Brodley (1997) o algoritmo por árvore de decisão apresenta uma tendência em penalizar classes com baixo número de observações no conjunto de treinamento, porém este grau de penalização depende da regra de divisão e do procedimento de poda empregado pelo algoritmo. Na **Figura 20** pode-se observar detalhes da imagens para cada classe.

TABELA 10: Acurácia do produtor e do usuário para a classificação por árvore de decisão

Classe	Acurácia do Produtor	Acurácia do Usuário
Pred. de araucária	86/90 * 100 = 96	86/88 * 100 = 97
Floresta sem araucária	90/90 * 100 = 100	90/90 * 100 = 100
Floresta com taquara	97/99 * 100 = 98	97/102 * 100 = 95
Taquara	87/90 * 100 = 97	87/94 * 100 = 93
Agricultura	90/90 * 100 = 100	90/90 * 100 = 100
Fruticultura 1	1/9 * 100 = 11	1/2 * 100 = 50
Fruticultura 2	18/18 * 100 = 100	18/18 * 100 = 100
Água	25/27 * 100 = 93	25/29 * 100 = 86
Nuvem	54/54 * 100 = 100	54/54 * 100 = 100
Sombra	72/72 * 100 = 100	72/72 * 100 = 100

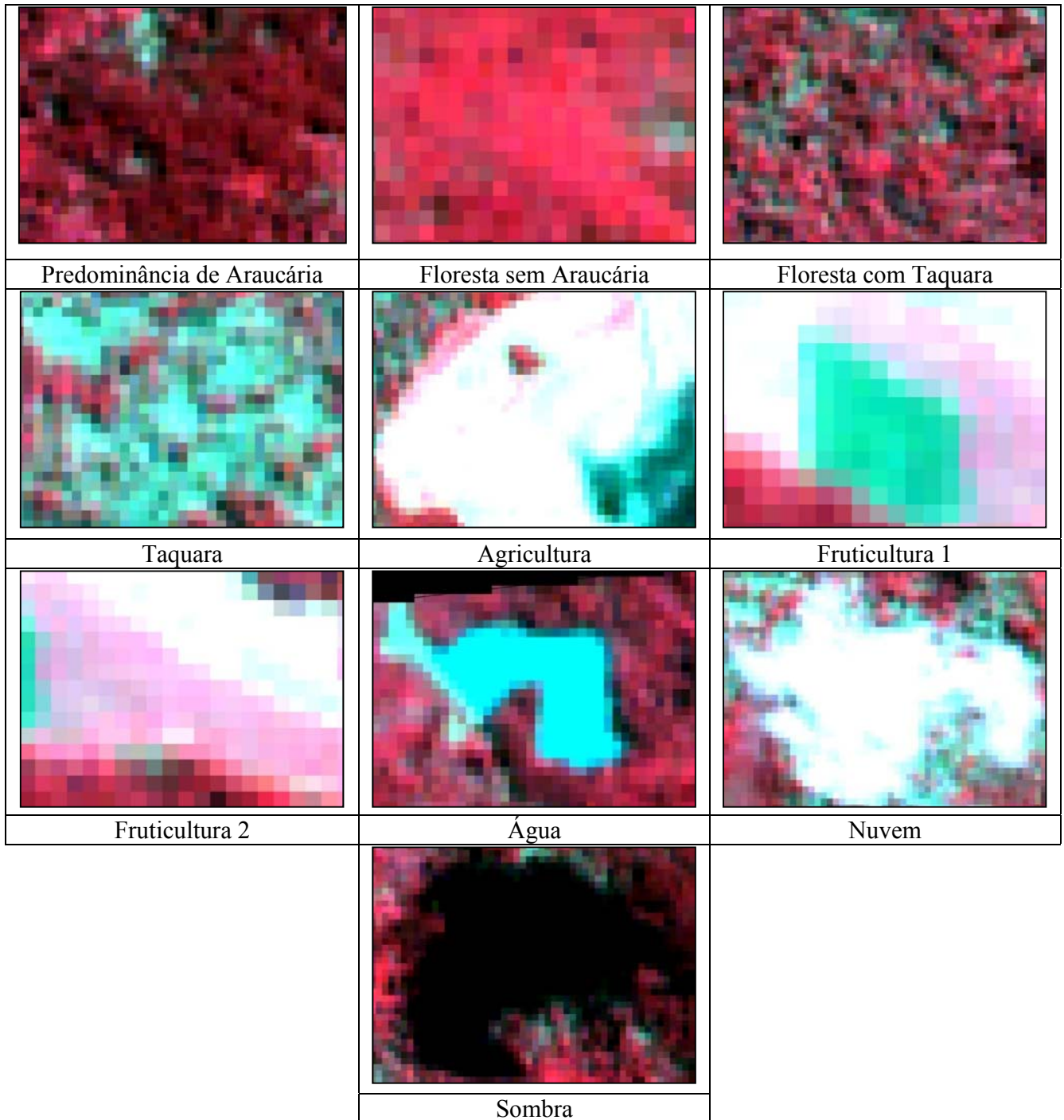


FIGURA 20: Imagens das classes de uso definidas para a Reserva

5.4 SELEÇÃO DE BANDAS PARA A SEGMENTAÇÃO

Conforme mencionado anteriormente, o *software* SPRING é limitado em relação ao número de bandas participantes no processo de segmentação, sendo necessário a redução do seu número. Assim, a Análise de *cluster* foi escolhida como sendo a técnica utilizada para seleção das bandas que entrariam no processo de segmentação.

Através da Análise de *cluster* foi possível selecionar seis bandas, das 61 existentes. O resultado da análise de *cluster* é um dendrograma, mostrado na **Figura 21**. Para a seleção foi utilizado um valor de 20% da linha de corte, sendo este critério subjetivo e específico para cada caso. A escolha desse limiar foi devida ao fato de algumas imagens apresentarem correlação baixa entre si. Como pode ser observado na **Figura 21**, as quatro imagens do lado esquerdo do gráfico (B66, B58, B50, B42) se unem em um único grupo somente acima de 50% da distância de ligação em comparação com os outros dois grupos do lado direito do gráfico, que se uniram a uma distância abaixo de 20%. Por esse motivo decidiu-se reduzir a distância da linha de corte de 50% para 20%.

As imagens selecionadas pela análise de *cluster* foram as seguintes:

- Banda 3 original (B3)
- Medida de co-ocorrência da correlação da banda 1 (B42)
- Medida de co-ocorrência da correlação da banda 2 (B50)
- Medida de co-ocorrência da correlação da banda 3 (B58)
- Medida de co-ocorrência da correlação da banda 4 (B66)
- Medida de co-ocorrência de média da banda 4 (B59)

As imagens de medida de co-ocorrência da correlação para as bandas 1, 2, 3 e 4 foram separadas cada uma em um *cluster*. A banda 3 original e a imagem de medida de co-ocorrência de média da banda 4, foram selecionadas como representativas de um grupo. Elas foram escolhidas porque obtiveram a maior correlação média com as bandas do seu respectivo grupo.

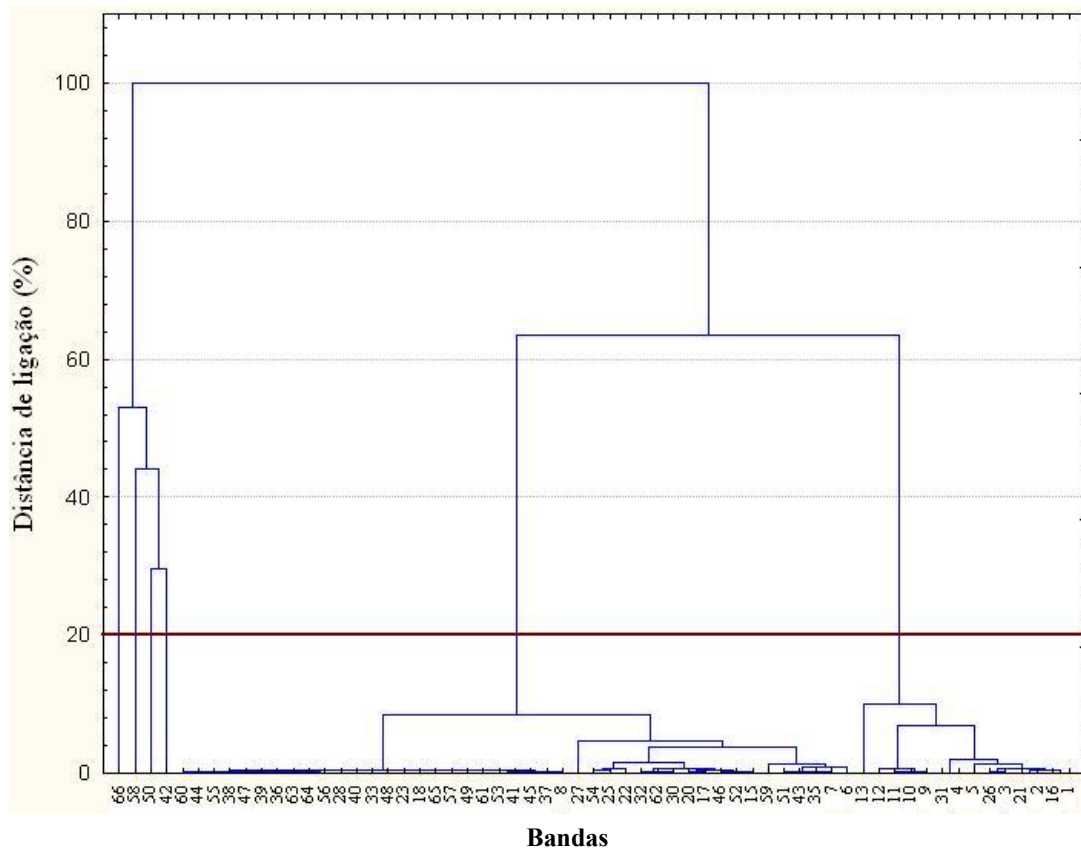


FIGURA 21: Dendrograma

5.5 SEGMENTAÇÃO DE IMAGENS

O conjunto de bandas selecionadas pela análise de *cluster* foi submetido ao processo de segmentação no *software* SPRING através do algoritmo “crescimento de regiões”. Foram testados vários limiares de similaridade e área, como mostrados na **Tabela 11**.

Uma fonte experiente para a avaliação das técnicas de segmentação é o olho humano (eCognition, 2000). Nenhum resultado de segmentação, mesmo avaliado quantitativamente, convencerá inteiramente se não satisfizer o olho humano. A manipulação consistente de contrastes locais é um pré-requisito para alcançar este objetivo (eCognition, 2000). Analisando visualmente as imagens segmentadas, e comparando-as com a imagem ALOS em composição colorida falsa cor (R4G3B2), a melhor combinação encontrada para valores de similaridade e área foi 5 e 180, respectivamente. Estes valores se mostraram apropriados para delimitar as regiões de interesse, gerando 206 polígonos, sendo que o menor deles possui 1,79 ha de área e o

maior possui 29 ha. Todos os limiars apresentaram boa distribuição de polígonos por toda a área estudada. As segmentações foram semelhantes, apresentando diferenças sutis entre si. Os limiars de área de 190, 200 e 300 não delimitaram algumas regiões de taquara. Os resultados dessas segmentações podem ser vistos no **Apêndice 3**. O resultado da segmentação escolhida aparece na **Figura 22**.

TABELA 11: Limiars de similaridade e área testados na segmentação

Similaridade	Área	N° polígonos
3	200	183
4	150	253
4	200	188
5	150	250
5	170	214
5	180	206
5	190	194
5	200	181
5	300	124

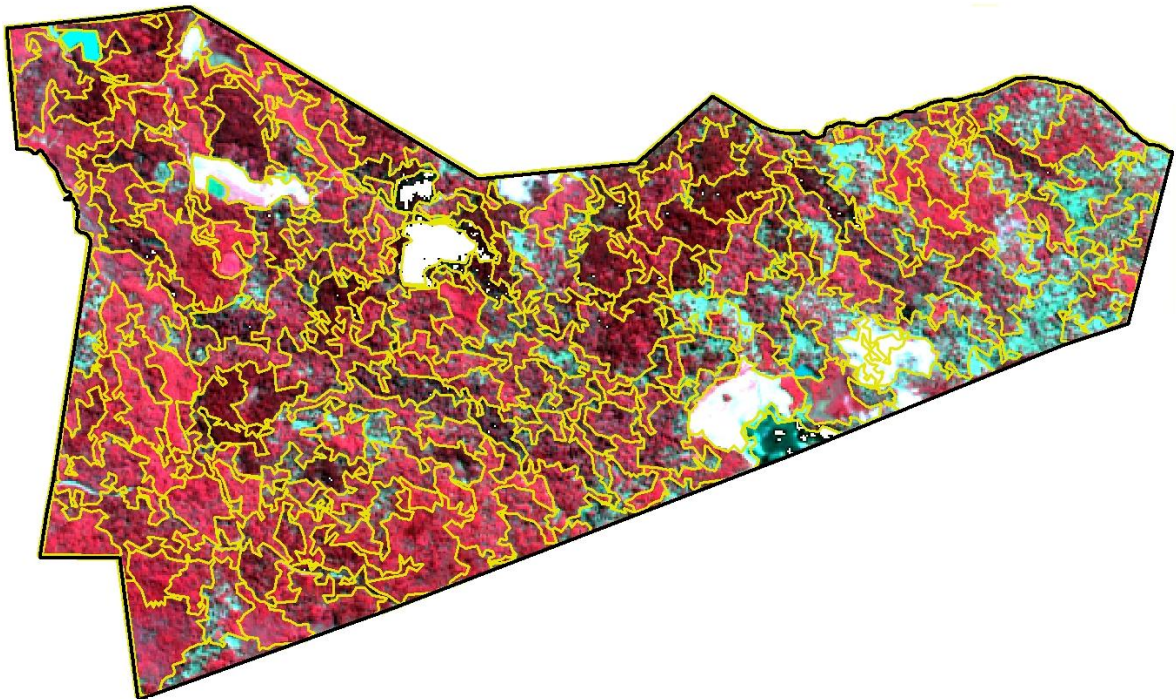


FIGURA 22: Imagem segmentada com limiars de similaridade e área 5 e 180, respectivamente

5.6 CLASSIFICAÇÃO DA IMAGEM SEGMENTADA

O método selecionado foi o de classificação supervisionada. Previamente à classificação foi realizada a segmentação. O limiar de similaridade e área escolhido foi de 5 e 180, respectivamente. As regiões formadas nesse processo serviram como amostras de treinamento para a classificação. Foram alocadas 32 regiões como amostras de treinamento em função da sobreposição das ROI's e avaliação da correspondência com as respectivas classes de uso do solo. O classificador utilizado foi o de *Bhattacharya*, com limiar de aceitação de 95%. É através desse limiar que o algoritmo classificador define se uma região será incorporada ou não a uma determinada classe (DLUGOSZ, 2005). O resultado dessa classificação pode ser visto na **Figura 23**.

Nesta técnica, o algoritmo de segmentação por crescimento de regiões não conseguiu separar as duas classes de fruticultura, incorporando-as às áreas de agricultura e taquara. Visualmente, esta classificação se apresentou mais “limpa”, sem a aparência de “*salt and pepper*”, característica das classificações digitais pixel a pixel.

A validação do modelo foi efetuada no *software* ENVI©, na opção de pós-classificação – matriz de confusão. Semelhante à classificação por árvore de decisão, as 71 amostras independentes foram utilizadas para calcular a matriz de erros, acurácia geral e coeficiente Kappa. O resultado dessa classificação gerou uma acurácia geral de 78,9216% e um Kappa de valor 0,7552 o que significa, neste caso, que a classificação efetuada é 75,5% melhor do que seria obtido por uma classificação (atribuição de classes) aleatória. A matriz de confusão obtida está apresentada na **Tabela 12**.

TABELA 12: Matriz de confusão da classificação por *Bhattacharya*

Classe	Pred. de Araucária	Sem araucaria	Floresta com taquara	Taquara	Agricultura	Água	Nuvem	Sombra	Total
Pred. de araucária	63	0	24	16	0	4	0	0	107
Floresta sem araucária	0	81	1	0	0	0	0	0	82
Floresta com taquara	27	9	47	3	0	0	0	0	86
Taquara	0	0	27	71	0	9	9	0	116
Agricultura	0	0	0	0	90	0	0	0	90
Água	0	0	0	0	0	14	0	0	14
Nuvem	0	0	0	0	0	0	45	0	45
Sombra	0	0	0	0	0	0	0	72	72
Total	90	90	99	90	90	27	54	72	612

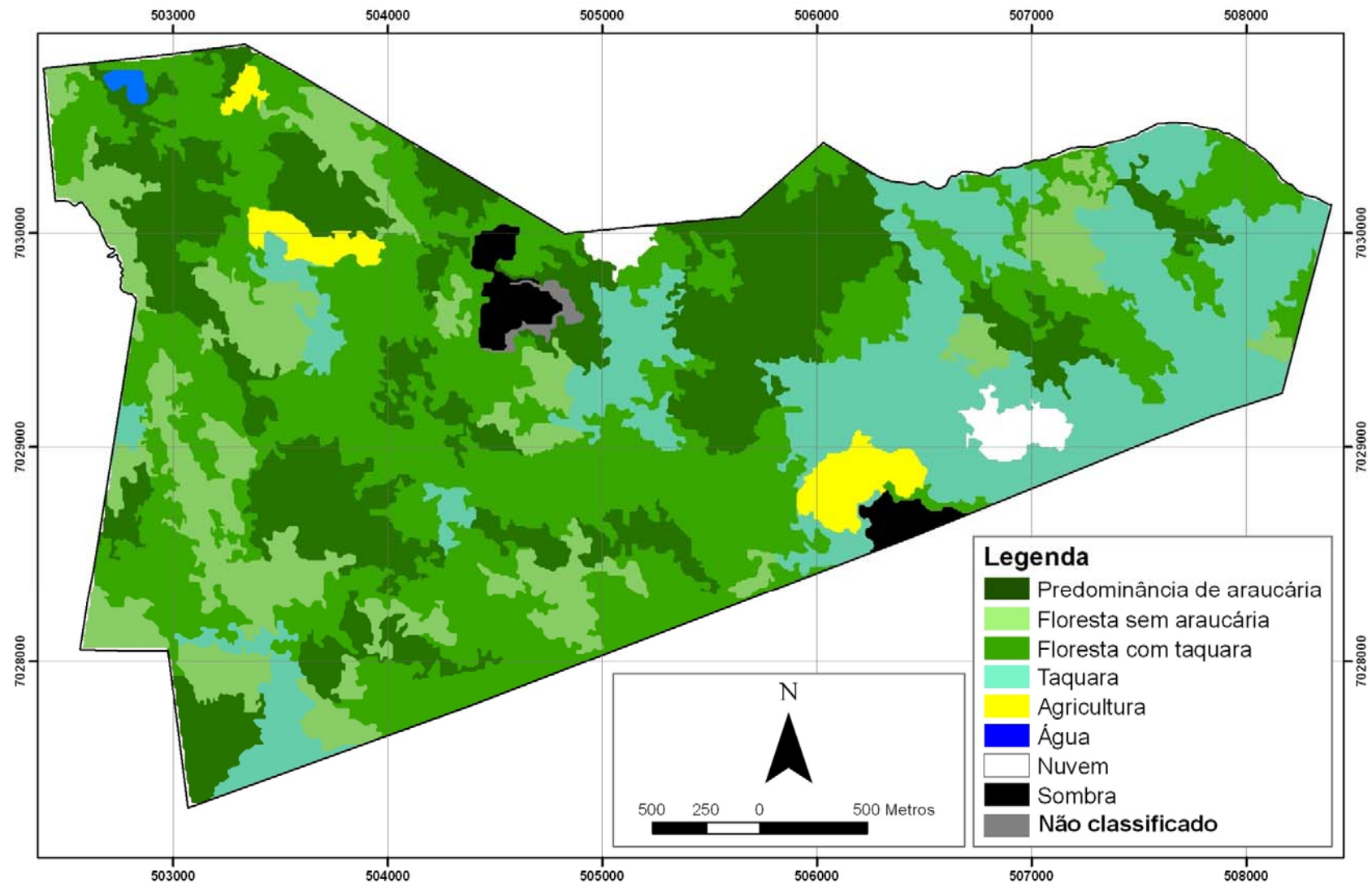


FIGURA 23: Resultado da classificação por regiões pelo classificador *Bhattacharya*

Como pode ser observado na **Tabela 12**, as classes que não apresentaram confusão nenhuma foram: agricultura e sombra. As classes Araucária e Floresta com taquara também apresentaram confusão nesta técnica. Esta última classe com grande confusão não somente com a Araucária, mas principalmente com taquara e em menor quantidade na classe Sem araucária.

As acurácias do produtor e do usuário estão mostradas na **Tabela 13**. Excluindo as classes que não apresentaram confusão (por isso a acurácia do produtor de 100%) a classe que apresentou o maior valor para a acurácia do produtor foi a Sem araucária (90%), seguida da classe Nuvem (83%) e Taquara (78%). O menor valor apresentado foi da classe Floresta com Taquara (47%).

Para a acurácia do usuário as classes com valores de 100% foram Agricultura, Água, Nuvem e Sombra, seguida da classe Sem araucária (99%). O menor valor obtido foi da classe Floresta com Taquara (55%).

TABELA 13: Acurácia do produtor e do usuário para a classificação por *Bhattacharya*

Classe	Acurácia do Produtor	Acurácia do Usuário
Pred. de araucária	$63/90 * 100 = 70$	$63/107 * 100 = 59$
Floresta sem araucária	$81/90 * 100 = 90$	$81/82 * 100 = 99$
Floresta com taquara	$47/99 * 100 = 47$	$47/86 * 100 = 55$
Taquara	$71/90 * 100 = 78$	$71/116 * 100 = 61$
Agricultura	$90/90 * 100 = 100$	$90/90 * 100 = 100$
Água	$14/27 * 100 = 52$	$14/14 * 100 = 100$
Nuvem	$45/54 * 100 = 83$	$45/45 * 100 = 100$
Sombra	$72/72 * 100 = 100$	$72/72 * 100 = 100$

5.7 COMPARAÇÃO ENTRE OS MÉTODOS DE CLASSIFICAÇÃO

Neste trabalho foram testadas duas técnicas de classificação de imagens para a cobertura da Terra. A primeira delas foi a Árvore de Decisão, na qual o conjunto de regras da árvore foi realizado no *software* WEKA e o processamento da imagem no *software* ENVI. Nesta técnica foram utilizadas 61 variáveis (13 espectrais 48 texturais) obtidas a partir de transformações das bandas originais. A segunda técnica foi a classificação por regiões, utilizando o algoritmo *Bhattacharya*, disponível no *software* SPRING. Nesta técnica é necessário primeiramente fazer a segmentação da imagem para que posteriormente as regiões formadas sejam utilizadas como

amostras de treinamento. Inicialmente se usaria as 61 variáveis utilizadas na árvore de decisão, porém o *software* SPRING possui limitação quanto ao número de bandas participantes no processo de segmentação, portanto foram selecionadas 6 bandas através da análise de *cluster*.

A comparação entre as duas técnicas se deu visualmente e através do coeficiente Kappa. O que se pode observar é que o Kappa se mostrou um pouco diferente para as duas classificações. Para a Árvore de Decisão o valor obtido foi de 0,97 e para a classificação por regiões foi de 0,76. Isso significa que as classificações por Árvore de Decisão e por regiões são respectivamente 97% e 76% melhores do que uma classificação aleatória. Estes valores indicam que a classificação pode ser considerada excelente e muito boa respectivamente, segundo a classificação de Landis e Koch (1977)⁶, adaptada por Figueiredo e Carvalho (2006), apresentada na **Tabela 2**. Apesar da técnica Árvore de Decisão apresentar um coeficiente Kappa superior, sua classificação se mostrou com aparência de “*salt and pepper*”, com pixels isolados, assemelhando-se a uma imagem com ruído, ao contrário da classificação por regiões que apresenta áreas com aspecto contínuo (**Figura 24**). As classes apresentaram grande diferença nas duas classificações. As áreas de taquara são muito maiores e evidentes na classificação por regiões. A floresta com taquara que era dominante na classificação por Árvore de Decisão não é tão destacada no outro método. Na **Figura 25** mostra-se em detalhe uma parte da Reserva com as duas classificações. Na classificação “b” pode-se perceber que as classes estão bem mais abrangentes. No lado esquerdo da imagem existe uma área grande da classe “predominância de araucária” que na classificação “a” são apenas pequenas manchas. Isso acontece porque na classificação por regiões as regiões formadas levam em consideração a vizinhança do pixel “semente” e que são unidos de acordo com um critério de similaridade definido previamente. Pode-se perceber também que as áreas “fruticultura 1 e 2” sumiram na classificação “b”, elas foram incorporadas em outras classes. O critério de similaridade não permitiu que elas fossem delimitadas.

Os dois métodos são técnicas supervisionadas, uma vez que necessitam de amostras de treinamento para “ensinar” o classificador. No caso do classificador *Bhattacharya*, como citado anteriormente, essas amostras são as regiões formadas no processo de segmentação. Neste sentido esta técnica de classificação se mostra mais trabalhosa, uma vez que é necessário testar e avaliar vários pares de similaridade e área antes de efetuar a classificação propriamente dita. Estes pares muitas vezes apresentam diferenças sutis entre si, demandando intenso trabalho de análise visual.

Para a classificação por Árvore de Decisão a etapa mais trabalhosa foi a inserção dos dados de treinamento no *software* WEKA. É necessário que os dados estejam em formato de arquivo texto com cabeçalho especificando todas as variáveis que entrarão no modelo (Apêndice 1). Para a classificação da imagem através da Árvore de Decisão no *software* ENVI© o procedimento é simples mas requer atenção no momento de inserir as equações em cada nó da árvore e também na seleção das imagens utilizadas no modelo.

Levando em consideração esses resultados e tendo como idéia o uso dessa classificação para o planejamento do manejo florestal da Reserva Florestal Embrapa Epagri, a técnica escolhida para auxiliar na delimitação de unidades de manejo seria a que utiliza crescimento de regiões. Esta técnica, além de possuir o coeficiente Kappa classificado como “muito bom”, apresenta uma classificação mais “limpa”, com regiões homogêneas.

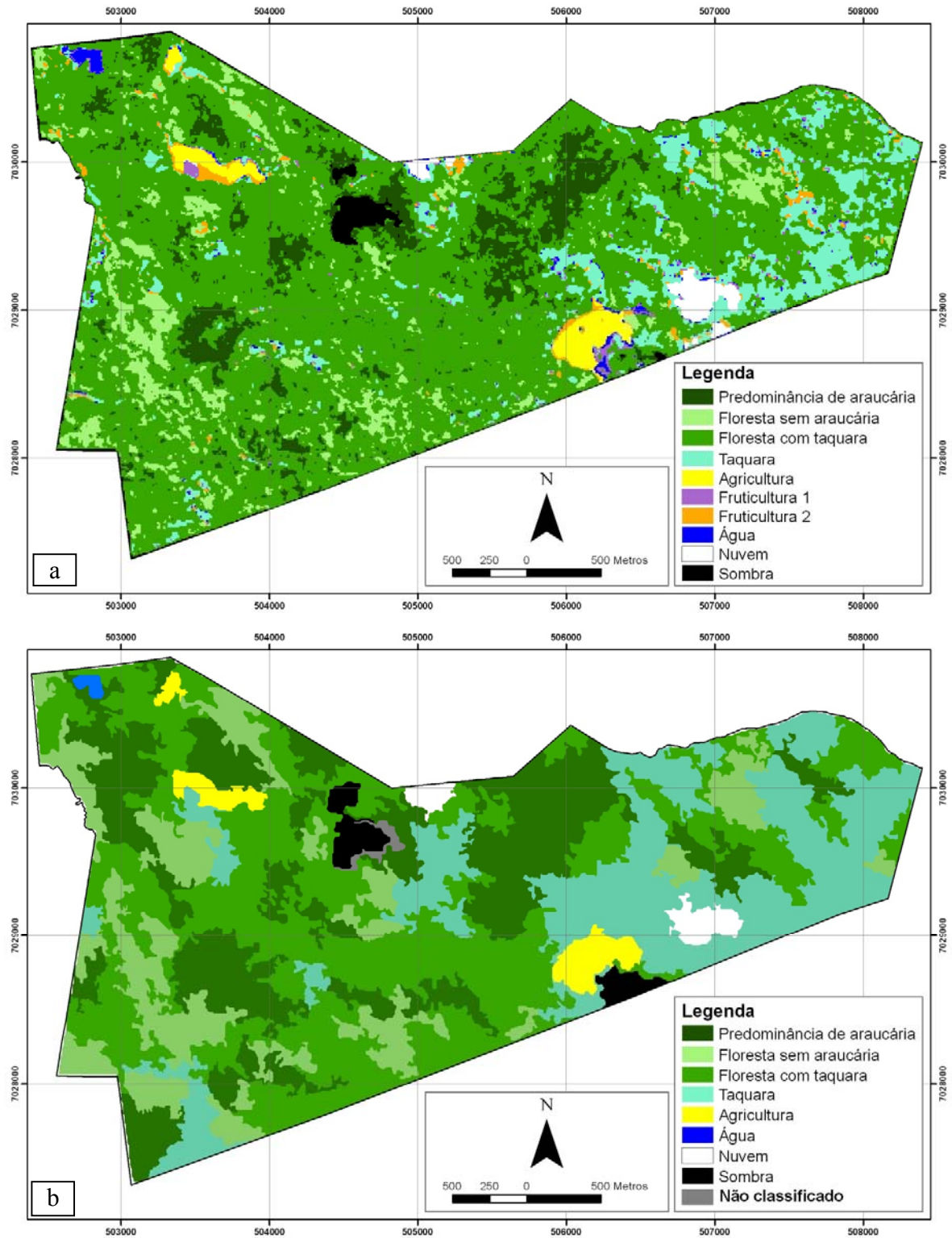


FIGURA 24: (a) Classificação por árvore de Decisão; (b) Classificação por regiões pelo classificador *Bhattacharya*

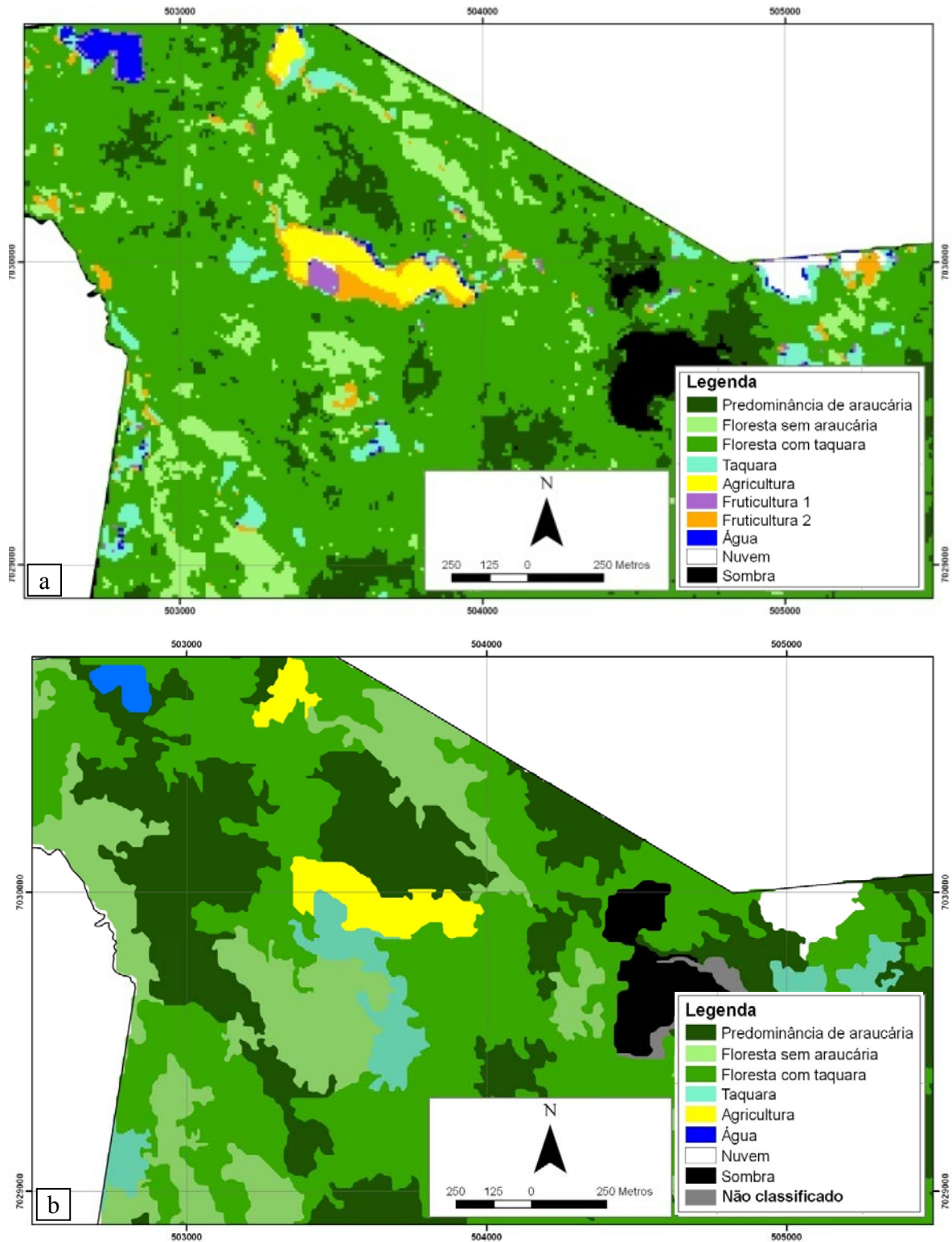


FIGURA 25: (a) Detalhes da classificação por Árvore de Decisão; (b) Detalhes da classificação por regiões pelo classificador *Bhattacharya*

6. CONCLUSÕES E RECOMENDAÇÕES

A análise dos resultados deste permitiu obter as seguintes conclusões:

- Com relação à fusão, no primeiro teste, de análise visual, as técnicas que apresentaram melhores resultados foram as Componentes Principais e a *Gram-Schmidt* com três bandas, apresentando cores que mais se aproximavam da imagem original. No segundo teste, de transferência de detalhes, as quatro técnicas de fusão se mostraram apropriadas. As técnicas apresentaram valores altos do coeficiente de correlação entre as bandas da fusão filtradas e da pancromática original filtrada, com exceção da banda 4 para as fusões *Gram-Schmidt* e Componentes Principais. No terceiro teste, de fidelidade espectral das imagens fusionadas com relação à multiespectral original, as técnicas *Gram-Schmidt* e Componentes Principais, tanto com três como com quatro bandas, se mostraram mais apropriadas. Os valores de média e desvio padrão para essas duas técnicas se aproximaram muito da imagem original, indicando a preservação do brilho e contraste. Comparando os três testes realizados e também a análise dos histogramas os métodos que apresentaram os melhores resultados foram o *Gram-Schmidt* e Componentes Principais;
- Na segmentação da imagem por crescimento de regiões os resultados mostraram-se adequados para a delimitação das classes de interesse. Os limiares de similaridade e área escolhidos foram de 5 e 180, respectivamente;
- Foram definidas 10 classes de uso do solo, sendo quatro destas referentes a tipologias florestais. A técnica de classificação por Árvore de Decisão delimitou essas 10 classes. Ela ainda foi extremamente rápida na geração das regras de decisão (1,09s). Já a classificação por *Bhattacharya* conseguiu delimitar somente oito classes, devido ao agrupamento de algumas delas (fruticultura 1 com taquara e fruticultura 2 com agricultura);
- Na aplicação da técnica Árvore de Decisão, das 19 variáveis que entraram no modelo, 14 são de textura, evidenciando a importância da geração de novas bandas através de filtros para a discriminação de feições em uma imagem. Na classificação por *Bhattacharya* essa proporção foi maior, pois, das seis bandas selecionadas pela análise de *cluster* para participar do processo de segmentação, cinco são de textura;

– No processo de classificação, a técnica Árvore de Decisão apresentou uma alta acurácia, tanto geral (97%) quanto para o coeficiente Kappa (0,966), em comparação com a técnica de classificação por regiões fornecida pelo algoritmo *Bhattacharya*, que apresentou acurácia geral de 79% e coeficiente Kappa de 0,755. Nas acurácias por classe (tanto do usuário quanto do produtor), a técnica Árvore de Decisão apresentou baixo valor em somente uma classe, ao contrário da classificação por *Bhattacharya* que apresentou baixos valores em quatro classes;

– Apesar de a classificação por árvore de decisão apresentar acurácia superior, a imagem ficou com aparência “*salt and pepper*”, com pixels isolados, assemelhando-se a uma imagem com ruído. Isso acontece em classificações realizadas pixel a pixel. Ao contrário, a classificação por *Bhattacharya* utiliza um algoritmo de crescimento de regiões, dando ao resultado da classificação um aspecto mais limpo e contínuo.

Para a sequência das pesquisas nesta área recomenda-se:

- Testar novas técnicas de fusão, utilizando também imagens provenientes de outros satélites e imagens de radar;
- Efetuar novos testes de segmentação de imagens utilizando diferentes limiares e diferentes *software*, como por exemplo, o e-Cognition;
- Utilizar variáveis como tipo de solo, declividade, variáveis dendrométricas e fitossociológicas na classificação por Árvore de Decisão para a delimitação de tipologias florestais.

REFERÊNCIAS BIBLIOGRÁFICAS

- ACCO, M. **Modelagem de dados do satélite Ikonos II para estimativa de micronutrientes na Floresta Ombrófila Mista Montana**. Dissertação (Mestrado em Engenharia Florestal) - Universidade Federal do Paraná-UFPR, Curitiba-PR, 68 p. 2004.
- ALVES, D.S., MOREIRA, J.C., KALIL, E.M., SOARES, J.V., FERNANDEZ, O., ALMEIDA, S., ORTIZ, J.D., AMARAL, S. Mapeamento do Uso da Terra em Rondônia Utilizando Técnicas de Segmentação e Classificação de Imagens TM. **Anais do VIII Simpósio Brasileiro de Sensoriamento Remoto**, Salvador-BA. p.71 – 79, 1996.
- ALVES, W.A.L., ARAÚJO, S.A. Avaliação da utilização de dois diferentes modelos de redes neurais na classificação de imagens com base em características texturais. **Exacta**. v. 4, n. 001, p. 77-86, 2006.
- BARBER, D. G., LEDREW, E. F. Sar sea ice using texture statistics: a multivariate approach. **Photogrammetric Engineering & Remote Sensing**. v. 57, n. 4, p. 385-395, abr. 1991.
- BARRET, E. C., CURTIS, L.F. **Introduction to Environmental Remote Sensing**: 4 ed. Cheltenham: Stanley Thornes Ltd. 457 p. 1999.
- BITTENCOURT, S., DALLA CORTE, A.P., SANQUETTA, C.R. Estrutura da Comunidade de *Pteridophyta* em uma Floresta Ombrófila Mista, Sul do Paraná, Brasil. **Silva Lusitana**. v.12, n. 2, p. 243 – 254, 2004.
- BRASIL. Congresso Nacional. **Lei nº 11.428, de 22 de dezembro de 2006**. Dispõe sobre a utilização e proteção da vegetação nativa do Bioma Mata Atlântica, e dá outras providências. In: BRASIL. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. Leis Ordinárias de 2006. [Brasília, DF], 2006. Publicada no DOU de 26.12.2006. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2006/Lei/_quadro2006.htm>. Acesso em: nov. 2007.
- BRITEZ, R. M. Aspectos ambientais a serem considerados na restauração da Floresta de Araucária no Estado do Paraná. **Pesquisa Florestal Brasileira**. n. 55, p. 37-43, 2007.
- CALDATO, S.; LONGHI, S.; FLOSS, P. Estrutura populacional de *Ocotea porosa* (Lauraceae) em uma Floresta Ombrófila Mista, Caçador, SC. **Ciência Florestal**. Santa Maria, v.9, n.1, p. 89-101, 1999.
- CANDEIAS, A.L.B.; SILVA, E.A. Extração de Estradas de uma imagem ETM+ Landsat usando Morfologia Matemática. **Anais do Congresso Brasileiro de Cadastro Técnico Multifinalitário (COBRAC)**. Universidade Federal de Santa Catarina. Florianópolis, 8p, 2004.
- CARREIRAS, J.M.B., PEREIRA, J.M.C., CAMPAGNOLO, M.L., SHIMABUKURO, Y.E. Assessing the extent of agriculture/pasture and secondary succession forest in the Brazilian Legal

Amazon using SPOT VEGETATION data. **Remote Sensing of Environment**, v. 101, n. 3, p. 283-298, April 2006.

CARVALHO JÚNIOR, O. A., COELHO, M.A.N., MARTINS, E.S., GOMES, R.A.T., COUTO JÚNIOR, A.F., OLIVEIRA, S.N., SANTANA, O.A. . Mapeamento da vegetação na floresta atlântica usando o classificador de árvore de decisão para integrar dados de sensoriamento remoto e modelo digital do terreno. **Revista Brasileira de Geofísica**. v.26, n. 3, p. 331-345, 2008.

CARVALHO, P. E. R. **Espécies florestais brasileiras: recomendações silviculturais, potencialidades uso da madeira**. Colombo: Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisas Florestais. EMBRAPA: CNPF. 640 p. 1994.

CCRS - Canada Centre for Remote Sensing **Fundamentals of Remote Sensing - Tutorial**. Ottawa, Ontario, Canadá (1999)

CENTENO, J. A. S. **Sensoriamento Remoto e Processamento de Imagens Digitais**. Departamento de Geomática - UFPR. Curitiba-PR. 210 p. 2003.

COLSTOUN, E. C. B., STORY, M.H., THOMPSON, C., COMMISSO, K., SMITH, T.G., IRONS, J.R. National Park vegetation mapping using multitemporal LANDSAT 7 data and a decision tree classifier. **Remote Sensing of Environment**. v.85, p. 316-327, 2003.

CONGALTON, R.G., GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. CRC Press, 2^a ed. 183 p. 2009.

CROSTA, A.P. **Processamento digital de imagens**. Campinas: IG/ UNICAMP, 170 p. 1992.

DISPERATI, A. A., OLIVEIRA FILHO, P.C. Delineamento de copas de árvores em fotografias aéreas de alta resolução, através de suas sombras periféricas: estudo de caso na Floresta Ombrófila Mista. **Revista Árvore**. Viçosa-MG. v. 29, n. 2, p. 195-202, 2005.

DLUGOSZ F.L., ROSOT M.A.D., ROSOT N.C., OLIVEIRA Y.M.M. Avaliação de tipologias da floresta ombrófila mista por meio de processamento digital de imagens de alta resolução. **Anais do III Evento de Iniciação Científica da Embrapa Florestas**. Colombo-PR. Cd-Rom, 2004.

DLUGOSZ, F. **Classificação orientada a regiões na discriminação de tipologias da Floresta Ombrófila Mista usando imagens orbitais Ikonos**. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Paraná-UFPR, Curitiba, PR. 135 p. 2005.

DLUGOSZ, F., ROSOT, N., ROSOT, A., OLIVEIRA, Y.M.M. Uso da segmentação por crescimento de regiões em imagens Ikonos na discriminação de tipologias de Floresta Ombrófila Mista. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**. Goiânia, p. 1493 – 1500. 2005.

DLUGOSZ, F., ROSOT, N., IURK, M.C. Avaliação quantitativa da segmentação por crescimento de regiões para a discriminação de tipologias da Floresta Ombrófila Mista em

imagens Ikonos. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, Brasil, p.1639 - 1646. 2007.

DUTRA, G. C., DAVIDE, A.C., CARVALHO, L.M.T., BOTELHO, S.A. Modelo de estratificação de áreas visando à recuperação da área no entorno da Usina Hidrelétrica do Funil (Rio Grande) - MG. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, Brasil, p. 2533-2539. 2007.

EASTMAN, J. R. **IDRISI for Windows: Manual do Usuário**. Porto Alegre: UFRS, 1998.

eCognition (Object Oriented Image Analysis). **User Guide 3**. Definiens Imaging. 2000.

FIGUEIREDO, S. M. M., CARVALHO, L.M.T. Avaliação da exatidão do mapeamento da cobertura da terra em Capixaba, Acre, utilizando classificação por árvore de decisão. **Cerne**. v.12, n. 001, p. 38-47, 2006.

FRANKLIN, S. E. PEDDLE, D. R. 1990. Classification of SPOT HRV imagery and texture features . **International Journal Remote Sensing**. v. 11, n. 3, p. 551-556, 1990.

FRIEDL, M. A., BRODLEY, C.E. Decision Tree Classification of Land Cover from Remotely Sensed Data. **Remote Sensing of Environment**. v.61, n. 3, p. 399-409, 1997.

FUCKNER, M. A., ROSA, V.G.C., ROSEMBACK, R. Análise comparativa da aplicação de transformações radiométricas e técnicas de fusão de imagens em área urbana. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**. Florianópolis. p. 6343-6350, 2007

FUCKNER, M.A.; ROSEMBACK, R.; ROSA, V.G.C. **Comparação de transformações radiométricas e técnicas de fusão de imagens CBERS-2/CCD e IRS-P6/LISS-IV: aplicação no município de Campinas – SP**. Monografia (para a disciplina de Processamento Digital de Imagens, curso de Pós-Graduação em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais-INPE, São José dos Campos-SP. 13p. 2005.

GANTZEL, O. L. **Avaliação das florestas de Araucaria angustifolia (Bert.) O. Ktze do Sul do Brasil, através de imagens do satélite Landsat II**. Dissertação (Mestrado em Ciências Florestais). Universidade Federal do Paraná-UFPR, Curitiba-PR. 105 p. 1979.

GONZALES, R. C., WOODS, R.E. **Digital Image Processing**. 2ª ed. Prentice Hall. New Jersey, 793 p., 2002.

GUALBERTO, L.P.S.; SOUZA FILHO, P.W.M. Reconhecimento de Ambientes Costeiros a partir de Análise Textural de Imagens SAR R99B (Banda L, Polarização VV). **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**. Florianópolis. p. 4889-4896, 2007.

GUERRA, M. P.; SILVEIRA, V.; REIS, M. S. dos; SCHNEIDER, L. Exploração, manejo e conservação da araucária (*Araucaria angustifolia*). In: SIMÕES, L.L.; LINO, C.F. (Ed.). **Sustentável Mata Atlântica: a exploração de seus recursos florestais**. São Paulo: SENAC, p. 85-101, 2000.

GUIDINI, M.P., VASCONCELOS, M.A.C. **Aplicação de Data Mining em Grandes Bases de Dados Geográficos através do uso da Biblioteca TerraLib.** Disponível em: <http://www.dpi.inpe.br/cursos/ser303/mauricio_guidini_data_mining.doc>. Acesso em 15 de jul. 2008.

HAN, J., KAMBER, M. **Data Mining: Concepts and Techniques:** Morgan Kaufmann Publishers. 743 p. 2006.

HARALICK, R. M., SHANMUGAM, K., DINSTEN, I.H. Textural Features for Image Classification. **Ieee Transactions On Systems, Man And Cybernetics.** v. smc-3, n. 6, p. 610-621, 1973.

HENDGES, E. R., ROSADO, R.C., PERREIRA, R.S. Correção atmosférica: uso de técnicas de processamento de imagens digitais na comparação de mapas temáticos. **Anais do III Colóquio Brasileiro de Ciências Geodésicas**, p.13. 2003.

HOWARD, J. A. **Remote Sensing of Forest Resources – Theory and Application.** London: Ed.Chapman & Hall. 436 p. 1991.

IBAMA. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis. **Portaria nº 37-N, de 03 de abril de 1992.** Lista Oficial de Espécies da Flores Brasileira Ameaçada de Extinção. 1992. 2p.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Manual Técnico de Uso da Terra.** 2006. 2ª edição. Disponível em: <<ftp://geofp.ibge.gov.br/documentos/recursosnaturais/usodaterra/manualusodaterra.pdf>>. Acesso em 08 de maio de 2009.

JAXA. **ALOS User Handbook.** Earth Observation Research Center, Japan Aerospace Exploration Agency. 2007. Disponível em: <<http://www.eorc.jaxa.jp/ALOS/en/doc/handbk.htm>>. Acesso em 10 de abr. 2008.

JAXA. **The 2nd ALOS Research Announcement:** Calibration and Validation, Utilization Research, and Scientific Research. Earth Observation Research Center, Japan Aerospace Exploration Agency. 2006a. Disponível em: <http://www.eorc.jaxa.jp/ALOS/en/ra/ra2_guide.htm>. Acesso em 10 de abr. 2008.

JAXA. **ALOS Product Format Description.** 2006b. Disponível em: <<http://www.eorc.jaxa.jp/ALOS/en/doc/format.htm>>. Acesso em 10 de abr. 2008.

JOHNSON, R. A., WICHERN, D.W. **Applied Multivariate Statistical Analysis.** 816 p. 1998.

KELLERMANN, B., ROSOT, M.A.D., OLIVEIRA, Y.M.M., TERRA, G. Avaliação da Regeneração Natural em um Fragmento de Floresta Ombrófila Mista – Caçador, SC. **Anais do VI Evento de Iniciação Científica da Embrapa Florestas.** Colombo-PR, Cd-Rom, 2007.

- KLEIN, R. M. O aspecto dinâmico do pinheiro-brasileiro. **Sellowia**. v.12, n.12: 17-48 p. 1960.
- KOCH, Z., CORRÊA, M.C. **Araucária: a floresta do Brasil Meridional**. Curitiba: Olhar Brasileiro. 148 p. 2002.
- KOZERA, C., DITTRICH, V., SILVA, S. Fitossociologia do componente arbóreo de um fragmento de Floresta Ombrófila Mista Montana, Curitiba, PR, BR. **Revista Floresta**, Curitiba, v. 36, n.2, p. 225-237, 2006.
- KURASZ, G. **Sistema de Informações Geográficas aplicado ao Zoneamento Ambiental da Reserva Florestal Embrapa/Epagri, Caçador-SC**. Dissertação (Mestrado em Ciências Florestais). Universidade Federal do Paraná-UFPR, Curitiba-PR. 137 p. 2005.
- LAMPRECHT, H. Ensayo sobre la estructura florística de la parte sur-oriental del bosque universitario “El Caimital”-Estado Barinas. **Revista Forestal Venezolana**, Mérida, v.7, n.10-11, p. 77-119, 1964.
- LANDIS, J., KOCH, G. G. The measurements of agreement for categorical data. **Biometrics**. v.33, n. 3, p. 159-179, 1977.
- LATORRE, M. L. **Desenvolvimento metodológico para a integração de dados de sensoriamento remoto multi resoluções para a representação da cobertura da terra em campos contínuos de vegetação e classificação por árvores de decisão**. Tese (Doutorado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais-INPE, São José dos Campos-SP. 236 p. 2007.
- LATORRE, M., CARVALHO JÚNIOR, O.A., CARVALHO, A.P.F., SHIMABUKURO, Y.E. Correção atmosférica: conceitos e fundamentos. **Espaço & Geografia**. v. 5, n. 1, p. 153-178, 2002.
- LILLESAND, T. M., KIEFER, R.W., CHIPMAN, J.W. **Remote Sensing and Image Interpretation**: John Wilwy & Sons, Inc. 756 p.2007.
- LIMA, E.C., SANQUETTA, C.R., KIRCHNER, F.F., FERRETTI, E.R., Qualidade da paisagem: estudo de caso na Floresta Ombrófila Mista. **Revista Floresta**, v. 34, n. 1, p. 45-56, 2004.
- LOBO, R. S., KALIL, R.C.E. **Um estudo de caso do impacto da imputação de dados Na mineração de dados**. Trabalho de Conclusão de Curso. (Curso de graduação em Ciência da Computação), Universidade Federal do Pará - UFPa, Belém-PA, 103 p. 2006.
- LOCK, R. E. N., KIRCHNER, F.F. Classificação de imagens multiespectrais LANDSAT TM e feições de textura: mapeamento da cobertura da terra. **Revista Floresta**. v.27, n. 1/2, p. 41-58, 1997.
- LONGHI, S.J. **Agrupamento e análise fitossociológica de comunidades florestais na sub-bacia hidrográfica do rio Passo Fundo-RS**. Tese (Mestrado em Engenharia Florestal), Universidade Federal do Paraná – UFPR. Curitiba-PR. 198 p. 1997.

MAACK, R. Geografia física do Estado do Paraná. Rio de Janeiro: Livraria José Olimpo. 450 p. 1981.

MARCELINO, E.V. **Mapeamento de áreas susceptíveis a escorregamento no município de Caraguatatuba (SP) usando técnicas de sensoriamento remoto.** Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais – INPE. São José dos Campos, SP. 228 p. 2004.

MARTINS, F. T., DUTRA, L.V., FREITAS, C.C., FONSECA, F.R., GUIMARÃES, R.J.P.S., MOURA, A.C.M., SCHOLTE, R.G.C., AMARAL, R.S., DRUMMOND, S.C., FREITAS, C.R., CARVALHO, O.S. Uso de árvore de decisão para predição da prevalência de esquistossomose no Estado de Minas Gerais, Brasil. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, Brasil, 21-26 abril, p.2841-2848, 2007.

MATHER, P.M. **Computer processing of remotely-sensed images: an introduction.** 2 ed. Great Britain: John Wiley, 291p. 1999.

MENESES, P. R. Fundamentos de Radiometria Óptica espectral. In: (Ed.). Mesense, P.R., Madeira Netto, J.S. (org) - **Sensoriamento Remoto: reflectância dos alvos naturais.** Brasília-DF: Universidade de Brasília-UnB, Embrapa Cerrados, p.15-40, 2001.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte-MG: Editora UFMG. 295 p. 2005.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e Metodologias de Aplicação.** São José dos Campos-SP: Instituto Nacional de Pesquisas Espaciais-INPE. 250 p. 2001.

NASCIMENTO, P. S. D. R. **Avaliação de técnicas de segmentação e classificação por regiões em imagens Landsat-TM visando o mapeamento de unidades de paisagem na Amazônia.** Dissertação (Mestrado em Sensoriamento Remoto). Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos – SP. 102 p. 1997.

NEGRELLE, R. SILVA, F. Fitossociologia de um trecho de floresta com *Araucaria angustifolia* (Bert.) O. Ktze. no município de Caçador – SC. **Boletim de Pesquisa Florestal**, n. 24/25, p. 37-54, 1992.

NETTO, J.D.S.M. Comportamento Espectral dos Solos. In: (Ed.). Mesense, P.R., Madeira Netto, J.S. (org) - **Sensoriamento Remoto: reflectância dos alvos naturais.** Brasília-DF: Universidade de Brasília-UnB, Embrapa Cerrados, p.127-154, 2001.

NOVO, E.M., COSTA, M.P., MANTOVANI, J.E. Avaliação do uso de medidas de textura na discriminação de ambientes aquáticos do reservatório da UHE Tucuruí a partir de dados RADARSAT. **Anais do IX Simpósio Brasileiro de Sensoriamento Remoto**, Santos, Brasil, p.155-165, 1998.

NUÑEZ, J.R. **Segmentação automática e classificação em vídeos esportivos**. Dissertação (Mestrado em Informática Aplicada) – Pontifícia Universidade Católica do Paraná – PUCPR. Curitiba-PR. 178 p. 2006.

OLIVEIRA, L.T. **Fusão de imagens de sensoriamento remoto e mineração de dados geográficos para mapear as fitofisionomias do bioma cerrado**. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras – UFLA. Lavras-MG. 131 p. 2004.

OLIVEIRA, Y.M.M. **Investigation of remote sensing for assessing and monitoring the Araucaria forest region of Brazil**. Tese (Doutorado em Ciências Florestais). University of Oxford, Oxford, UK. 247 p. 1999.

OLIVEIRA, Y.M.; ROSOT, A.D.; RIVERA, H. Aplicação de geotecnologias na elaboração de planos de manejo de áreas protegidas: O caso da Reserva Florestal Embrapa/Epagri. In: **Seminário de Geotecnologias Aplicadas à Engenharia Florestal**. Cd-Rom. 2006. 1

OLIVEIRA, Y.M.; ROTTA, E. Levantamento da estrutura horizontal de uma mata de araucária no primeiro planalto paranaense. **Boletim de Pesquisa Florestal**, n. 4, p. 1-46, 1982.

PAL, M., MATHER, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. **Remote Sensing of Environment**. v. 86, p. 554-565, 2003.

PEDDLE, D.R., FRANKLIN, S. E. Image Texture processing and data integration for surface pattern discrimination. **Photogrammetric Engineering & Remote Sensing**. v. 57, n. 4, p. 413-420, abr. 1991.

PINHO, C.M.D.; RENNÓ, C.D.; KUX, H.J.H. Avaliação de técnicas de fusão aplicadas à imagem Quickbird. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**, Goiânia, Brasil, p.4225-4232, 2005.

PONZONI, F. J. Comportamento Espectral da Vegetação. In: (Ed.). Mesense, P.R., Madeira Netto, J.S. (org) - **Sensoriamento Remoto: reflectância dos alvos naturais**. Brasília-DF: Universidade de Brasília-UnB, Embrapa Cerrados, p.156-199, 2001.

PONZONI, F. J., SHIMABUKURO, Y.E. **Sensoriamento Remoto no estudo da Vegetação**. São José dos Campos-SP: Parêntese. 127 p. 2007.

REITZ, R., KLEIN, R., REIS, A. **Projeto madeira de Santa Catarina**. Convênio SUDESUL/Governo do Estado de Santa Catarina/IBDF. 320 p. 1978.

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. **Anais do VIII Simpósio Brasileiro de Sensoriamento Remoto**. Salvador, Brasil., INPE, p. 863-869, 1996.

RIVERA, H. **Ordenamento territorial de áreas florestais utilizando avaliação multicritério apoiada Por geoprocessamento, fitossociologia e análise multivariada**. Dissertação (Mestrado em Ciências Florestais). Universidade Federal do Paraná - UFPR, Curitiba-PR. 225 p. 2007.

ROBINSON, J. M., MORRISON, J., MUEHRCKE, P., KIMERLING, A., GUPRILL, S. **Elements of Cartography**. New York: John Wiley & Sons. 674 p. 1995.

RODRÍGUEZ, A.C.M.; HERNANDEZ FILHO, P.; MOREIRA, M.A. Mapeamento multitemporal do uso e cobertura do solo do município de São Sebastião - SP, utilizando técnicas de segmentação e classificação de imagens TM - Landsat e HRV – Spot. **Anais do GISBRASIL** Salvador, Cd-Rom, 2000.

RONDON NETO, R.; KOZERA, C.; ANDRADE, R.; CECY, A.; HUMMES, A. FRITZSONS, E.; CALDEIRA, M.; MACIEL, M.; SOUZA, M. Caracterização florística e estrutural de um fragmento de Floresta Ombrófila Mista, em Curitiba, PR – Brasil. **Revista Floresta**. Curitiba, v. 32, n. 1, p. 3-16, 2002.

ROSOT, M.A., CRUZ, P., OLIVEIRA, Y.M.M., RIVERA, H., MATTOS, P. . Desarrollo de un modelo de plan de manejo para áreas protegidas en bosques con araucaria en el sur de Brasil. . **Anais do 2º Congresso Latinoamericano IUFRO**. p.110-111, 2006.

ROSOT, M.A.D.; RIVERA, H.; OLIVEIRA, Y.M.M. de; DLUGOSZ, F.L.; ROSOT, N.C. Discriminating sub-typologies in Araucarian forests in Brazil: a comparison between remote sensing and phytosociological approaches. **International Symposium on Remote Sensing of Environment**, San Jose, n. 32, 2007.

ROSOT, N. C. **Integração de imagens de sensores de microondas e ópticos para fins de mapeamento e classificação de reflorestamentos no sul do Brasil**. Tese (Doutorado em Engenharia de Produção). Universidade Federal de Santa Catarina-UFSC, Florianópolis-SC, 190 p. 2001.

RSI (Research Systems, Inc). **ENVI User's Guide**. ENVI© Version 4.2 August, 2005.

SAMPAIO, C. D. S. **Abordagem metodológica baseada nos dados multitemporais modis EVI/NDVI para classificação da cobertura vegetal na região do Parque nacional da Chapada dos Veadeiros/GO**. Universidade de Brasília-UnB, Brasília-DF, 73 p. 2007.

SANQUETTA, C. R. Controle de taquaras como alternativa para a recuperação da Floresta de Araucária. **Pesquisa Florestal Brasileira**. n. 55, p. 45-53, 2007.

SCHNEIDER, M.J.; BELLON, O.R.P.; ARAKI, H. Experimentos em fusão de imagens de alta resolução. **Bol. Ciênc. Geod**. Curitiba v. 9, n. 1, p. 75-88, 2003.

SCHOWENGERDT, R. A. **Techniques for Image Processing and Classification in Remote Sensing**. University of Arizona. 249p. 1983.

SCHOWENGERDT, R.A. **Remote Sensing: Models and Methods for Image Processing**. Academic Press Inc, 3rd edition, 515p. 2007.

SCHWARTZ, W. R., PEDRINI, H. Segmentação de imagens de terrenos baseada na associação de características de texturas com dependência espacial modelada por campo aleatório de Markov. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**, Goiânia, p.4311-4318, 2005.

SHIBA, M.H., SANTOS, R.L., QUINTANILHA, J.A., KIM, H.Y. . Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**, Goiânia, p.4319-4326, 2005.

SILVA, F., MARCONI, L. Fitossociologia de uma Floresta de Araucária em Colombo-PR. **Boletim de Pesquisa Florestal**, n. 20, p. 23-28, 1990.

SILVA, G. F. N., TAVARES JÚNIOR, S.S., EVANGELISTA, R.A.O., OLIVEIRA JÚNIOR, M.C.M. . Integração digital e análise espacial aplicados ao estudo da suscetibilidade incêndios florestais em Apiaú e Ribeiro Campos - Roraima. **Revista Floresta**, Curitiba-PR. v. 38, n. 4, p. 683-697, 2008.

SILVA, M. P. S. **Mineração de Dados : Conceitos, Aplicações e Experimentos com Weka. Livro da Escola Regional de Informática Rio de Janeiro - Espírito Santo**. Porto Alegre: Sociedade Brasileira de Computação, v. 1, p. 1-20, 2004

SPRING – Manuais. **Tutorial de Geoprocessamento**. Disponível em: <http://www.dpi.inpe.br/spring/portugues/tutorial/registro_img.html>. Acesso em: 25 de nov. 2008a.

SPRING. **Sistema de ajuda on-line**. Disponível em: <file:///C:/Arquivos%20de%20programas/spring501_Port/helpport/c_clapix.htm?versao=nclassica#intro>. Acesso em: 12 de out. 2008b.

SPRING. **Sistema de ajuda on-line**. Disponível em: <file:///C:/Arquivos%20de%20programas/spring501_Port/helpport/registro.htm?versao=nclassic#transf_pc>. Acesso em: 12 de out. 2008c.

SULSOFT Serviços de Processamento de Dados Ltda. 2007. **Guia do ENVI em Português**. Disponível em: <<http://www.sulsoft.com.br/>>. Acesso em: 17 de dez. 2008.

TELLES JR., M.A.B.G., ROSA, A.N.C.S. Fusão de imagens IRS-1C e Landsat 5 utilizando método aditivo de sub-bandas. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**, Goiânia, p. 4327-4334, 2005.

TERRA, G., SILVA MATOS, D. M., ROSOT, M.A.D., OLIVEIRA, Y.M.M. . Considerações sobre a FRAGMENTAÇÃO da Floresta Ombrófila Mista e o Ciclo de Vida das Taquaras. **Seminário sobre Recuperação de Áreas Degradadas em Floresta de Araucária**. Curitiba-PR 2006.

THOMÉ, N. **Ciclo da madeira: história da devastação da Floresta de Araucária e do desenvolvimento da indústria madeireira em Caçador e na região do Contestado no século XX**. Caçador: Universal. 1995. 206 p.

TSO, B., MATHER, P.M. **Classification Methods for Remotely Sensed Data**. New York: Taylor & Francis. 332 p, 2001.

USÓ, A. M. **Unsupervised band selection and segmentation in hiper/multiespectral images**. (Ph.D). Departament de Llenguatges i Sistemes Informàtics, Universitat Jaume I, Castellón, 224 p. 2008.

VASCONCELOS, C.N., NOVO, E.M.L.M. Mapeamento do uso e cobertura da terra a partir da segmentação e classificação de imagens – fração solo, sombra e vegetação derivadas do modelo linear de mistura aplicado a dados do sensor TM/Landsat5, na região do reservatório de Tucuruí - PA. **Acta Amazonica**. v. 34, n. 3, p. 487-493, 2004.

VENTURIERI, A., SANTOS, J.R. Técnicas de Classificação de Imagens para Análise de Cobertura Vegetal. In: (Ed.). ASSAD, E.D. e SANO, E.E. **Sistema de Informações Geográficas**. Brasília: Embrapa-SPI/Embrapa-CPAC. Capítulo 18, p. 351-371, 1998.

VERONA, J. D., SHIMABUKURO, Y.E. Classificação da cobertura vegetal da região norte da FLONA do Tapajós utilizando a segmentação de imagens-fração. **Anais do XII Simpósio Brasileiro de Sensoriamento Remoto**, Goiânia, p. 1685-1689, 2005.

VIEIRA, I.C.G., ALMEIDA, A.S., DAVIDSON, E.A., STONE, T.A., CARVALHO, C.J.R., GUERRERO, J.B. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. **Remote Sensing of Environment**. v. 87, n. 4, p. 470-481, November 2003.

WALD, L. Some terms of reference in data fusion. **IEEE Transaction on Geoscience and Remote Sensing**. v. 37, n. 3, p.1190-1193, 1999.

WATZLAWICK, L. F. **Estimativa de biomassa e carbono em Floresta Ombrófila Mista e plantações florestais a partir de dados de imagens do satélite Ikonos II**. Tese (Doutorado em Ciências Florestais) Universidade Federal do Paraná-UFPR, Curitiba-PR. 119 p. 2003.

WITTEN, I. H. e E. FRANK. **Data Mining: Practical Machine Learning Tools and Techniques**: Morgan Kaufmann Publishers. 525 p. 2005.

XU, M., WATANACHATURAPORN, P., VARSHNEY, P.K., ARORA, M.K. Decision tree regression for soft classification of remote sensing data. **Remote Sensing of Environment**. v. 97, p. 322-336, 2005.

YI, J. L. R., AULICINO, L.C.M.A., SHIMABUKURO, Y.E., RUDORFF, B.F., DUARTE, V., MOREIRA, M.A., MARTINI, P.R., SOUZA, I.M. . Segmentação de imagens-fração derivadas do sensor TM-Landsat para o mapeamento do uso do solo no município de Sapezal (MT). **Anais do IX Simpósio Brasileiro de Sensoriamento Remoto**, Santos, p. 71-76, 1998.

ZAGAGLIA, C. R., CARVALHO, E.V.T., SANTOS, R.N.A. Mapeamento de Manguezais do Brasil: Subsídio ao Projeto GEF Mangue. **Anais do XIII Simpósio Brasileiro de Sensoriamento Remoto**, Florianópolis, p.4343-4350, 2007.

APÊNDICES

APÊNDICE A - Cabeçalho do arquivo ARFF

```

@RELATION 'VARIABLES_ESPECTRAIS_ALOS'
@attribute ID_Amostras {Agricultura_2_1,Agricultura_2_2,Agricultura_2_3,...}
@attribute B1 real
@attribute B2 real
@attribute B3 real
@attribute B4 real
@attribute PC_1 real
@attribute PC_2 real
@attribute PC_3 real
@attribute PC_4 real
@attribute NDVI real
@attribute SAVI_1 real
@attribute SAVI_05 real
@attribute SAVI_025 real
@attribute RVI real
@attribute Oc_DR_B1 real
@attribute Oc_mean_B1 real
@attribute Oc_var_B1 real
@attribute Oc_entropy_B1 real
@attribute Oc_DR_B2 real
@attribute Oc_mean_B2 real
@attribute Oc_var_B2 real
@attribute Oc_entropy_B2 real
@attribute Oc_DR_B3 real
@attribute Oc_mean_B3 real
@attribute Oc_var_B3 real
@attribute Oc_entropy_B3 real
@attribute Oc_DR_B4 real
@attribute Oc_mean_B4 real
@attribute Oc_var_B4 real
@attribute Oc_entropy_B4 real
@attribute Cooc_mean_B1 real
@attribute Cooc_var_B1 real
@attribute Cooc_homog_B1 real
@attribute Cooc_contrast_B1 real
@attribute Cooc_dissim_B1 real
@attribute Cooc_entropy_B1 real
@attribute Cooc_2_moment_B1 real
@attribute Cooc_correl_B1 real
@attribute Cooc_mean_B2 real
@attribute Cooc_var_B2 real
@attribute Cooc_homog_B2 real
@attribute Cooc_contrast_B2 real
@attribute Cooc_dissim_B2 real

```

```
@attribute Cooc_entropy_B2 real
@attribute Cooc_2_moment_B2 real
@attribute Cooc_correl_B2 real
@attribute Cooc_mean_B3 real
@attribute Cooc_var_B3 real
@attribute Cooc_homog_B3 real
@attribute Cooc_contrast_B3 real
@attribute Cooc_dissim_B3 real
@attribute Cooc_entropy_B3 real
@attribute Cooc_2_moment_B3 real
@attribute Cooc_correl_B3 real
@attribute Cooc_mean_B4 real
@attribute Cooc_var_B4 real
@attribute Cooc_homog_B4 real
@attribute Cooc_contrast_B4 real
@attribute Cooc_dissim_B4 real
@attribute Cooc_entropy_B4 real
@attribute Cooc_2_moment_B4 real
@attribute Cooc_correl_B4 real
@attribute Classe
{Agricultura,Agua,Araucaria,Floresta_com_taq,Fruticultura1,Fruticultura2,Nuvem,Sem_arau
c,Sombra,Taquara}
@data
```

APÊNDICE B - Árvore de decisão originada no WEKA

```

Cooc_mean_B2 <= 16.1111
| Oc_mean_B1 <= 70.7778
| | Cooc_mean_B3 <= 7.4444: Sombra (252.0)
| | Cooc_mean_B3 > 7.4444
| | | Cooc_var_B4 <= 13.4321
| | | | Oc_mean_B1 <= 70.5556: Araucaria (433.0)
| | | | Oc_mean_B1 > 70.5556
| | | | | Cooc_mean_B1 <= 18.8889: Floresta_com_taq (2.0)
| | | | | Cooc_mean_B1 > 18.8889: Araucaria (12.0)
| | | | Cooc_var_B4 > 13.4321: Floresta_com_taq (5.0/1.0)
| | Oc_mean_B1 > 70.7778
| | | Cooc_mean_B4 <= 35.7778
| | | | Cooc_mean_B2 <= 12.7778
| | | | | Oc_DR_B2 <= 6: Araucaria (4.0)
| | | | | Oc_DR_B2 > 6: Floresta_com_taq (4.0)
| | | | | Cooc_mean_B2 > 12.7778: Floresta_com_taq (462.0)
| | | Cooc_mean_B4 > 35.7778
| | | | Cooc_dissim_B4 <= 4.5556
| | | | | Oc_mean_B4 <= 67.2222
| | | | | | Cooc_mean_B2 <= 14.5556
| | | | | | | Oc_mean_B2 <= 49.4444: Floresta_com_taq (2.0)
| | | | | | | Oc_mean_B2 > 49.4444
| | | | | | | | Oc_DR_B4 <= 17: Sem_arauc (33.0)
| | | | | | | | Oc_DR_B4 > 17
| | | | | | | | | B4 <= 65: Floresta_com_taq (2.0)
| | | | | | | | | B4 > 65: Sem_arauc (3.0)
| | | | | | | | Cooc_mean_B2 > 14.5556: Floresta_com_taq (6.0)
| | | | | | | Oc_mean_B4 > 67.2222: Sem_arauc (399.0)
| | | | | Cooc_dissim_B4 > 4.5556
| | | | | | Cooc_contrast_B3 <= 0.7778: Sem_arauc (14.0)
| | | | | | Cooc_contrast_B3 > 0.7778: Floresta_com_taq (14.0/1.0)
Cooc_mean_B2 > 16.1111
| Cooc_mean_B4 <= 34.8889
| | PC_2 <= 35.6389
| | | NDVI <= 97.6068: Fruticultural1 (30.0/1.0)
| | | NDVI > 97.6068
| | | | Oc_DR_B4 <= 27
| | | | | Cooc_var_B3 <= 15.1358: Taquara (450.0/1.0)
| | | | | Cooc_var_B3 > 15.1358: Agua (5.0/1.0)
| | | | | Oc_DR_B4 > 27
| | | | | | B2 <= 60: Agua (2.0)
| | | | | | B2 > 60: Fruticultural1 (7.0)
| | | PC_2 > 35.6389: Agua (82.0)
| Cooc_mean_B4 > 34.8889

```



```
| | Oc_mean_B1 <= 118.5556
| | | Oc_mean_B3 <= 60: Fruticultura2 (35.0)
| | | Oc_mean_B3 > 60
| | | | PC_3 <= 12.2333: Nuvem (7.0)
| | | | PC_3 > 12.2333: Agricultura (361.0/1.0)
| | Oc_mean_B1 > 118.5556: Nuvem (227.0)
```

Number of Leaves : 26

Size of the tree : 51

APÊNDICE C - Imagens segmentadas e seus respectivos limiares de similaridade e área

