



UNIVERSIDADE DE BRASÍLIA - UnB  
INSTITUTO DE GEOCIÊNCIAS – IG  
PÓS-GRADUAÇÃO EM GEOCIÊNCIAS APLICADAS E GEODINÂMICA

**AVALIAÇÃO DE IMAGENS POLARIMÉTRICAS DO SENSOR ALOS-  
2/PALSAR-2 E DE TÉCNICAS DE MINERAÇÃO DE DADOS NA  
CLASSIFICAÇÃO DE USO E COBERTURA DE TERRAS DO  
CERRADO**

Flávio Fortes Camargo

Tese de Doutorado nº 42

Brasília - DF

Dezembro de 2018



UNIVERSIDADE DE BRASÍLIA - UnB  
INSTITUTO DE GEOCIÊNCIAS – IG  
PÓS-GRADUAÇÃO EM GEOCIÊNCIAS APLICADAS E GEODINÂMICA

**Avaliação de imagens polarimétricas do sensor ALOS-2/PALSAR-2 e de técnicas de mineração de dados na classificação de uso e cobertura de terras do Cerrado**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Geociências Aplicadas da Universidade de Brasília, como requisito para obtenção do título de Doutor.

Área de concentração: *Geoprocessamento Aplicado à Análise Ambiental*

Flávio Fortes Camargo  
Orientador: Edson Eyji Sano

Brasília – DF  
Dezembro de 2018

## **BANCA EXAMINADORA**

---

Edson Eyji Sano  
Embrapa Cerrados - Orientador

---

Tati de Almeida  
Universidade de Brasília – Examinador Interno

---

Cláudia Maria de Almeida  
Instituto Nacional de Pesquisas Espaciais – Examinador Externo

---

José Cláudio Mura  
Instituto Nacional de Pesquisas Espaciais – Examinador Externo

Brasília – DF  
Dezembro de 2018

## **FICHA CATALOGRÁFICA**

Camargo, Flávio F.  
Avaliação de imagens polarimétricas do sensor ALOS-2/PALSAR-2 e de técnicas de mineração de dados na classificação de uso e cobertura de terras do Cerrado. / Flávio Fortes Camargo; orientação de Edson Eyji Sano. Brasília, 2018  
195p.  
Tese de Doutorado (D) – Universidade de Brasília / Instituto de Geociências, 2018.  
1. ALOS-2/PALSAR-2; 2. polarimetria SAR; 3. mineração de dados; 4. uso e cobertura de terras; 5. cerrado.

## **REFERÊNCIA BIBLIOGRÁFICA**

Camargo, Flávio Fortes. Avaliação de imagens polarimétricas do sensor ALOS-2/PALSAR-2 e de técnicas de mineração de dados na classificação de uso e cobertura de terras do Cerrado. Brasília, Instituto de Geociências, Universidade de Brasília, 2018, 195p. Tese de Doutorado.

## **CESSÃO DE DIREITOS**

NOME DO AUTOR: Flávio Fortes Camargo

TÍTULO DA TESE DE DOUTORADO: Avaliação de imagens polarimétricas do sensor ALOS-2/PALSAR-2 e de técnicas de mineração de dados na classificação de uso e cobertura de terras do Cerrado.

GRAU: Doutor ANO: 2018

É concedida à Universidade de Brasília permissão para reproduzir cópias desta tese de doutorado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Esse direito poderá ocorrer somente após a publicação dos artigos contidos no documento. O autor reserva-se a outros direitos de publicação e nenhuma parte desta tese de doutorado pode ser reproduzida sem a autorização por escrito do autor.

---

Flávio Fortes Camargo  
flavio.fortes.camargo@gmail.com

*“(...) sail on, sing a song, carry on  
‘Cause We Rock...”*  
– Ronnie James Dio

-----  
*“Yo vengo de todas partes, Y hacia todas partes voy:  
Arte soy entre las artes, En los montes, monte soy.”*  
– Jose Marti

Dedico aos meus pais (*in memoriam*).

## AGRADECIMENTOS

Ao Programa de Pós-Graduação em Geociências Aplicadas e Geodinâmica do Instituto de Geociências da Universidade de Brasília, pelo apoio e suporte necessários ao desenvolvimento desta tese. Agradecimento especial à banca de qualificação pelas sugestões de melhorias, as quais busquei seguir à risca no desenvolvimento da pesquisa. Agradeço também ao Dr. Ricardo Brites, em particular, que sempre se dispôs a sanar dúvidas e dar sugestões, mesmo que por *e-mail*. Também agradeço ao Dr. Gustavo Baptista por ter me orientado no primeiro semestre da pós-graduação.

Ao meu orientador Dr. Edson Eyji Sano, pela orientação e suporte incondicionais aos experimentos realizados no âmbito desta tese e também em experimentos correlatos, não tratados neste documento, todavia importantes também e submetidos a simpósios (<https://ieeexplore.ieee.org/document/7325979/>). Destaco como suporte imprescindível a cessão de imagens ALOS-2/PALSAR-2 por meio do programa *Kyoto & Carbon Protocol* da Agência de Exploração Aeroespacial do Japão (JAXA).

Ao Departamento Nacional de Infraestrutura de Transportes (DNIT), nas figuras do seu Diretor Geral e do Superintendente Regional do Estado de São Paulo, pela importante concessão de afastamento para realização de pós-graduação *stricto sensu* no país (nos moldes do Artigo 96A da Lei Federal 8.112/1990).

Ao Dr. José Cláudio Mura e sua equipe do INPE, pela receptividade, ajuda e orientações gerais sobre polarimetria SAR, em especial nos tópicos relativos ao processamento dos dados complexos. Agradecimento especial às aulas da disciplina “Processamento de Imagens SAR”, aulas estas que presenciei como ouvinte no segundo semestre de 2016.

À Dra. Cláudia Maria de Almeida do INPE, também pela receptividade, ajuda e sugestões relativas aos algoritmos de aprendizado de máquina para fins de classificação de imagens de sensoriamento remoto, em especial aos classificadores do aplicativo WEKA.

Ao Dr. Camilo Daleles Rennó, também do INPE, pela cessão da planilha para realização dos testes de hipóteses, bem como pelas sugestões sobre amostragem e validação.

Aos colegas de pós-graduação do INPE, em especial, aos doutores Arnaldo Queiroz e Cleber Oliveira e, também, ao MSc. Thiago Rodrigues, com os quais conversei bastante sobre polarimetria e ortorretificação de imagens SAR.

À minha querida namorada e companheira Iana Belli, que me apoia incondicionalmente nos meus desafios acadêmicos e profissionais. Agradeço a ela por toda a paciência ao longo destes três anos que estamos juntos.

## RESUMO

O Cerrado ocupa uma área de aproximadamente 2.000.000 km<sup>2</sup> (23% do território brasileiro), concentrando-se principalmente no Planalto Central. A sua composição fitofisionômica abrange formações florestais, savânicas e campestres, as quais vêm sofrendo mudanças severas com a introdução da agricultura extensiva e mecanizada de grãos para exportação. Tecnologias de sensoriamento remoto têm sido empregadas para monitoramento da cobertura vegetal do Cerrado, utilizando principalmente imagens ópticas de sensores orbitais (casos de iniciativas como TerraClass e MapBiomas). Apesar dos avanços metodológicos e tecnológicos, muito ainda se tem a explorar dos sensores de radar de abertura sintética (utiliza-se o acrônimo em inglês SAR - *synthetic aperture radar*) orbitais para fins de mapeamento do uso e cobertura do Cerrado e discriminação de suas fitofisionomias. Nesta tese, foram realizados dois experimentos de classificação de imagens SAR do sensor ALOS-2/PALSAR-2 em duas áreas de Cerrado. O primeiro experimento (artigo 1) foi realizado em área de estudo de 356 km<sup>2</sup> localizada na porção norte do Distrito Federal (DF). A abordagem metodológica combinou segmentação multiresolução, atributos de segmentos e procedimentos interativos e iterativos de aprendizado de máquina. Foram gerados 397 atributos com base nas imagens em amplitude obtidas nas polarizações HH e HV. Esses atributos foram processados no *software* WEKA 3.8 com o emprego dos algoritmos de classificação J48 (árvore de decisão – AD J48), *Random Forest* (Floresta Aleatória - RF) e Rede Neural Artificial *Multilayer Perceptron* (Perceptron Multicamada - MLP). Todos os classificadores do primeiro experimento apresentaram índices de concordância Kappa superiores a 0,70, com destaque para o desempenho da Rede Neural Artificial, com índice Kappa de 0,87. O algoritmo RF apresentou desempenho inferior aos resultados verificados na literatura, o que pode ser atribuído a limitações no tamanho e distribuição espacial da amostragem realizada. No segundo experimento (artigo 2), um fluxo mais abrangente de trabalho foi proposto para a classificação de imagens SAR polarimétricas obtidas também pelo satélite ALOS-2/PALSAR-2, visando ao mapeamento de uso e cobertura de terras. A área de estudo, com 3.660 km<sup>2</sup>, localiza-se no centro-leste do estado de Goiás e no norte e nordeste do DF. A abordagem metodológica combinou atributos polarimétricos, segmentação multiresolução, atributos de segmentos e procedimentos interativos e iterativos de aprendizado de máquina. Também foi utilizado, para fins de comparação, um classificador *pixel-a-pixel* polarimétrico baseado em medidas de distâncias obtidas da distribuição Wishart (Polarimétrico Wishart – PW). Foram gerados 125 atributos utilizando imagens multipolarimétricas e relacionados com



componentes de decomposição de alvos (teoremas de van Zyl, Freeman-Durden, Yamaguchi e Cloude-Pottier), parâmetros polarimétricos incoerentes (índices de biomassa e razões de polarização) e imagens polarizadas em amplitude (HH, HV, VH e VV). Esses atributos foram processados no *software* WEKA 3.8 com o emprego dos algoritmos de classificação *Naive Bayes* (NB), AD J48, RF, MLP e *Support Vector Machine* (Máquina de Suporte Vetorial - SVM). Os classificadores RF, MLP e SVM apresentaram os melhores desempenhos (índices Kappa entre 0,66 e 0,68), tendo sido considerados estatisticamente iguais em ambos os cenários propostos (legendas com nove e cinco classes). Os classificadores NB e AD J48 também apresentaram resultados estatisticamente iguais em ambos os cenários. O AD J48 foi mais adequado para identificação de áreas urbanas e cobertura vegetal natural. O classificador PW apresentou o menor desempenho entre todos os classificadores, porém, com potencial para boa identificação de cobertura vegetal arbórea com os dados SAR na banda L. Os fluxos de trabalho propostos em ambos os experimentos foram ágeis e replicáveis, inclusive com potencial para imagens orbitais provenientes de sensores de radar que operam em comprimentos de onda diferentes daquele do satélite ALOS-2/PALSAR-2.

**Palavras-Chave:** 1. ALOS-2/PALSAR-2; 2. polarimetria SAR; 3. aprendizado de máquina; 4. uso e cobertura de terras; 5. cerrado.

## ABSTRACT

The Brazilian tropical savanna (Cerrado) occupies an area of approximately 2 million km<sup>2</sup> (23% of the Brazilian territory) mainly in the Brazilian Central Plateau. The Cerrado vegetation includes forestlands, shrublands and fields that have undergone severe changes with the introduction of extensive and mechanized agricultural production of grains for exportation. Remote sensing technologies have been used to monitor the Cerrado's vegetation cover using mainly optical satellite images (e.g., the TerraClass and MapBiomass projects). Despite the methodological and technological advances, there are lots of effort to be done using synthetic aperture radar (SAR) satellites to map land use and land cover and to discriminate Cerrados's phytophysionomies. In this thesis, two experiments were carried out with ALOS-2/PALSAR-2 SAR images, in two different study areas over the Brazilian Cerrado. The first experiment (Paper # 1) was carried out in a study area (356 km<sup>2</sup>) located in the northern portion of Brasília, Federal District. The methodological approach proposed in this first experiment combined multiresolution segmentation, object attributes and machine learning procedures. A set of 397 attributes was generated based on the amplitude, HH- and HV-polarized images. These attributes were processed in the WEKA 3.8 software using the J48 (decision tree - DT J48), Random Forest (RF) and Multilayer Perceptron Artificial Neural Network (MLP) classifiers. Classification results attained Kappa indices higher than 0.70, especially the MLP algorithm, with a Kappa index of 0.87. The RF algorithm presented lower performance in comparison with the results presented in the literature, probably due to the reduced number and poor spatial distribution of the training samples. In the second experiment (Paper # 2), a more comprehensive methodology was proposed for the classification of ALOS-2/PALSAR-2 polarimetric SAR images, aiming at the land use and land cover mapping. The study area (3,660 km<sup>2</sup>) encompassed the mid-east of the Goiás State and the north and northeast of the Federal District. The methodological approach of this second experiment combined polarimetric attributes, multiresolution segmentation, segment attributes and machine learning procedures. A polarimetric pixel-to-pixel classifier (Polarimetric Wishart classifier - PW) was also employed for comparison purposes. The PW classifier is based on distance measures calculated using the Wishart distribution. A set of 125 attributes were generated using multipolarimetric images, including the target decomposition components (van Zyl, Freeman-Durden, Yamaguchi, and Cloude-Pottier procedures), incoherent polarimetric parameters (biomass indices and polarization ratios) and polarized amplitude images (HH, HV, VH, and VV polarizations). These attributes were processed in

the WEKA 3.8 software using the Naive Bayes (NB), AD J48, RF, MLP, and Support Vector Machine (SVM) classification algorithms. The RF, MLP, and SVM classifiers presented the best performances and they were considered statistically equal in both proposed scenarios (nine and five thematic classes). Classifiers NB and AD J48 also presented statistically equal results in both scenarios, with AD J48 being more adequate to identify urban areas and natural vegetation coverages. The PW classifier presented the lowest performance among all classifiers. Despite its low performance, the PW classifier presented high potential for classifying forestlands by means of L-band images. The two workflows proposed in this thesis are agile and have potential to be replicated for other satellite images operating at wavelengths other than that from the ALOS-2/PALSAR-2 satellite.

**Keywords:** 1. ALOS-2/PALSAR-2; 2. SAR polarimetry; 3. machine learning; 4. land use and land cover; 5. Brazilian savanna.

## SUMÁRIO

|  |              |
|--|--------------|
| <b>AGRADECIMENTOS</b> .....  | <b>vii</b>   |
| <b>RESUMO</b> .....  | <b>viii</b>  |
| <b>ABSTRACT</b> .....  | <b>x</b>     |
| <b>LISTA DE FIGURAS</b> .....  | <b>xvii</b>  |
| <b>LISTA DE TABELAS</b> .....  | <b>xxi</b>   |
| <b>LISTA DE QUADROS</b> .....  | <b>xxii</b>  |
| <b>LISTA DE SIGLAS E ABREVIATURAS</b> .....                          | <b>xxiii</b> |
| <b>LISTA DE SÍMBOLOS</b> .....                                       | <b>xxvii</b> |
| <b>INTRODUÇÃO</b> .....  | <b>29</b>    |
| 1.1 - Hipótese .....   | 30           |
| 1.2 - Objetivos .....  | 30           |
| 1.2.1 – Objetivo geral .....   | 30           |
| 1.2.2 – Objetivos específicos .....                                  | 30           |
| 1.3 - Estrutura da tese .....  | 31           |
| <b>REFERÊNCIAS</b> .....   | <b>31</b>    |
| <b>FUNDAMENTAÇÃO TEÓRICA</b> .....                                   | <b>33</b>    |
| 2.1 - O bioma Cerrado.....   | 33           |
| 2.2 – Características gerais dos radares de abertura sintética ..... | 35           |
| 2.3 – Princípios básicos de polarimetria SAR.....                    | 39           |
| 2.3.1 – Polarização de ondas eletromagnéticas.....                   | 39           |
| 2.3.2 – Matriz de espalhamento S.....                                | 43           |
| 2.3.3 – Matrizes de covariância e de coerência .....                 | 44           |
| 2.3.4 – Teoremas de decomposição de alvos.....                       | 46           |

|  |           |
|--|-----------|
| 2.3.4.1 – Teorema de van Zyl .....                                 | 46        |
| 2.3.4.2 – Teorema de Freeman-Durden .....                          | 47        |
| 2.3.4.3 – Teorema de Yamaguchi .....                               | 48        |
| 2.3.4.4 – Teorema de Cloude-Pottier (H-A- $\alpha$ ).....          | 49        |
| 2.3.4.4.1 – Entropia polarimétrica (H) .....                       | 50        |
| 2.3.4.4.2 – Anisotropia polarimétrica (A) .....                    | 50        |
| 2.3.4.4.3 – Ângulo $\alpha$ .....                                  | 51        |
| 2.4 – Classificador supervisionado polarimétrico Wishart .....     | 51        |
| 2.5 – Conceitos básicos sobre mineração de dados.....              | 52        |
| 2.5.1. <i>Software</i> WEKA e classificadores utilizados .....     | 54        |
| REFERÊNCIAS .....  | 56        |
| <b>ARTIGO 1.....</b>   | <b>58</b> |
| 3.1 - Introdução .....   | 60        |
| 3.2 - Material e Métodos .....                                     | 62        |
| 3.2.1 - Material.....  | 62        |
| 3.2.1.1 - Área de estudo .....                                     | 62        |
| 3.2.1.2 - O <i>software</i> WEKA e classificadores utilizados..... | 63        |
| 3.2.1.3 - A missão ALOS-2/PALSAR-2.....                            | 64        |
| 3.2.1.4 – Dados e chaves de interpretação utilizados.....          | 64        |
| 3.2.2 - Método.....  | 67        |
| 3.2.2.1 - Fluxograma .....   | 67        |
| 3.2.2.2 - Segmentação de imagens .....                             | 68        |
| 3.2.2.3 - Classificação e validação .....                          | 71        |
| 3.3 - Resultados e Discussão .....                                 | 73        |

|   |           |
|---|-----------|
| 3.4 - Conclusões .....                                  | 80        |
| Agradecimentos.....                                     | 82        |
| Referências .....                                       | 82        |
| <b>ARTIGO 2.....</b>                                    | <b>88</b> |
| 4.1 - Introdução .....                                  | 90        |
| 4.2 - A missão ALOS-2/PALSAR-2 .....                    | 95        |
| 4.3 - Área de estudo.....                               | 95        |
| 4.4 - Material e Métodos .....                          | 97        |
| 4.4.1 - Material.....                                   | 97        |
| 4.4.2 - Fluxo metodológico.....                         | 99        |
| 4.4.3 - Métodos .....                                   | 99        |
| 4.4.3.1 - Legenda e chaves de interpretação .....       | 99        |
| 4.4.3.2 - Pré-processamentos .....                      | 105       |
| Calibração radiométrica .....                           | 105       |
| Filtragem do ruído <i>speckle</i> .....                 | 106       |
| 4.4.3.3 - Parâmetros polarimétricos incoerentes .....   | 107       |
| 4.4.3.4 - Decomposições polarimétricas .....            | 109       |
| 4.4.3.5 - Ortorretificação.....                         | 113       |
| 4.4.3.6 - Segmentação e extração de atributos.....      | 115       |
| 4.4.3.7 - Classificações .....                          | 117       |
| Classificadores do aplicativo WEKA .....                | 117       |
| Classificador supervisionado polarimétrico Wishart..... | 119       |
| 4.4.3.8 - Validações .....                              | 122       |
| 4.5 - Resultados e Discussão .....                      | 123       |

|  |            |
|--|------------|
| 4.6 - Conclusões .....   | 151        |
| Agradecimentos.....  | 153        |
| Referências .....  | 153        |
| <b>CONSIDERAÇÕES FINAIS .....</b>  | <b>161</b> |
| <b>APÊNDICE A: RELAÇÃO DE ATRIBUTOS UTILIZADOS NAS CLASSIFICAÇÕES<br/>REALIZADAS NO WEKA 3.8.....</b>  | <b>163</b> |
| <b>APÊNDICE B: ÁRVORE DE DECISÃO GERADA PELO CLASSIFICADOR J48<br/>(LEGENDA COM NOVE CLASSES) .....</b>  | <b>167</b> |
| <b>APÊNDICE C: ATRIBUTOS EMPREGADOS PELA ÁRVORE DE DECISÃO J48<br/>(APÊNDICE B).....</b>   | <b>170</b> |
| <b>APÊNDICE D: RELAÇÃO DE ATRIBUTOS OBTIDOS UTILIZANDO O<br/>PROCEDIMENTO DE SELEÇÃO DO ALGORITMO <i>INFOGAIN</i> (LEGENDA COM<br/>NOVE CLASSES) .....</b> | <b>171</b> |
| <b>APÊNDICE E: MATRIZES DE CONFUSÃO REFERENTES AO CAPÍTULO 4<br/>(LEGENDA COM NOVE CLASSES) .....</b>  | <b>174</b> |
| <b>APÊNDICE F: MATRIZES DE CONFUSÃO REFERENTES AO CAPÍTULO 4<br/>(LEGENDA COM CINCO CLASSES).....</b>  | <b>178</b> |
| <b>APÊNDICE G: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4:<br/>SIGNIFICÂNCIAS DOS ÍNDICES KAPPA (LEGENDA COM NOVE CLASSES) ...</b>                       | <b>181</b> |
| <b>APÊNDICE H: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4:<br/>SIGNIFICÂNCIAS DOS ÍNDICES KAPPA (LEGENDA COM CINCO CLASSES) .</b>                        | <b>183</b> |
| <b>APÊNDICE I: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4:<br/>COMPARAÇÕES DE ÍNDICES KAPPA (LEGENDA COM NOVE CLASSES).....</b>                          | <b>185</b> |
| <b>APÊNDICE J: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4:<br/>COMPARAÇÕES DE ÍNDICES KAPPA (LEGENDA COM CINCO CLASSES).....</b>                         | <b>189</b> |
| <b>ANEXO A: COMPROVANTE DE SUBMISSÃO E SITUAÇÃO DO ARTIGO # 1....</b>  | <b>193</b> |

**ANEXO B: DADOS DE PRECIPITAÇÃO PLUVIOMÉTRICA DA ESTAÇÃO METEOROLÓGICA AUTOMÁTICA DO INMET (01/05/2016 - 14/05/2016) – ESTAÇÃO ECOLÓGICA ÁGUAS EMENDADAS (BRASÍLIA/DF). ..... 195**



## LISTA DE FIGURAS

|   |    |
|---|----|
| Figura 2.1. Abrangência geográfica do bioma Cerrado: Estende-se desde o norte do estado do Paraná até próximo à linha do equador, no litoral do Maranhão. Fonte: Sano et al. (2010)....   | 33 |
| Figura 2.2. Esquema em perfil das fitofisionomias do bioma Cerrado. Fonte: EMBRAPA (2018). .....  | 34 |
| Figura 2.3. Região das micro-ondas do espectro eletromagnético. Fonte: Lee e Pottier (2009). .....  | 35 |
| Figura 2.4. Geometria de imageamento SAR. Fonte: Lee e Pottier (2009). .....  | 36 |
| Figura 2.5. Representações lateral e frontal do campo elétrico de uma onda eletromagnética. (a) Onda com polarização vertical (V). (b) Onda com polarização horizontal (H). Ilustrações do comprimento de onda ( $\lambda$ ), amplitude (A), fase ( $\phi$ ) variando entre 0 a $2\pi$ , campo elétrico vertical ( $E_v$ ), campo elétrico horizontal ( $E_h$ ). Fonte: van der Sanden (1997).....        | 40 |
| Figura 2.6. Trajetória temporal de uma onda monocromática plana e em dada direção fixa. Fonte: Lee e Pottier (2009).....  | 41 |
| Figura 2.7. Elipse de polarização para representação do vetor campo elétrico. Fonte: Richards (2009). .....   | 42 |
| Figura 2.8. Etapas do processo descoberta de conhecimento em bases de dados (DCBD): limpeza e integração; seleção e transformação; mineração de dados; e avaliação e apresentação. Fonte: Han et al. (2012). .....  | 52 |
| Figura 2.9. Fluxo de processos DCBD em imagens de sensoriamento remoto. Fonte: adaptado de Hsu et al. (2002). .....   | 53 |
| Figura 3.1. Localização da área de estudo no Distrito Federal. A imagem corresponde a uma composição colorida RGB (RGB/HV-HH-HH) da cena do satélite ALOS-2/PALSAR-2, obtida em passagem de 29 de abril de 2015. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S. .....   | 63 |
| Figura 3.2. Fluxograma das principais etapas do trabalho. ....  | 68 |
| Figura 3.3. Exemplo de um resultado da segmentação de imagens do satélite ALOS-2/PALSAR-2 envolvendo as polarizações HH e HV. Os seguintes parâmetros de segmentação foram considerados: escala = 350; forma = 0,9; cor = 0,1; compactação = 0,1; suavização = 0,9. Os segmentos foram sobrepostos a uma composição colorida RGB (RGB/HV-HH-HH). <i>Datum</i> horizontal: WGS 84; Projeção: UTM 23S. .... | 69 |
| Figura 3.4. Exemplos de atributos extraídos por meio de segmentos gerados por segmentação multiresolução: média por objeto da polarização HH em amplitude (a); média por objeto da  |    |

|  |     |
|--|-----|
| polarização HH em intensidade (b); GLCM entropia por objeto (todas as direções) (c); e desvio-padrão por objeto da polarização HH em amplitude (d). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....   | 71  |
| Figura 3.5. Localização das amostras de treinamento (em azul) e de validação (em amarelo) na porção sul da área de estudo. A imagem refere-se à composição colorida RGB das polarizações HH, HV. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....  | 72  |
| Figura 3.6. Resultados da classificação de imagens pelos algoritmos AD J48 (a), RF (b) e RNA MLP (c).....  | 75  |
| Figura 3.7. Árvore de decisão gerada pelo classificador J48. ....  | 76  |
| Figura 4.1. Localização da área de estudo, entre Goiás e Distrito Federal. A imagem corresponde a uma composição colorida RGB (RGB/HH-HV-VV) da cena do satélite ALOS-2/PALSAR-2, obtida em 14 de maio de 2016. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....   | 96  |
| Figura 4.2. Fluxograma das principais etapas do trabalho. ....   | 101 |
| Figura 4.3. Composição colorida RGB LANDSAT-8/OLI (RGB/B4-B5-B3): Distribuição dos pontos amostrais utilizados na validação, pontos visitados em campo e trechos percorridos de automóvel (trabalho de campo realizado em setembro de 2017). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....  | 102 |
| Figura 4.4. Filtragem do ruído <i>speckle</i> utilizando o filtro <i>Refined Lee</i> e janela 7 x 7. ....  | 107 |
| Figura 4.5. Filtragem do ruído <i>speckle</i> utilizando o filtro <i>Refined Lee</i> e janela 9 x 9. ....  | 107 |
| Figura 4.6. Composições coloridas RGB apresentando: a) imagens polarizadas em amplitude (HH/HV/VV); b); teorema de Freeman-Durden ( <i>double/volume/surface</i> ); c) teorema de Cloude-Pottier (H/A/ $\alpha$ ); d) teorema de van Zyl ( <i>double/volume/surface</i> ); e) Teorema de Yamaguchi sem a componente <i>Helix</i> ( <i>double/volume/surface</i> ) e, f) Teorema de Yamaguchi com a componente <i>Helix</i> ( <i>double/helix/surface</i> ). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S..... | 112 |
| Figura 4.7. Parte da cena ALOS-2/PALSAR-2 SLC 1.1 original e comprimida em alcance. ....   | 114 |
| Figura 4.8. Parte da cena ALOS-2/PALSAR-2 SLC 1.1 ortorretificada, usando o modelo <i>range doppler</i> do SNAP 6.0. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S. ....   | 114 |
| Figura 4.9. Exemplo da segmentação obtida em uma região situada na porção central da área de estudo. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S. ....   | 116 |
| Figura 4.10. Gráfico <i>box-plot</i> gerado com o conjunto amostral total de treinamento contendo o atributo média de $\sigma^0$ (dB) na polarização HH. A = Pastagens; B= Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado sujo; D = Culturas temporárias; E   |     |

|  |     |
|--|-----|
| = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. ....   | 118 |
| Figura 4.11. Exemplo da distribuição das amostras de treinamento e de validação. <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S. ....   | 119 |
| Figura 4.12. Assinaturas polarimétricas (co-polarização) de cada classe, obtidas com as amostras de treinamento: a) Pastagens; b) Formações florestais; c) Cerrado Típico antropizado/Cerrado Ralo/Campo Sujo; d) Culturas temporárias; e) Solo exposto; f) Silvicultura; g) Corpo d'água continental; h) Áreas com edificações (com reflexão de canto); i) Áreas com edificações (com espalhamento volumétrico); e j) Campo Limpo. .... | 121 |
| Figura 4.13. Variação dos valores de acurácia global por conjunto de treinamento (legenda com nove classes). ....  | 126 |
| Figura 4.14. Variação do índice Kappa por conjunto de treinamento (legenda com nove classes). ....   | 127 |
| Figura 4.15. Representação gráfica das acurácias do usuário e do produtor para cada classificador utilizado (legenda com nove classes). ....   | 133 |
| Figura 4.16. Representação gráfica dos valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com nove classes). ....   | 134 |
| Figura 4.17. Representação gráfica das acurácias do usuário e do produtor para cada classificador utilizado (legenda com cinco classes). ....  | 136 |
| Figura 4.18. Representação gráfica dos valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com cinco classes). ....  | 137 |
| Figura 4.19. Índice Kappa e acurácia global por classificador. Legenda com nove classes e conjunto amostral com 200 amostras por classe. ....  | 138 |
| Figura 4.20. Índice Kappa e acurácia global por classificador. Legenda com cinco classes e conjunto amostral com 200 amostras por classe. ....   | 138 |
| Figura 4.21. Classificações obtidas com os algoritmos PW (a) e NB (b). Legenda com nove classes. ....  | 139 |
| Figura 4.22. Classificações obtidas com os algoritmos AD J48 (a) e RF (b). Legenda com nove classes. ....  | 140 |
| Figura 4.23. Classificações obtidas com os algoritmos MLP (a) e SVM (b). Legenda com nove classes. ....  | 141 |
| Figura 4.24. Classificações obtidas com os algoritmos PW (a) e NB (b). Legenda com cinco classes. ....   | 142 |

|  |     |
|--|-----|
| Figura 4.25. Classificações obtidas com os algoritmos AD J48 (a) e RF (b). Legenda com cinco classes. ....   | 143 |
| Figura 4.26. Classificações obtidas com os algoritmos MLP (a) e SVM (b). Legenda com cinco classes. ....   | 144 |
| Figura 4.27. a) Classificação obtida com o algoritmo SVM e conjunto de treinamento de 200 amostras por classe (legenda com nove classes). b) Classificação gerada pelo programa TerraClass (INPE, 2017). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....  | 146 |
| Figura 4.28. a) Classificação obtida com o algoritmo SVM e conjunto de treinamento de 200 amostras por classe (legenda com cinco classes). b) Classificação gerada pelo programa TerraClass (INPE, 2017). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....   | 147 |
| Figura 4.29. Representação gráfica dos resultados dos testes Z (legenda com nove classes). Na abscissa estão os diferentes classificadores utilizados, e na ordenada, os p-valores calculados. Classificadores que apresentarem p-valores maiores que $\alpha$ (0,05) são considerados iguais aos seus correspondentes na abscissa. .... | 149 |
| Figura 4.30. Representação gráfica dos resultados do teste Z (legenda com cinco classes). Na abscissa estão os diferentes classificadores utilizados e na ordenada os p-valores calculados. Classificadores que apresentarem p-valores maiores que $\alpha$ (0,05) são considerados iguais aos seus correspondentes na abscissa. ....    | 150 |

## LISTA DE TABELAS

|  |     |
|--|-----|
| Tabela 3.1. Matriz de confusão da classificação de imagens obtida pelo algoritmo AD J48. .   | 77  |
| Tabela 3.2. Matriz de confusão da classificação de imagens obtida pelo algoritmo RF. ....  | 78  |
| Tabela 3.3. Matriz de confusão obtida pelo classificador RNA MLP. ....   | 80  |
| Tabela 4.1. Valores de acurácia global por conjunto de treinamento (legenda com nove classes). ....  | 125 |
| Tabela 4.2. Índice Kappa por conjunto de treinamento (legenda com nove classes). ....  | 126 |
| Tabela 4.3. Acurácias do usuário e do produtor para cada classificador utilizado (legenda com nove classes). ....                            | 132 |
| Tabela 4.4. Valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com nove classes). ....  | 132 |
| Tabela 4.5. Acurácias do usuário e do produtor para cada classificador utilizado (legenda com cinco classes). ....                           | 135 |
| Tabela 4.6. Valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com cinco classes). .... | 135 |
| Tabela 4.7. Matriz de p-valores obtidos nas comparações entre os classificadores utilizados. Legenda com nove classes. ....                  | 148 |
| Tabela 4.8. Matriz de p-valores obtidos nas comparações entre os classificadores utilizados. Legenda com cinco classes. ....                 | 149 |

## LISTA DE QUADROS

|  |     |
|--|-----|
| Quadro 3.1. Detalhes dos diferentes modos de aquisição do satélite ALOS-2/PALSAR-2. ...  | 64  |
| Quadro 3.2. Chave de interpretação visual de composições coloridas RGB das polarizações HH, HV (RGB/HV-HH-HH). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S.....  | 66  |
| Quadro 3.3. Categorias e correspondentes tipos de atributos que foram considerados neste estudo. ....  | 70  |
| Quadro 3.4. Teste de hipótese da classificação obtida pelo algoritmo AD J48. ....  | 77  |
| Quadro 3.5. Teste de hipótese da classificação obtida pelo algoritmo RF.....   | 78  |
| Quadro 3.6. Teste de hipótese comparando o desempenho dos classificadores AD J48 e RF.   | 79  |
| Quadro 3.7. Teste de hipótese da classificação de imagens obtida pelo classificador RNA MLP.....   | 79  |
| Quadro 3.8. Teste de hipótese comparando o desempenho dos classificadores RF e RNA MLP.....  | 80  |
| Quadro 4.1. Detalhes dos diferentes modos de aquisição do satélite ALOS-2/PALSAR-2. ...  | 95  |
| Quadro 4.2. Relação de imagens utilizadas na pesquisa. ....  | 97  |
| Quadro 4.3. Relação de dados e de material auxiliar utilizados na pesquisa. ....   | 98  |
| Quadro 4.4. Classes da legenda e respectivas chaves de interpretação visual utilizando composições coloridas RGB ALOS-2/PALSAR-2 (RGB/HH-HV-VV) e LANDSAT-8/OLI (RGB/B5-B3-B4). <i>Datum</i> horizontal: WGS84; Projeção: UTM 23S..... | 103 |
| Quadro 4.5. Relação de parâmetros polarimétricos incoerentes empregados nesta pesquisa.  | 109 |
| Quadro 4.6. Relação de teoremas de decomposição de alvos utilizados nesta pesquisa. ....   | 111 |
| Quadro 4.7. Imagens e parâmetros utilizados na segmentação.....  | 116 |
| Quadro 4.8. Categorias e correspondentes tipos de atributos que foram considerados neste estudo. ....  | 116 |
| Quadro 4.9. Conjuntos de segmentos utilizados nos classificadores do <i>software</i> WEKA 3.8. ....  | 118 |
| Quadro 4.10. Conjuntos de <i>pixels</i> utilizados na classificação e validação. ....  | 120 |
| Quadro 4.11. Métricas de validação temática utilizadas nesta pesquisa. ....  | 122 |
| Quadro 4.12. <i>Ranking</i> final dos classificadores para as legendas de nove e cinco classes. ...  | 150 |

## LISTA DE SIGLAS E ABREVIATURAS

|          |   |
|----------|---|
| ALOS     | <i>Advanced Land Observing Satellite</i><br>(Satélite Avançado de Observação da Terra)  |
| AD J48   | Árvore de Decisão J48   |
| ARFF     | <i>Attribute-Relation File Format</i><br>(Formato de Arquivo de Relação de Atributos)   |
| ASCII    | <i>American Standard Code for Information Interchange</i><br>(Código Padrão Americano para Intercâmbio de Informações)                          |
| BMI      | <i>Biomass Index</i> (Índice de Biomassa)   |
| BOS      | <i>Basic Observation Scenario</i> (Cenário Básico de Observação)  |
| BSA      | <i>Backscatter Alignment</i> (Alinhamento de Retroespalhamento)   |
| CART     | <i>Classification and Regression Tree</i><br>Árvore de Regressão e Classificação  |
| C4.5     | Classificador por árvore de decisão desenvolvido em linguagem C   |
| CEOS     | Formato de arquivo que segue as especificações do <i>Committee on Earth Observation Satellites</i> (Comitê de Satélites de Observação da Terra) |
| CODEPLAN | Companhia de Planejamento do Distrito Federal   |
| CSI      | <i>Canopy Structure Index</i> (Índice de Estrutura de Dossel)   |
| EMBRAPA  | Empresa Brasileira de Pesquisa Agropecuária   |
| EMQ      | Erro Médio Quadrático   |
| AP       | Acurácia do Produtor  |
| AU       | Acurácia do Usuário   |
| ESA      | <i>European Space Agency</i> (Agência Espacial Européia)  |
| ETM+     | <i>Enhanced Thematic Mapper Plus</i> (Mapeador Temático Melhorado)  |
| GLCM     | <i>Gray Level Co-occurrence Matrix</i> (Matriz de Co-ocorrência de Nível de Cinza)  |
| GLDV     | <i>Gray Level Difference Vector</i> (Vetor de Diferença de  |

|           |  |
|-----------|--|
|           | Nível de Cinza)  |
| GLOVIS    | <i>Global Visualization Viewer</i> (Visualizador de Visualização Global)                               |
| IAF       | Índice de Área Foliar  |
| IBGE      | Instituto Brasileiro de Geografia e Estatística  |
| INFOGAIN  | Algoritmo de seleção de atributos com base no cálculo do valor de entropia                             |
| INMET     | Instituto Nacional de Meteorologia   |
| INPE      | Instituto Nacional de Pesquisas Espaciais  |
| JAXA      | <i>Japan Aerospace Exploration Agency</i> (Agência de Exploração Aeroespacial do Japão)                |
| JERS      | <i>Japanese Earth Resource Satellite</i> (Satélite Japonês de Recursos Terrestres)                     |
| KCP       | Índice Kappa Condicional do Produtor   |
| KCU       | Índice Kappa Condicional do Usuário  |
| KDD       | <i>Knowledge Discovery in Databases</i> (Descoberta de Conhecimento em Bases de Dados)                 |
| LANDSAT   | <i>Land Remote Sensing Satellite</i> (Satélite de Sensoriamento Remoto Terrestre)                      |
| LIDAR     | <i>Light Detection and Ranging</i> (Detecção e Telemetria por Luz)                                     |
| LVC       | Laboratório de Visão Computacional da Pontifícia Universidade Católica do Rio de Janeiro               |
| MAPBIOMAS | Projeto multi-institucional de monitoramento anual do uso e cobertura de terras dos biomas brasileiros |
| MDE       | Modelo Digital de Elevação   |
| MLP       | <i>Multilayer Perceptron</i> (Perceptron Multicamada)  |
| NB        | <i>Naive Bayes</i>   |
| NEXTGIS   | <i>Opensource Geospatial Solutions Company</i> (Empresa de Soluções Geoespaciais de Código Aberto)     |
| OLI       | <i>Operational Land Imager</i> (Imageador Operacional da Superfície Terrestre)                         |
| PALSAR    | <i>Phased Array type L-band Synthetic Aperture Radar</i> (Radar de Abertura Sintética Banda L do tipo  |



|                  |  |
|------------------|--|
|                  | Matriz Faseada)  |
| PAM              | Produção Agrícola Municipal  |
| PCIDSK           | Formato nativo de arquivo do aplicativo PCI Geomatica                            |
| PCT              | Ponto de Controle no Terreno   |
| PNB              | Parque Nacional de Brasília  |
| PolInSAR         | Polarimetria e Interferometria SAR   |
| PW               | Classificador Polarimétrico com distribuição Wishart                             |
| QGIS             | <i>Quantum GIS</i>   |
| QuickMapServices | <i>Plugin for basemaps using QGIS</i> (Extensão para mapas base usando o QGIS)   |
| RADAMBRASIL      | Projeto Radar da Amazônia e Brasil   |
| RADAR            | <i>Radio Detection and Ranging</i> (Detecção e Telemetria por Rádio)             |
| RF               | <i>Random Forest</i> (Floresta Aleatória)  |
| RFDI             | <i>Radar Forest Degradation Index</i> (Índice de Degradação Florestal por Radar) |
| RLOS             | <i>Radar Line of Sight</i> (Linha de Visado do Radar)                            |
| RNA              | Rede Neural Artificial   |
| RVI              | <i>Radar Vegetation Index</i> (Índice de Vegetação por Radar)                    |
| SAR              | <i>Synthetic Aperture Radar</i> (Radar de Abertura Sintética)                    |
| SARINGEST        | Algoritmo de importação de dados SAR polariméricos para o formato PCIDSK         |
| SEGETH           | Secretaria de Estado de Gestão do Território e Habitação do Distrito Federal     |
| SENTINEL         | Missão de observação da Terra da Agência Espacial Européia (ESA)                 |
| SLC              | <i>Single Look Complex</i> (Dado Complexo Obtido por Visada Única)               |
| SMO              | <i>Sequential Minimal Optimization</i> (Otimização Sequencial Mínima)            |
| SNAP             | <i>Sentinel Application Platform</i> (Plataforma de Aplicação Sentinel)          |
| SPRING           | Sistema de Processamento de Informações Georreferenciadas                        |

|       |   |
|-------|---|
| SRTM  | <i>Shuttle Radar Topography Mission</i> (Missão Topográfica do Radar Shuttle)                     |
| SVM   | <i>Support Vector Machine</i> (Máquina de Suporte Vetorial)                                       |
| TM    | <i>Thematic Mapper</i> (Mapeador Temático)  |
| USGS  | <i>United States Geological Survey</i> (Serviço Geológico dos Estados Unidos da América)          |
| UTM   | Universal Transversa de Mercator  |
| VSI   | <i>Volume Scattering Index</i> (Índice de Espalhamento Volumétrico)                               |
| WEKA  | <i>Waikato Environment for Knowledge Analysis</i> (Waikato Ambiente para Análise do Conhecimento) |
| WGS84 | <i>World Geodetic System 1984</i> (Sistema Geodésico Mundial 1984)                                |

## LISTA DE SÍMBOLOS

|                         |  |
|-------------------------|--|
| $\langle \dots \rangle$ | Média espacial em matrizes                               |
| $\alpha$                | Ângulo alfa (componente do teorema de Cloude-Pottier)    |
| $a$                     | Amplitude da fase do sinal elétrico                      |
| $A$                     | Anisotropia (componente do teorema de Cloude-Pottier)    |
| $L$                     | Banda L PALSAR-2 ( $\lambda = 23,5$ cm)                  |
| $\Psi$                  | Base matricial 2x2                                       |
| $CF$                    | Coefficiente de calibração para $\sigma^0$               |
| $\sigma^0$              | Coefficiente de retroespalhamento                        |
| $\lambda$               | Comprimento da onda eletromagnética                      |
| dB                      | Decibéis   |
| $\delta$                | Diferença de fase  |
| $H$                     | Entropia (componente do teorema de Cloude-Pottier)       |
| $\hat{K}$               | Estimador KHAT do índice Kappa                           |
| $\phi$                  | Fase da onda elétrica                                    |
| $A$                     | Fator de conversão para calibração em $\sigma^0$         |
| $\omega$                | Frequência angular da onda elétrica                      |
| $G_t$                   | Ganho da antena de transmissão                           |
| $I$                     | Imagem Imaginária  |
| $Q$                     | Imagem Real  |
| $K$                     | Índice de concordância Kappa                             |
| $\tau$                  | Índice de concordância Tau                               |
| $[S]$                   | Matriz de espalhamento S                                 |
| $[C]$                   | Matriz de covariância                                    |
| $[T]$                   | Matriz de coerência                                      |
| HH                      | Polarizações horizontais, emitida e recebida pela antena |
| HV                      | Polarização horizontal emitida e                         |

|           |   |
|-----------|---|
| VH        | polarização vertical recebida pela antena<br>Polarização vertical emitida e polarização horizontal recebida pela antena |
| VV        | Polarizações verticais, emitida e recebida pela antena  |
| $P_r$     | Potência do sinal recebido (Watts)  |
| $P_t$     | Potência do sinal transmitido (Watts)   |
| $t$       | Tempo   |
| $c$       | Velocidade da luz (constante)   |
| $\vec{E}$ | Vetor campo elétrico  |
| $\vec{r}$ | Vetor posição (distância)   |
| $\vec{k}$ | Vetor de propagação (a direção da onda)   |

### INTRODUÇÃO

O Cerrado cobre extensa área no interior do Brasil. Em termos de latitude, estende-se desde o norte do estado do Paraná até próximo à linha do equador, no litoral do Maranhão (RIBEIRO e WALTER, 2008; SANO et al., 2010). Aproximadamente até 1970, o uso de terras no Cerrado se concentrava na pecuária extensiva e baseada nos campos ou pastagens naturais. Nas últimas décadas, o uso e ocupação sofreu mudanças severas com a introdução da agricultura extensiva, caracterizada principalmente pela produção mecanizada de grãos para exportação (JEPSON, 2005; KLINK e MACHADO 2005; SANO et al., 2010).

Atualmente, o sensoriamento remoto tem se mostrado uma ferramenta eficiente para o monitoramento do Cerrado. Com base em segmentação e interpretação visual de imagens do sensor LANDSAT-7/ETM+ do ano de 2002, Sano et al. (2010) realizaram um mapeamento em escala de semidetalhe (1:250.000) do Cerrado com acurácia temática global de 70%. Mais recentemente, os projetos institucionais TerraClass (INPE, 2017) e MapBiomias (MAPBIOMAS, 2017) têm desenvolvido mapeamentos e inventários sistemáticos do Cerrado também em escalas de semidetalhe, empregando imagens ópticas.

Ainda, segundo Ribeiro e Walter (2008), o bioma Cerrado apresenta uma estação chuvosa bem demarcada que ocorre geralmente entre os meses de outubro a março. Devido a essa característica climatológica, a aquisição de imagens ópticas fica restrita ao período de estiagem (dada a baixa cobertura de nuvens), limitando assim o entendimento da sazonalidade anual do referido bioma.

Neste sentido, avanços tecnológicos atuais têm demonstrado que as imagens de radar representam uma fonte viável de dados para mapeamento sistemático de uso e cobertura de terras do Cerrado. Os sensores de radar de abertura sintética (SAR) apresentam as vantagens de imageamento noturno e diurno e capacidade de penetrar nuvens. Em comprimentos de onda mais longos, ocorre ainda a penetração nos dosséis florestais, favorecendo assim a melhor caracterização da biomassa (SANO et al., 2001; 2005, 2009; BITENCOURT et al., 2007; OUCHI, 2013). Dentre as iniciativas existentes usando imagens de radar, destacam-se os mapeamentos de vegetação

realizados pelo projeto RADAMBRASIL (1:1.000.000) na década de 1970, com mosaicos semicontrolados obtidos por plataforma aérea (IBGE, 2018). Recentemente, diversos autores têm analisado o potencial das imagens SAR orbitais no estudo e mapeamento da vegetação de Cerrado (SANO et al., 2001; 2005; 2009; BITENCOURT et al., 2007).

Neste contexto, esta pesquisa apresenta uma abordagem nova de classificação de imagens ALOS-2/PALSAR-2 utilizando atributos polarimétricos (teoremas de decomposição de alvos e parâmetros polarimétricos) e técnicas de mineração de dados (algoritmos de aprendizado de máquina) visando contribuir para a análise, entendimento e mapeamento, em escala de semidetalhe, do uso e cobertura de terras no Cerrado.

## **1.1 - Hipótese**

O emprego combinado de técnicas de segmentação de imagens e de mineração de dados aplicadas à classificação de dados SAR polarizados ou polarimétricos representam alternativa viável e de alta acurácia temática para fins de identificação semiautomática de padrões de uso e cobertura de terras em áreas de Cerrado.

## **1.2 - Objetivos**

### **1.2.1 – Objetivo geral**

O objetivo da presente tese é realizar um estudo comparativo entre técnicas de mineração de dados para fins de identificação de padrões de uso e cobertura de terras em área de Cerrado usando imagens polarimétricas do sistema sensor ALOS-2/PALSAR-2 e empregando, para efeito comparativo, resultados obtidos com classificador SAR polarimétrico Wishart.

### **1.2.2 – Objetivos específicos**

- a) Gerar dados polarizados e polarimétricos a serem empregados na segmentação e na mineração de dados;
- b) Testar parâmetros de segmentação para a melhor delimitação dos padrões de uso e cobertura de terras;

- c) Realizar mineração de dados sobre os segmentos e seus atributos empregando diferentes algoritmos de aprendizado de máquina;
- d) Treinar e aplicar classificador polarimétrico baseado na distribuição de probabilidades Wishart para fins de comparação dos resultados da mineração de dados;
- e) Construir um *ranking* dos classificadores de acordo com seus desempenhos obtidos dos experimentos realizados;
- f) Elencar limitações e potencialidades dos dados e classificadores avaliados.

### 1.3 - Estrutura da tese

No Capítulo 1, realizou-se uma introdução geral, apresentando o contexto da investigação, a hipótese e os objetivos da presente pesquisa.

No Capítulo 2, foi realizada a revisão bibliográfica pertinente sobre as principais teorias e conceitos tratados nesta pesquisa.

O Capítulo 3 consiste no primeiro artigo, no qual se analisaram os desempenhos de três algoritmos de aprendizado de máquina na classificação baseada em segmentos de imagens ALOS-2/PALSAR-2 em amplitude e em duas polarizações (HH e HV).

O Capítulo 4 apresenta o segundo artigo, que trata da classificação de imagens multipolarizadas e multipolarimétricas (HH, HV, VH e VV) do sensor ALOS-2/PALSAR-2, empregando atributos polarimétricos, segmentação e mineração de dados. Foram testados cinco algoritmos de aprendizado de máquina e um classificador *pixel-a-pixel* polarimétrico que utiliza a distribuição Wishart.

No Capítulo 5, têm-se as considerações finais da tese.

### REFERÊNCIAS

BITENCOURT, M. D.; MESQUITA Jr., H. N.; KUNTSCHIK, G.; ROCHA, H. R.; FURLEY, P. A. Cerrado vegetation study using optical and radar remote sensing: two Brazilian case studies. **Canadian Journal of Remote Sensing**, v. 33, n. 6, p. 468–480, 2007.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Desbravar, conhecer, mapear: memórias do projeto Radam/RadamBrasil**. Rio de Janeiro: IBGE, 2018, 364 p.

INPE. Instituto Nacional de Pesquisas Espaciais. **Projeto TerraClass Cerrado: mapeamento do uso e cobertura vegetal do Cerrado**. 2017. Disponível em: <<http://www.dpi.inpe.br/tccerrado/download.php>>. Acesso em: 01 jul. 2017.

JEPSON, W. A disappearing biome? Reconsidering land-cover change in the Brazilian savanna. **The Geographical Journal**, v. 171, n. 2, p. 99–111, 2005.

KLINK, C. A.; MACHADO, R. B. Conservation of the Brazilian Cerrado. **Conservation Biology**, v. 19, p. 707–713, 2005.

MAPBIOMAS. **Mapeamento anual da cobertura e uso do solo no Brasil**. 2017. Disponível em: <<http://mapbiomas.org>>. Acesso em: 15 jun. 2017.

OUCHI, K. Recent trend and advance of synthetic aperture radar with selected topics. **Remote Sensing**, v. 5, n. 2, p. 716–807, 2013.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. (Eds.). **Cerrado: Ecologia e Flora**, Planaltina: Embrapa Cerrados, 2008, pp. 151–199.

SANO, E. E.; PINHEIRO, G. C. C.; MENESES, P. R. Assessing JERS-1 synthetic aperture radar data for vegetation mapping in the Brazilian savanna. **Journal of Remote Sensing Society of Japan**, v. 21, p. 158–167, 2001.

SANO, E. E.; FERREIRA, L. G.; HUETE, A. R. Synthetic aperture radar (L-band) and optical vegetation indices for discriminating the Brazilian savanna physiognomies: A comparative analysis. **Earth Interactions**, v. 9 (Paper no. 15), p. 1–15, 2005.

SANO, E. E.; SANTOS, E. M.; MENESES, P. R. Análise de imagens do satélite ALOS PALSAR para o mapeamento de uso e cobertura da terra do Distrito Federal. **Geociências**, v. 28, n. 4, p. 441–451, 2009.

SANO, E. E.; ROSA, R.; BRITO, J. L. S.; FERREIRA, L. G. Land cover mapping of the tropical savanna region in Brazil. **Environmental Monitoring and Assessment**, v. 166, p. 116–124, 2010.



## FUNDAMENTAÇÃO TEÓRICA

## 2.1 - O bioma Cerrado

O Cerrado é o segundo maior bioma brasileiro em área, superado apenas pela Floresta Amazônica, ocupa mais de 2.000.000 km<sup>2</sup> (23% do território brasileiro) e está localizado majoritariamente no Planalto Central (Figura 2.1). As altitudes variam entre 300 m e 1.600 m, onde predominam os Latossolos. O clima é caracterizado pela presença de invernos secos e verões chuvosos, com média anual de precipitação de 1.500 mm e variação entre 750 mm e 2.000 mm. As chuvas concentram-se entre os meses de outubro a março (RIBEIRO e WALTER, 2008).

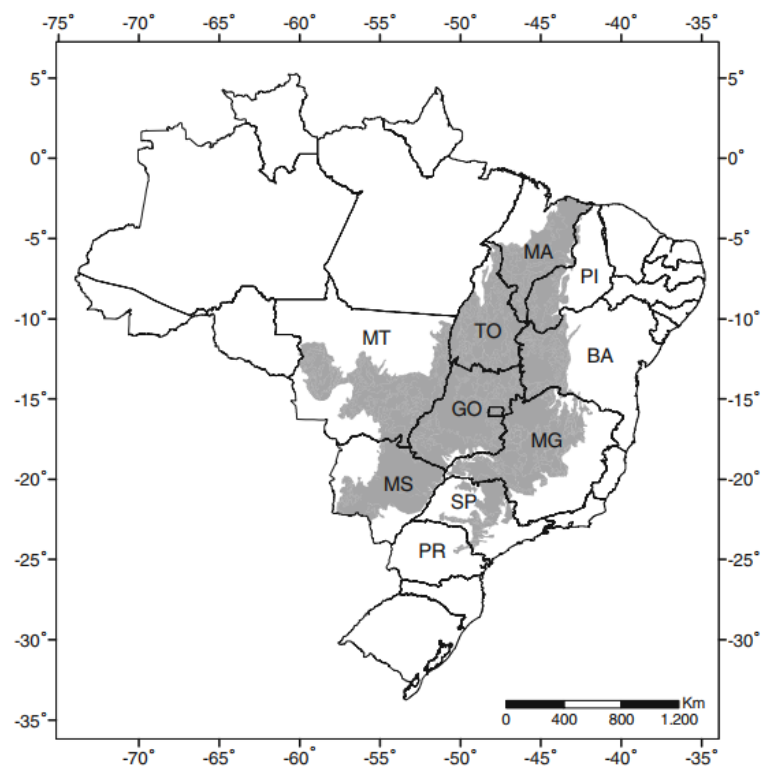


Figura 2.1. Abrangência geográfica do bioma Cerrado: Estende-se desde o norte do estado do Paraná até próximo à linha do equador, no litoral do Maranhão. Fonte: Sano et al. (2010).

A composição fisionômica da vegetação abrange formações florestais, savânicas e campestres. Segundo Ribeiro e Walter (2008), as formações florestais são constituídas de Mata Ciliar, Mata de Galeria, Mata Seca e Cerradão. Por sua vez, a formação savânica é o próprio Cerrado em sentido estrito (áreas com árvores e arbustos espalhados sobre um estrato graminoso e sem um dossel contínuo), que pode ser subdividido em Cerrado Denso, Cerrado Típico, Cerrado Ralo e Cerrado Rupestre, além de Vereda, Parque de Cerrado e Palmeiral. Por fim, as formações campestres correspondem a Campo Sujo, Campo Limpo e Campo Rupestre. A Figura 2.2 apresenta um esquema em perfil das fitofisionomias do bioma Cerrado.

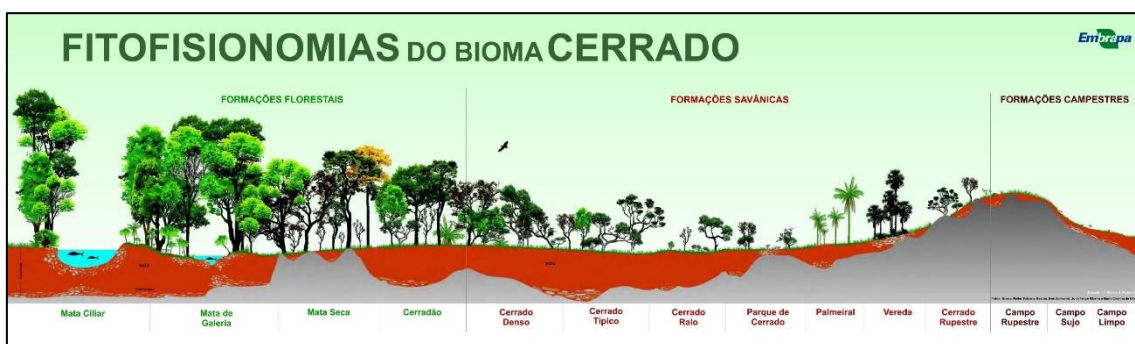


Figura 2.2. Esquema em perfil das fitofisionomias do bioma Cerrado. Fonte: EMBRAPA (2018).

## 2.2 – Características gerais dos radares de abertura sintética

Em sentido amplo, radar (que corresponde ao acrônimo em inglês: *radio detection and ranging* - detecção e telemetria por rádio) é um instrumento que mede o tempo de propagação que uma radiação eletromagnética, emitida pela sua própria antena, leva para atingir um alvo e retornar. Determina-se a posição do alvo por meio da conversão do tempo de propagação em distância radial e, também, estabelecendo sua direção como a de apontamento da antena (VAN DER SANDEN, 1997; REIGBER, 2001; RICHARDS, 2009). Mais especificamente, radar de abertura sintética (do acrônimo inglês *synthetic aperture radar* – SAR) é um sistema de imageamento ativo que opera na região das micro-ondas do espectro eletromagnético, geralmente entre as faixas das bandas P e Ka (Figura 2.3) (LEE e POTTIER, 2009). Por se tratar de um sistema ativo, SARs têm capacidade de imageamento diurno e noturno, penetração de nuvens e, com algumas limitações, penetração em nuvens de chuva (OUCHI, 2013).

Para fins de imageamento por sensoriamento remoto, sistemas SAR são geralmente dispostos em plataformas aéreas ou orbitais e operam sua antena em uma geometria de visada lateral, com iluminação perpendicular à linha de direção de voo. Por se tratar de um sistema que utiliza o tempo de propagação do sinal, a visada lateral é utilizada para evitar a ambiguidade da onda eletromagnética e possibilitar a geração de imagens com o posicionamento correto dos alvos. Desse modo, esses sistemas emitem pulsos na região das micro-ondas que interagem com os alvos no terreno e retornam retroespalhados para sua antena (LEE e POTTIER, 2009; RICHARDS, 2009).

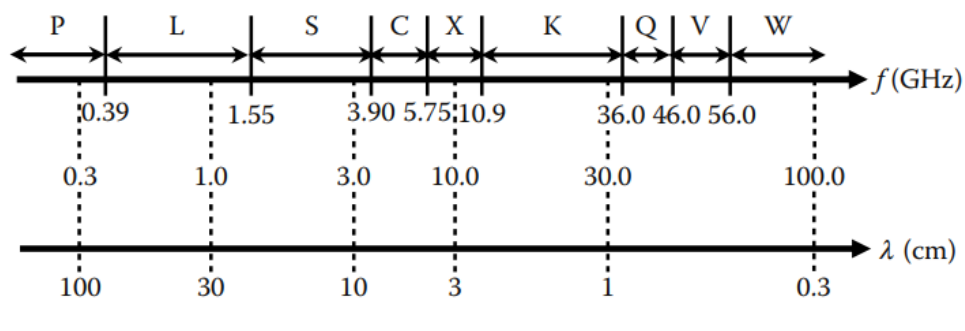


Figura 2.3. Região das micro-ondas do espectro eletromagnético. Fonte: Lee e Pottier (2009).

A mais comum configuração de imageamento é a monoestática. Nesta configuração, o transmissor e o receptor dos pulsos de micro-ondas compartilham a mesma antena (LEE e POTTIER, 2009; RICHARDS, 2009). O conhecimento do tipo de

sistema utilizado (monoestático ou biestático) é fundamental, dado que, por exemplo, implica diretamente na seleção adequada de metodologias de processamento polarimétrico e na interpretação dos resultados obtidos. Nesse sentido, cabe reforçar que nesta tese foram utilizados dados SAR polarimétricos do sistema monoestático ALOS-2/PALSAR-2.

A geometria de aquisição de imageamento de um sistema SAR é apresentada na Figura 2.4. O sistema SAR apresentado situa-se a uma altura  $H$  e apresenta deslocamento a uma velocidade  $V_{SAR}$ . A posição da antena é referida como em azimute (*along-track*) ( $y$ ). O pulso eletromagnético da antena é emitido de forma inclinada (*slant-wise*) em direção ao terreno e perfaz com este, um ângulo de incidência  $\theta$ . O eixo radial ou linha de visada do radar (*radar-line-of-sight – RLOS*) é denominado de *slant-range* ( $r$  – alcance inclinado). A área coberta pelo pulso eletromagnético da antena, em  $x$  (*ground range – alcance no terreno*) e em  $y$  (azimute), é denominada de *footprint* da antena. A plataforma se desloca e gera uma faixa de imageamento (*radar swath*) delimitada pelos alcances próximo e distante (*near and far ranges*) (LEE e POTTIER, 2009).

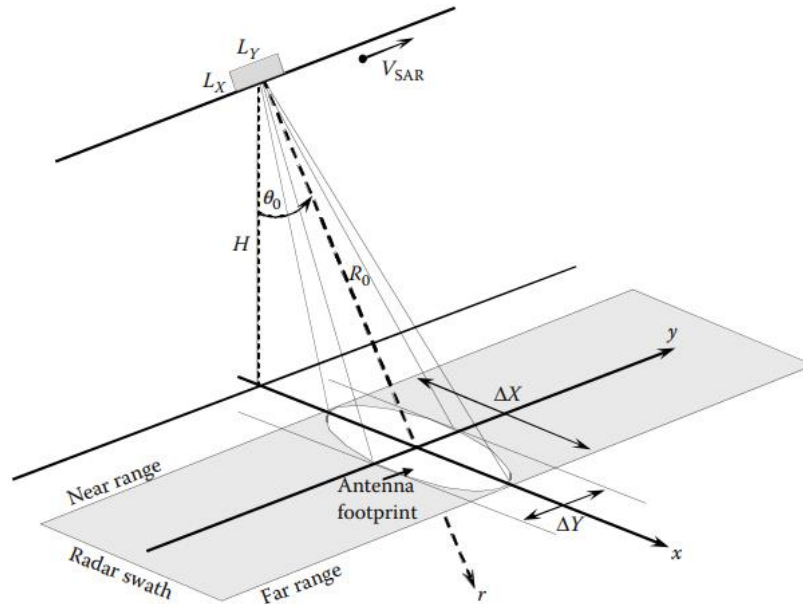


Figura 2.4. Geometria de imageamento SAR. Fonte: Lee e Pottier (2009).

Com relação à resolução em alcance, para que o sistema SAR seja capaz de ter alta resolução espacial e identificar e discriminar alvos sutis, é necessário que o pulso

eletromagnético emitido seja relativamente curto e modulado linearmente em frequência com uma alta largura de banda (Bp), denominado de *chirp* ( $\tau = 1/Bp$ ) (RICHARDS, 2009). Desse modo, a projeção do alcance inclinado no terreno ( $r_g$ ) é obtida da razão entre o tempo de duração do pulso eletromagnético comprimido ( $\tau$ ), multiplicado pela velocidade da luz ( $c$ ), e o seno do ângulo de incidência ( $\theta$ ) (RICHARDS, 2009), conforme a formulação matemática abaixo (Eq. 2.1):

$$r_g = \frac{\tau c}{2 \text{sen} \theta} \quad (2.1)$$

Por outro lado, a resolução em azimute é obtida por meio da sintetização de uma antena virtual suficientemente grande para propiciar alta resolução espacial. Disto se justifica o acrônimo SAR anteriormente explicado (LEE e POTTIER, 2009; RICHARDS, 2009). De acordo com Richards (2009), a resolução em azimute de um SAR independe da altitude da plataforma e pode ser obtida simplesmente pela metade do comprimento da antena ( $l_a$ ), conforme formulação a seguir (Eq. 2.2):

$$r_a = \frac{l_a}{2} \quad (2.2)$$

Ainda com relação à geometria de imageamento SAR apresentada na Figura 2.4, cabe destaque ao ângulo de incidência  $\theta$ , por se tratar de um parâmetro que interfere de forma significativa no retroespalhamento de uma imagem. Os valores desse ângulo aumentam gradativamente do início da faixa imageada (*near range*) em direção ao final desta (*far range*). Quanto maior for este ângulo, maior será o sombreamento. Por outro lado, quanto menor for esse ângulo, maiores serão os efeitos de deslocamento do relevo (*foreshortening* ou encurtamento de rampa e *layover* ou inversão de relevo) (REIGBER, 2001; LEE e POTTIER, 2009; RICHARDS, 2009).

Com relação à imagem SAR formada, considerando a predominância de alvos distribuídos, cada célula de resolução armazena informações de amplitude e fase da onda (dado complexo), as quais são obtidas por meio da soma vetorial coerente dos sinais de retorno de diferentes espalhadores no terreno. Isso promove um tipo de ruído característico nas imagens denominado de *speckle*. Esse ruído pode ser minimizado por meio de técnicas de filtragem e de processamento *multilook* (multivisada), ocorrendo

alguma deterioração da resolução espacial da imagem (REIGBER, 2001; LEE e POTTIER, 2009; RICHARDS, 2009).

Geralmente, para fins de visualização e interpretação, somente as imagens em amplitude são utilizadas, pois contêm o brilho e a intensidade do sinal e possibilitam a identificação dos objetos. As imagens com a fase apresentam uma distribuição aleatória e são utilizadas em aplicações de polarimetria e interferometria, por exemplo (REIGBER, 2001; LEE e POTTIER, 2009; RICHARDS, 2009).

No que tange às propriedades dos alvos imageados, rugosidade e conteúdo de umidade devem ser considerados. A rugosidade da superfície afeta o brilho das imagens e tem relação com o tipo de alvo ou cobertura do terreno. Salienta-se que ângulo de incidência e comprimento de onda devem ser considerados na análise da rugosidade (RICHARDS, 2009). De modo geral, a rugosidade pode ser classificada em três categorias: lisa (a superfície reflete a maior parte da energia incidente de forma especular); intermediária (a energia incidente é refletida pela superfície de forma difusa e não homogênea); e rugosa (a energia incidente é retroespalhada em todas as direções e de forma mais homogênea) (RICHARDS, 2009).

Por outro lado, as propriedades dielétricas dos alvos são alteradas de acordo com o teor de umidade. Isso afeta o modo como esse material aparece em uma imagem de radar. O mesmo alvo ou material pode ser identificado de forma variada em momentos ou locais diferentes, de acordo com o teor de umidade. Regra geral, elevado teor de umidade ocasiona uma constante dielétrica alta, e, por conseguinte, um alto sinal de retorno. Por outro lado, um teor de umidade baixo ocasiona uma constante dielétrica baixa e um baixo sinal de retorno (RICHARDS, 2009).

A equação de radar estabelece a relação fundamental entre as características do sistema de radar, do alvo e do sinal que é transmitido pela antena. Considerando alvos distribuídos (áreas homogêneas de florestas, por exemplo), deve-se empregar uma integração em área e da forma apresentada na seguinte formulação matemática (Eq. 2.3) (RICHARDS, 2009):

$$P_r = \iint_{pixel} \frac{P_t G_t G_r \lambda^2 \sigma^0 r_a r_g}{(4\pi)^3 R^4} \quad (2.3)$$

Em que:

$P_r$  = Potência do sinal recebido (watts);

$P_t$  = Potência do sinal transmitido (watts);

$G_t$  = Ganho da antena de transmissão;

$G_r$  = Ganho da antena de recepção;

$\lambda$  = Comprimento de onda do sistema SAR;

$\sigma^0$  = Coeficiente de retroespalhamento;

$r_a$  = resolução em azimute;

$r_g$  = resolução em alcance;

$R$  = Distância em *range* entre o radar e o alvo (considerando transmissão e emissão de energia de forma isotrópica).

$P_r$  é diretamente proporcional a  $P_t$ ,  $\sigma^0$ ,  $A$ ,  $G$ ,  $\lambda$  e a  $1/(4\pi)^3 R^4$ . O coeficiente de retroespalhamento  $\sigma^0$  consiste na razão entre a energia espalhada e o sinal original enviado pela antena. Os valores do  $\sigma^0$  são expressos em decibéis (dB) e estão diretamente relacionados aos parâmetros do alvo, enquanto os demais termos da equação de radar se relacionam com os parâmetros do sensor (LEWIS e HENDERSON, 1998; RICHARDS, 2009).

## 2.3 – Princípios básicos de polarimetria SAR

### 2.3.1 – Polarização de ondas eletromagnéticas

Nos radares, ondas eletromagnéticas podem ser transmitidas e recebidas em diferentes modos de polarização (Figura 2.5). Essas ondas podem ser apresentadas em até quatro diferentes modos de polarização: HH (transmite horizontal, recebe horizontal); VV (transmite vertical, recebe vertical); HV (transmite horizontal, recebe vertical); e VH (transmite vertical, recebe horizontal). Esses modos de polarização são obtidos por meio da filtragem dos pulsos emitidos e recebidos. Em termos práticos, polarizações HH e VV são ditas como polarizações lineares (*co-pol*) e HV e VH como polarizações cruzadas (*cross-pol*) (VAN DER SANDEN, 1997).

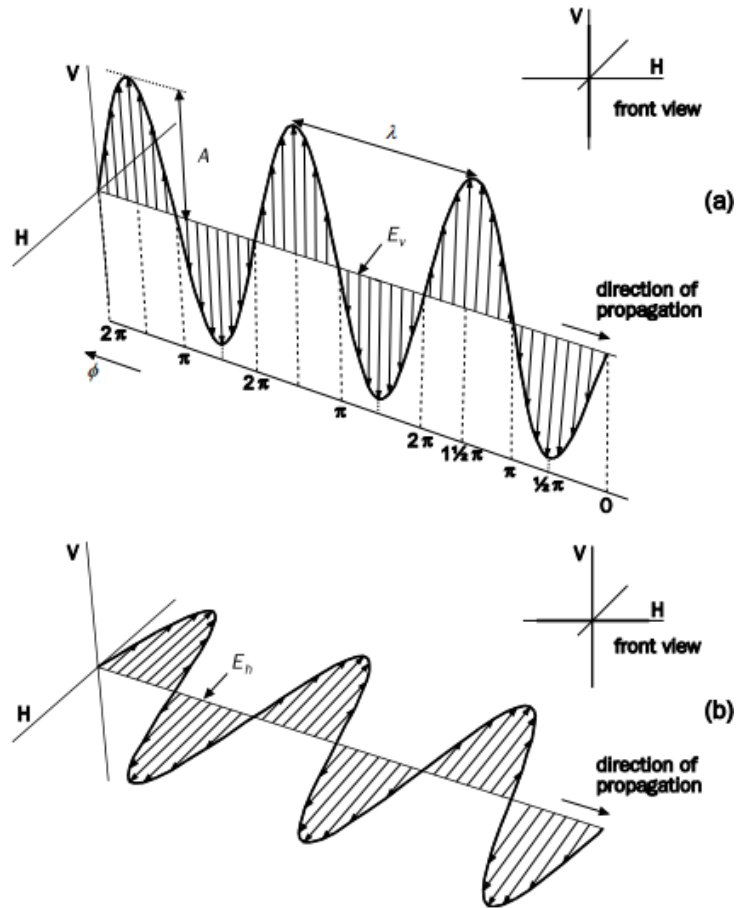


Figura 2.5. Representações lateral e frontal do campo elétrico de uma onda eletromagnética. (a) Onda com polarização vertical (V). (b) Onda com polarização horizontal (H). Ilustrações do comprimento de onda ( $\lambda$ ), amplitude ( $A$ ), fase ( $\phi$ ) variando entre 0 a  $2\pi$ , campo elétrico vertical ( $E_v$ ), campo elétrico horizontal ( $E_h$ ). Fonte: van der Sanden (1997).

Dados SAR polarimétricos apresentam maior quantidade de informação sobre os alvos terrestres, se comparados aos dados SAR polarizados convencionais (VAN DER SANDEN, 1997; OUCHI, 2013). Isso ocorre pelo fato de que a polarimetria SAR tem a capacidade de descrever e armazenar dados e informações sobre as mudanças de estado de polarização do vetor campo elétrico alterado pela estrutura e constante dielétrica de objetos na superfície terrestre (OUCHI, 2013; LEE e POTTIER, 2009).

De acordo com Richards (2009), o vetor campo elétrico pode ser definido pela seguinte forma exponencial complexa (Eq. 2.4):

$$\vec{E}(\vec{r}, t) = \vec{E} \exp(j(\omega t - \vec{k} \cdot \vec{r})) \quad (2.4)$$



Em que:

$\vec{E}$  = vetor campo elétrico;

$\vec{r}$  = vetor posição (distância);

t = tempo;

$\omega$  = frequência angular da onda elétrica; e

$\vec{k}$  = vetor de propagação (a direção da onda).

Em situações em que os alvos se encontram a grandes distâncias da plataforma e fonte de emissão, a onda eletromagnética pode ser assumida como plana. Nesse caso,  $\vec{E}$  (amplitude constante) se propaga como uma onda de frente plana e, então, pode ser definido sobre um sistema plano de referência (Figura 2.6). Nessa configuração,  $\vec{k}$  se apresenta perpendicular ao referido plano e  $\vec{E}$  pode ser decomposto nas componentes horizontal e vertical, conforme equação abaixo (Eq. 2.5) (HELLMANN, 2001; RICHARDS, 2009).

$$\vec{E}(\vec{r}, t) = \vec{E}_h(\vec{r}, t) + \vec{E}_v(\vec{r}, t) \quad (2.5)$$

Em que:

$\vec{E}$  = vetor campo elétrico;

$\vec{r}$  = vetor posição (a distância);

t = tempo;

$\vec{E}_h$  = vetor campo elétrico horizontal; e

$\vec{E}_v$  = vetor campo elétrico vertical.

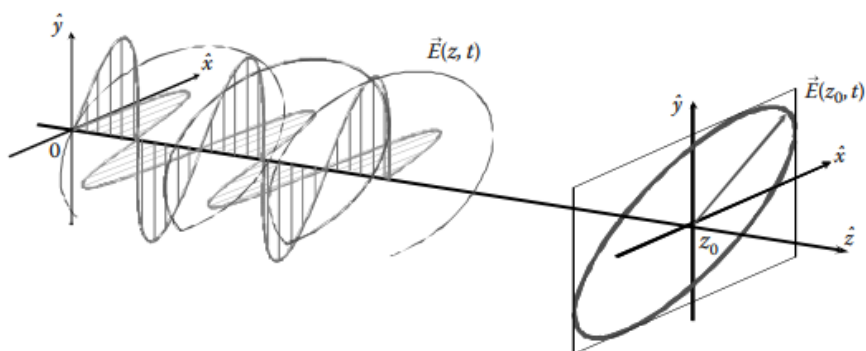


Figura 2.6. Trajetória temporal de uma onda monocromática plana e em dada direção fixa. Fonte: Lee e Pottier (2009).

Ainda de acordo com Hellmann (2001), Lee e Pottier (2009) e Richards (2009), a forma mais comum de se representar  $\vec{E}$  é a equação de uma elipse (Eq. 2.6):

$$\frac{\vec{E}_h(\vec{r},t)^2}{a_h^2} + \frac{\vec{E}_v(\vec{r},t)^2}{a_v^2} - 2\cos\delta \frac{\vec{E}_h(\vec{r},t)\vec{E}_v(\vec{r},t)}{a_h a_v} = \sin^2\delta \quad (2.6)$$

Em que:

$a_h$  = amplitude na componente horizontal;

$a_v$  = amplitude na componente vertical;

$\vec{E}_h$  = vetor campo elétrico horizontal;

$\vec{E}_v$  = vetor campo elétrico vertical; e

$\delta = \delta_v - \delta_h$  (diferença de fase).

Por sua vez, a elipse de polarização (Figura 2.7) descreve o estado de polarização de  $\vec{E}$ . Há duas propriedades fundamentais dessa forma geométrica que estão diretamente relacionadas ao estado de polarização da radiação. A primeira propriedade é a elipticidade ou excentricidade (razão entre os semieixos maior e menor ou  $f/e$ ) que descreve o quão diferente a elipse em questão é de um círculo ou de uma linha reta. A outra propriedade representa a inclinação (ângulo  $\tau$ ) da elipse em relação ao plano horizontal (RICHARDS, 2009). No caso de sistemas monoestáticos, utiliza-se o sistema de coordenadas *backscatter alignment* (BSA – alinhamento de retroespalhamento), o qual aponta na direção de deslocamento da onda e com giro para o lado direito (regra da mão direita) (LEE e POTTIER, 2009).

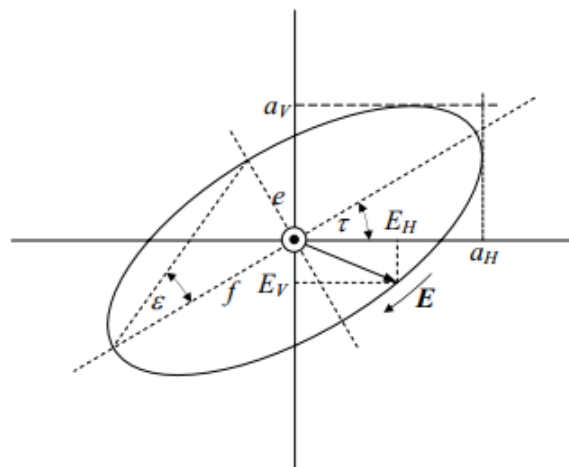


Figura 2.7. Elipse de polarização para representação do vetor campo elétrico. Fonte: Richards (2009).

### 2.3.2 – Matriz de espalhamento S

A matriz de espalhamento S (ou matriz de Jones) consiste em uma transformação linear, que descreve as alterações ocorridas entre o vetor elétrico transmitido pela antena SAR e o vetor elétrico retroespalhado pelos alvos da superfície imageada. Essa matriz armazena toda a informação relativa ao espalhamento e, por conseguinte, permite a descrição do alvo e dos processos físicos atuantes (HELLMANN, 2001). Pode-se utilizar uma representação vetorial contendo as informações da matriz de espalhamento S (Eq. 2.7) (HELLMANN, 2001; LEE e POTTIER, 2009).

$$[S] = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \rightarrow \vec{k} = \frac{1}{2} \text{Traço}([S]\Psi) = [k_0, k_1, k_2, k_3]^T \quad (2.7)$$

Em que:

[S] = matriz de espalhamento S;

$S_{ij}$  = elemento complexo de [S];

$k_i$  = elemento complexo do vetor de Jones ( $\vec{k}$ );

Traço [S] = a soma dos elementos da diagonal de [S]; e

$\Psi$  = base matricial 2x2.

Considerando a situação de retroespalhamento em configuração monoestática e, também, o teorema da reciprocidade ( $S_{hv} = S_{vh}$ ), para  $\Psi$ , utiliza-se a base de Pauli (Eq. 2.8) (HELLMANN, 2001; LEE e POTTIER, 2009):

$$\Psi_P = \left\{ \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \quad (2.8)$$

Por sua vez, o vetor de Pauli é dado por (Eq. 2.9) (HELLMANN, 2001; LEE e POTTIER, 2009):

$$\vec{K}_P = \frac{1}{\sqrt{2}} [(S_{hh}+S_{vv})(S_{hh}-S_{vv})(2S_{hv})]^T \quad (2.9)$$

Uma das bases também utilizadas para  $\Psi$  é a base de Borgeaud (Eq. 2.10):

$$\Psi_B = \left\{ 2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, 2\sqrt{2} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, 2 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\} \quad (2.10)$$

A base de Borgeaud causa uma expansão lexicográfica da matriz de espalhamento  $S$ . O vetor de Borgeaud é dado por (Eq. 2.11) (HELLMANN, 2001; LEE e POTTIER, 2009):

$$\vec{K}_B = [S_{hh}\sqrt{2}S_{hv}S_{vv}]^T \quad (2.11)$$

As transformações de bases de Pauli e de Borgeaud são empregadas respectivamente na geração das matrizes de covariância e de coerência a serem detalhadas na sequência deste documento.

### 2.3.3 – Matrizes de covariância e de coerência

Lee e Pottier (2009) apresentaram o conceito de alvos distribuídos, ou seja, em imageamento SAR nem todos os alvos são estacionários ou fixos. Muitos alvos variam ao longo do tempo. Alvos naturais, por exemplo, modificam-se por conta do vento e de alterações nos gradientes de temperatura e pressão. Nesse sentido, a onda plana retroespalhada para a antena do sistema SAR deve ser considerada parcialmente polarizada. Isso significa que esta onda não mais se encontra coerente, monocromática e totalmente polarizada na forma descrita pela elipse de polarização (LEE e POTTIER, 2009). Deste modo, a matriz de espalhamento  $S$  apresenta limitações na representação desses alvos incoerentes, sendo necessária a utilização das matrizes de covariância e de coerência (e seus autovetores e autovalores), também denominadas de matrizes de potência, para o entendimento de áreas com predomínio de alvos distribuídos (HELLMANN, 2001; LEE e POTTIER, 2009).

A matriz de covariância polarimétrica é formada pelo produto vetorial entre o vetor de Borgeaud e seu complexo conjugado transposto (Eq. 2.12) (HELLMANN, 2001):

$$[C]_{4 \times 4} = \langle \vec{k}_B \vec{k}_B^{*T} \rangle = \begin{bmatrix} \langle |S_{hh}|^2 \rangle & \langle S_{hh} S_{hv}^* \rangle & \langle S_{hh} S_{vh}^* \rangle & \langle S_{hh} S_{vv}^* \rangle \\ \langle S_{hv} S_{hh}^* \rangle & \langle |S_{hv}|^2 \rangle & \langle S_{hv} S_{vh}^* \rangle & \langle S_{hv} S_{vv}^* \rangle \\ \langle S_{vh} S_{hh}^* \rangle & \langle S_{vh} S_{hv}^* \rangle & \langle |S_{vh}|^2 \rangle & \langle S_{vh} S_{vv}^* \rangle \\ \langle S_{vv} S_{hh}^* \rangle & \langle S_{vv} S_{hv}^* \rangle & \langle S_{vv} S_{vh}^* \rangle & \langle |S_{vv}|^2 \rangle \end{bmatrix} \quad (2.12)$$

Em que:

$[C]_{4 \times 4}$  = matriz 4x4 de covariância (elementos da diagonal representam a intensidade do sinal na respectiva polarização);

$\langle \dots \rangle$  = média espacial;

$k_B$  = vetor de Borgeaud; e

$\vec{k}_B^{*T}$  = complexo conjugado transposto do vetor de Borgeaud.

Ainda de acordo com Hellmann (2001), outro conceito ou forma de tratar a natureza não-determinística ou incoerente da onda retroespalhada consiste na matriz de coerência polarimétrica. Essa matriz é formada pelo produto vetorial entre o vetor de Pauli e o seu complexo conjugado (Eq. 2.13):

$$[T]_{4 \times 4} = \langle \vec{k}_P \vec{k}_P^{*T} \rangle \quad (2.13)$$

Dado o teorema de reciprocidade, tem-se que  $S_{hv} = S_{vh}$ . Nesse caso, o vetor de Pauli pode ser reduzido a três dimensões sem que haja perda de informações (Eq. 2.14) (HELLMANN, 2001):

$$[T]_{3 \times 3} = \langle \vec{k}_{P_3} \vec{k}_{P_3}^{*T} \rangle = \frac{1}{2} \begin{bmatrix} \langle |A|^2 \rangle & AB^* & AC^* \\ \langle A^* B \rangle & \langle |B|^2 \rangle & BC^* \\ \langle A^* C \rangle & \langle B^* C \rangle & \langle |C|^2 \rangle \end{bmatrix} \quad (2.14)$$

Em que:

$[T]_{3 \times 3}$  = matriz 3x3 de coerência;

$k_P$  = vetor de Pauli;

$\vec{k}_P^{*T}$  = complexo conjugado transposto do vetor de Pauli;

$A = S_{hh} + S_{vv}$ ;

$B = S_{hh} - S_{vv}$ ; e

$C = 2S_{hv}$ .

### 2.3.4 – Teoremas de decomposição de alvos

A técnica de decomposição de alvos tem por objetivo representar o mecanismo médio de espalhamento como uma soma de elementos independentes, associando, a cada um desses elementos, um mecanismo físico de espalhamento (CLOUDE e POTTIER, 1996). De acordo com Cloude e Pottier (1996) e Hellmann (2001), os teoremas de decomposição de alvos podem ser classificados em dois grandes grupos: decomposições do tipo coerentes e decomposições do tipo incoerentes.

As decomposições coerentes assumem a ocorrência de espalhadores determinísticos e, por conseguinte, que a onda retroespalhada é polarizada. Geralmente, esse tipo de decomposição de alvos utiliza a matriz de espalhamento  $S$  para representar os estados de polarização da onda retroespalhada. Por outro lado, as decomposições incoerentes consideram *a priori* que o espalhamento não é do tipo determinístico e, portanto, a onda retroespalhada é parcialmente polarizada. Nessa situação, utilizam-se as matrizes de covariância e de coerência para caracterizar a onda retroespalhada (LEE e POTTIER, 2009; RICHARDS, 2009).

Geralmente, nas aplicações de sensoriamento remoto, por se tratarem de áreas e alvos distribuídos, a premissa de ocorrência de alvos determinísticos puros não é válida (RICHARDS, 2009). Desse modo, utilizam-se com frequência as matrizes de reflexão de potência já destacadas. Com base nesse entendimento, nesta tese, utilizaram-se somente métodos incoerentes de decomposição de alvos, sendo empregados, assim, os seguintes teoremas de decomposição de alvos: van Zyl (com três componentes); Freeman-Durden (com três componentes); Yamaguchi (com quatro componentes); e Entropia (H)-Anisotropia (A)-ângulo  $\alpha$  (teorema de Cloude-Pottier).

#### 2.3.4.1 – Teorema de van Zyl

A decomposição de van Zyl é constituída de três componentes responsáveis por uma descrição geral dos tipos de espalhamentos predominantes com base na matriz de covariância  $3 \times 3$  para terrenos naturais azimutalmente simétricos (reflexão simétrica) e na situação de imageamento monoestático. As componentes de espalhamento são: *even number of reflections* (número inteiro de reflexões divisível por 2); *odd number of reflections* (número inteiro de reflexões sem divisão exata por 2); e *diffuse scattering*

(espalhamento difuso) (VAN ZYL, 1989; FREEMAN e DURDEN, 1998; LEE e POTTIER, 2009).

De acordo com van Zyl (1989), com base em dados de área teste em São Francisco (Califórnia), observou-se que a componente *even number of reflections* estava mais associada a áreas urbanas, enquanto a componente *odd number of reflections* predominou em áreas de oceano. Por fim, as áreas vegetadas apresentaram mais associação com o espalhamento do tipo difuso (*difuse scattering*). Freeman e Durden (1998) destacaram que uma das principais limitações do teorema de van Zyl é a descrição puramente matemática do espalhamento, sem associação ou descrição do tipo físico de espalhamento predominante. Desse modo, com base no modelo de van Zyl (1989), Freeman e Durden (1998) apresentaram um novo modelo, o qual é descrito na próxima seção deste documento.

#### **2.3.4.2 – Teorema de Freeman-Durden**

A decomposição de alvos de Freeman-Durden foi desenvolvida principalmente para a interpretação e entendimento do retroespalhamento de alvos florestais, nos quais o modelo de reflexão simétrica se adapta adequadamente. Desse modo, essa decomposição tem-se demonstrado útil na discriminação entre áreas florestadas e não-florestadas, floresta inundada e não-inundada, contribuindo ainda na estimativa dos efeitos da floresta inundada nas assinaturas polarimétricas dos alvos (FREEMAN e DURDEN, 1998; LEE e POTTIER, 2009).

Essa decomposição de alvos modela a matriz de covariância como sendo a contribuição de três mecanismos de espalhamento (FREEMAN e DURDEN, 1998; LEE e POTTIER, 2009; RICHARDS, 2009):

- i) espalhamento volumétrico: por exemplo, as copas das árvores são modeladas como um conjunto de espalhadores do tipo dipolo orientados aleatoriamente;
- ii) *double bounce* (batida dupla ou par): espalhadores modelados como refletores do tipo diedro;
- iii) superficial: espalhadores modelados como do tipo *Bragg*, que ocorrem em superfícies ligeiramente rugosas e em superfícies onduladas (superfícies oceânicas, por exemplo).

De acordo com a proposta de Freeman e Durden (1998):

i) a potência correspondente ao espalhamento do tipo volumétrico é dada por (Eq. 2.15):

$$P_v = \frac{8}{3} f_v \quad (2.15)$$

ii) a potência correspondente ao espalhamento do tipo *double bounce* é dada por (Eq. 2.16):

$$P_d = f_d(1 + |S_{hh} + S_{vv}|^2) \quad (2.16)$$

iii) e a potência correspondente ao espalhamento do tipo superficial é dada por (Eq. 2.17):

$$P_s = f_s(1 + |S_{vv} + S_{hh}|^2) \quad (2.17)$$

Em que:

$P$  = potência;

$f$  = contribuição da componente de espalhamento;

$S_{hh}$  = elemento complexo da matriz de espalhamento S (polarização HH); e

$S_{vv}$  = elemento complexo da matriz de espalhamento S (polarização VV).

### 2.3.4.3 – Teorema de Yamaguchi

Yamaguchi et al. (2005) desenvolveram um modelo para decomposição de alvos com base na modelagem da matriz de covariância e constituído de quatro componentes de espalhamento: *double bounce* (batida dupla), superficial, volumétrico e hélice. Os três primeiros componentes baseiam-se nos seus correspondentes apresentados no modelo de Freeman e Durden (1998). Já o quarto componente foi introduzido por Yamaguchi et al. (2005) e modela o espalhamento do tipo hélice que se trata de um alvo com forma e/ou estrutura artificial, isto é, refletor não simétrico capaz de gerar polarizações circulares para a esquerda ou para a direita e é mais adequado para descrição de alvos não-naturais.

Yamaguchi et al. (2005) realizaram ainda uma alteração na modelagem do espalhamento volumétrico com o intuito de melhor representar o espalhamento em ambientes florestados. O modelo teórico convencional do espalhamento volumétrico consiste em uma nuvem de dipolos de orientação aleatória com modelagem baseada em uma função de densidade de probabilidade uniforme e com diferentes ângulos de orientação. Todavia, em ambientes predominadas por estruturas verticais (como vegetação florestal), os espalhamentos de troncos e galhos apresentam ângulos



preferenciais de distribuição. Desse modo, Yamaguchi et al. (2005) realizaram uma modificação na função de densidade de probabilidade referente ao modelo de Freeman-Durden (1998), considerando-se o aspecto da orientação dominante de troncos e ramos de árvores.

#### 2.3.4.4 – Teorema de Cloude-Pottier (H-A- $\alpha$ )

O teorema de Cloude-Pottier analisa os autovalores e os autovetores da matriz de coerência, propondo uma decomposição dessa matriz em elementos que representam espalhamentos parciais, sendo assim capaz de abranger todos os tipos de mecanismos de espalhamento. O enfoque nas análises dos autovalores e autovetores da matriz de coerência traz a vantagem de independe da base utilizada e de ser sempre diagonalizável (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009). Na proposta de Cloude e Pottier (1997), a matriz de coerência é decomposta na soma de outras três matrizes parciais de coerência, sendo que cada matriz individual é ponderada pelo seu respectivo autovalor. Deste modo, tem-se (Eq. 2.18) (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009):

$$[T_3] = \sum_{i=1}^3 T_{3_i} = \sum_{i=1}^3 \lambda_i (\vec{e}_i \cdot \vec{e}_i^{*T}) = \lambda_1 (\vec{e}_1 \cdot \vec{e}_1^{*T}) + \lambda_2 (\vec{e}_2 \cdot \vec{e}_2^{*T}) + \lambda_3 (\vec{e}_3 \cdot \vec{e}_3^{*T}) \quad (2.18)$$

Em que:

$[T_3]$  = matriz de coerência 3 x 3;

$\lambda_i$  = autovalor; e

$e_i$  = autovetor.

Cada uma das matrizes individuais representa uma contribuição de um tipo específico de espalhamento determinístico. O peso de cada contribuição é dado pelo autovalor, enquanto o tipo de mecanismo de espalhamento é dado pelo autovetor. Os principais indicadores extraídos desse teorema de decomposição de alvos são: entropia (H), anisotropia (A) e a medida do ângulo  $\alpha$  (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009).

#### 2.3.4.4.1 – Entropia polarimétrica (H)

A entropia polarimétrica indica o grau de aleatoriedade da distribuição dos mecanismos de espalhamento (Eq. 2.19) (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009):

$$H = -\sum_{i=1}^3 P_i \log_3 P_i \quad (2.19)$$

Em que:

$$P_i = \frac{\lambda_i}{\sum_{k=1}^3 \lambda_k}$$

Os valores de  $P_i$  são interpretados como a intensidade relativa do processo de espalhamento  $i$ . Os valores de H variam entre 0 e 1.  $H = 0$  indica que a matriz de coerência tem apenas um autovalor ( $\lambda_i$ ) diferente de 0, representando apenas um processo determinístico de espalhamento. Por outro lado,  $H = 1$  significa que os autovalores da matriz de coerência são todos iguais, indicando um processo de espalhamento do tipo ruído aleatório que despolariza totalmente a onda incidente (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009).

#### 2.3.4.4.2 – Anisotropia polarimétrica (A)

A anisotropia polarimétrica é um indicador da importância relativa dos mecanismos de espalhamento secundários, sendo mais representativo quando ocorre média entropia (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009). A anisotropia polarimétrica é dada por (Eq. 2.20):

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (2.20)$$

Segundo Cloude e Pottier (1997) e Lee e Pottier (2009), para alta entropia, a anisotropia não traz nenhuma informação adicional, dado que os autovalores são aproximadamente iguais. Para baixa entropia, os auto-valores são próximos a zero. A

média entropia significa que mais de um mecanismo de espalhamento contribui para o retroespalhamento, mas não esclarece quantos mecanismos. A anisotropia fornece ainda informações adicionais, tais como: a alta anisotropia, que indica que apenas o segundo mecanismo de espalhamento é importante; e a baixa anisotropia, que indica que existe também a contribuição do terceiro mecanismo de espalhamento (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009).

#### 2.3.4.4.3 – Ângulo $\alpha$

A partir da matriz de coerência, cada autovetor pode também ser parametrizado em termos de cinco ângulos:  $\alpha$ ,  $\beta$ ,  $\varphi$ ,  $\delta$  e  $\gamma$ , com destaque para o ângulo  $\alpha$ , com variação entre  $0^\circ$  e  $90^\circ$ , uma vez que este identifica o tipo de mecanismo físico de espalhamento do alvo. Para  $\alpha$  próximo a  $0^\circ$ , predomina espalhamento isotrópico *odd bounce* (batida ímpar), em  $45^\circ$ , predomina o espalhamento do tipo dipolo e, em  $90^\circ$ , há o predomínio de espalhamento do tipo *double bounce* (batida dupla ou par) (CLOUDE e POTTIER, 1997; LEE e POTTIER, 2009).

### 2.4 – Classificador supervisionado polarimétrico Wishart

O classificador supervisionado polarimétrico Wishart (PW) utiliza a matriz de covariância para extrair medidas de distância a partir da distribuição de probabilidades multivariada Wishart. Com base em amostras coletadas em dados polarimétricos, *pixels* podem ser atribuídos a uma classe que apresente a menor medida de distância Wishart. O classificador PW opera igualmente em imagens polarimétricas *single* ou *multilook*, independente do número de polarizações ou mesmo da aplicação ou não de filtros para remoção do ruído *speckle* (LEE e POTTIER, 2009).

De acordo com Lee e Grunes (1994) e Lee e Pottier (2009), para dados polarimétricos em que se assumem classes com iguais probabilidades de ocorrência, as distâncias podem ser calculadas utilizando-se a Eq. 2.21:

$$d_3(Z, \omega_m) = n \ln|C_m| + n \text{Tr}(C_m^{-1}Z) \quad (2.21)$$

Em que:

$d_3(Z, \omega_m)$  = medida de distância Wishart;

$Z$ : = estimador de máxima verossimilhança da matriz de covariância do *pixel*;

$\omega_m$  = estimador da matriz de covariância da classe  $m$ ;

$C_m$  = matriz de covariância do centro da classe  $m$ ;

$Tr(C_m^{-1}Z)$  = matriz transposta de  $(C_m^{-1}Z)$ .

## 2.5 – Conceitos básicos sobre mineração de dados

Mineração de dados (MD) consiste em uma nova fronteira do conhecimento no âmbito das ciências da computação e também pode ser denominada de descoberta de conhecimento em bases de dados (DCBD) (do inglês *Knowledge Discovery in Databases - KDD*). De forma ampla, MD abrange um conjunto de técnicas e tecnologias computacionais e métodos para análise automática e inteligente de grandes volumes de dados, visando à extração de padrões (HAN et al., 2012). Apesar de ser elemento fundamental no processo DCDB, a MD é elencada também como uma etapa nesse fluxo analítico, conforme destacaram Han et al. (2012). A Figura 2.8 ilustra o DCDB e apresenta a MD como uma importante etapa.

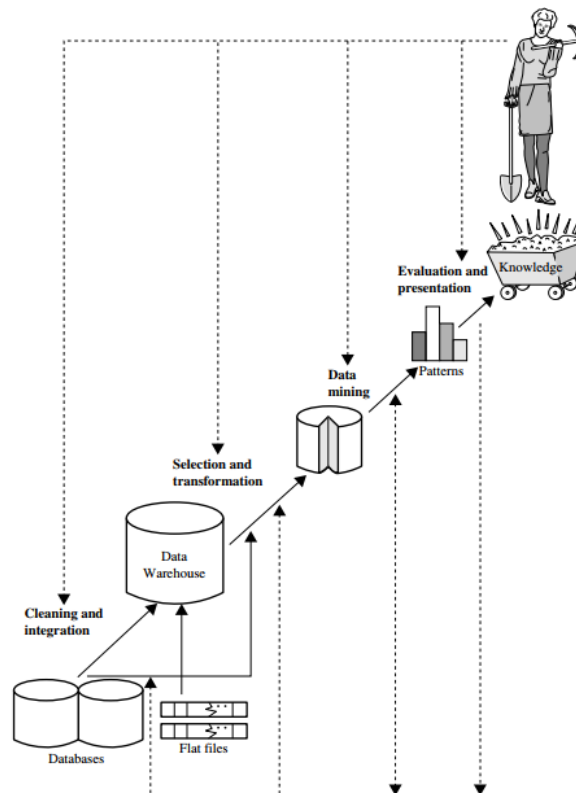


Figura 2.8. Etapas do processo descoberta de conhecimento em bases de dados (DCBD): limpeza e integração; seleção e transformação; mineração de dados; e avaliação e apresentação. Fonte: Han et al. (2012).

Hsu et al. (2002) apresentaram uma abordagem da MD voltada ao sensoriamento remoto (*image mining* – mineração de imagens) (Figura 2.9) e destacaram que, nesse caso, a mineração de dados deve explorar outros aspectos, como processamento e representação em diferentes níveis, técnicas de reconhecimento de objetos e padrões, técnicas de recuperação e indexação de imagens, técnicas de classificação e agrupamento, regras de associação e inteligência artificial.

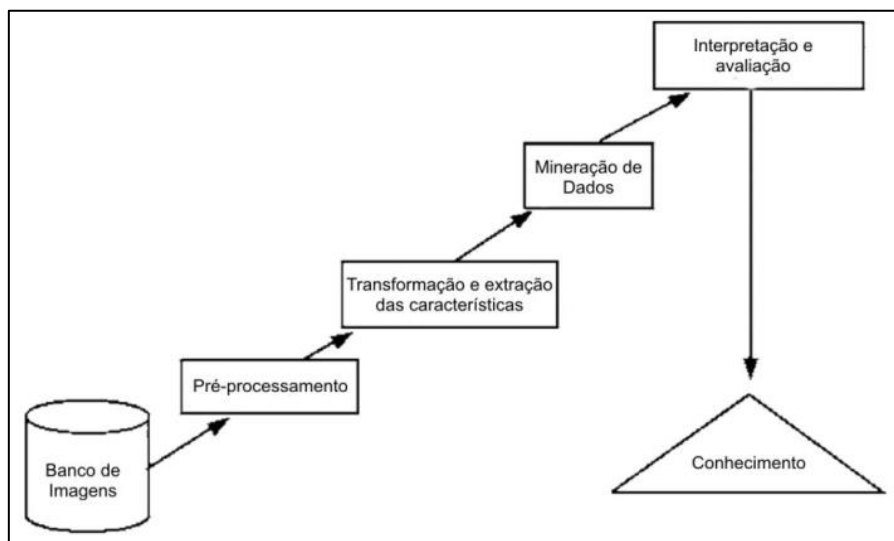


Figura 2.9. Fluxo de processos DCBD em imagens de sensoriamento remoto. Fonte: adaptado de Hsu et al. (2002).

Nesta tese, para fins de mineração de dados e reconhecimento de padrões nas imagens ALOS-2/PALSAR-2, utilizou-se o pacote de mineração de dados WEKA (*Waikato Environment for Knowledge Analysis* – Waikato Ambiente para Análise do Conhecimento) e os seguintes algoritmos de classificação: *naive* Bayes (NB), árvore de decisão J48 (AD J48), *random forest* (Floresta Aleatória - RF), rede neural artificial *multilayer perceptron* (Perceptron Multicamadas - MLP) e *support vector machine* (Máquina de Suporte Vetorial - SVM). O SVM corresponde a um classificador com otimização e treinamento por meio do algoritmo *sequential minimal optimization* (Otimização Sequencial Mínima - SMO).

### 2.5.1. *Software* WEKA e classificadores utilizados

O *software* WEKA consiste em um ambiente computacional de domínio público e é composto de diversos algoritmos de aprendizado de máquina específicos para tarefas de MD. O WEKA foi desenvolvido pela Universidade de Waikato, na Nova Zelândia. Os algoritmos disponíveis nesse ambiente foram implementados em linguagem Java, e os códigos são abertos (WITTEN e FRANK, 2005; BOUCKAERT et al., 2010).

Com relação aos algoritmos de classificação utilizados, o algoritmo de classificação NB emprega o teorema desenvolvido no século XVIII pelo clérigo Inglês Thomas Bayes. Esse teorema trata de probabilidade e teoria da decisão e emprega o conceito de probabilidade condicional para previsão de eventos (WITTEN e FRANK, 2005; HAN et al., 2012). O classificador NB assume os atributos utilizados para uma dada classe como independentes (*class-conditional independence*) de modo a simplificar e agilizar os processamentos e cálculos envolvidos na classificação. Por causa dessa característica, é denominado de *naive*. Esse classificador apresenta bom desempenho se os atributos forem independentes e pode ser tão preciso quanto as ADs e algumas redes neurais. Além disso, apresenta flexibilidade para tratar de ausência de dados numéricos (WITTEN e FRANK, 2005; HAN et al., 2012).

A AD J48 consiste em uma representação gráfica computacional aplicada a tarefas interativas e iterativas de aprendizado de máquina e emprega diferentes conjuntos de dados e instâncias independentes. Ela utiliza a abordagem “dividir-e-conquistar” (*divide-and-conquer*), também denominada de indução *top-down* (de cima para baixo) das ADs (WITTEN e FRANK, 2005).

Uma AD é constituída de nós, nós-folhas e ramos. Os nós testam atributos de instâncias que geralmente são comparados com determinadas constantes. Nós-folhas se encontram nas extremidades de uma AD e são responsáveis por atribuir classes a todas as instâncias que alcançam as folhas. A seleção dos atributos e seus limiares para a divisão dos ramos fundamenta-se no cálculo do ganho de informação ou entropia. O atributo que apresentar maior valor de ganho ou entropia será utilizado na divisão do ramo (SILVA et al., 2008; SHARMA et al., 2013). Conforme já destacado, no *software* WEKA 3.8, está disponível o algoritmo AD J48, o qual consiste na recodificação em linguagem Java do indutor C4.5, originalmente escrito em linguagem C (BOUCKAERT et al., 2010).

O RF é um classificador do tipo *ensemble* (de conjunto) e *bagging* (ensacamento). *Ensemble* porque reúne e emprega um conjunto de classificadores do tipo *Classification and Regression Trees* (Árvores de Regressão e Classificação - CART) e realiza diversas classificações. *Bagging* porque as árvores são criadas a partir de subconjuntos do mesmo conjunto amostral, com reposição. Os usuários devem previamente definir o número de atributos e o número de árvores. Cada árvore realiza sua classificação e, ao final, a cada instância, será atribuída a respectiva classe com o maior número de votos oriundos de cada árvore (BELGIU e DRAGUT, 2016).

Por sua vez, um perceptron é um tipo de rede neural artificial que consiste em um hiperplano contido em um espaço de instâncias. Inspirado no funcionamento do cérebro humano, o algoritmo MLP emprega uma grande quantidade de neurônios artificiais interconectados de forma a decompor problemas complexos em problemas menores. Esses problemas menores são solucionados no nível do neurônio artificial, que corresponde a uma unidade de processamento virtual (WITTEN e FRANK, 2005).

As redes neurais artificiais são constituídas de uma camada de entrada, nenhuma, uma ou várias camadas escondidas e uma camada de saída. Segundo Witten e Frank (2005), o grande desafio está no aprendizado, ou seja, no treinamento da estrutura artificial de neurônios, o que inclui o ajuste iterativo dos pesos dos nós entre as camadas internas. O algoritmo MLP utiliza o método *backpropagation* (retropropagação) para treinamento da rede. A partir dos resultados da camada de saída, esse método calcula os erros e ajusta os pesos dos nós em cada neurônio artificial. Esse ajuste é feito com base no algoritmo de otimização *gradient descent* (gradiente descendente) que emprega funções sigmoidais (KASPERCZUK e DARDZIŃSKA, 2016).

SVMs são algoritmos de classificação que têm a capacidade de gerar variados hiperplanos e, por conseguinte, separações não-lineares em um espaço multidimensional de atributos. O algoritmo SVM emprega um modelo linear que busca o hiperplano marginal máximo, de modo a encontrar a melhor separação possível entre instâncias. As instâncias mais próximas ao hiperplano marginal máximo são denominadas de *support vectors* (vetores de suporte). Para a construção de hiperplanos não-lineares, os SVMs utilizam funções *kernel* (polinomial, gaussiana e sigmoideal) para transformar um espaço de atributos linear para outro não-linear. O algoritmo SMO é utilizado para o treinamento, sendo responsável pela resolução do problema de programação quadrática que surge ao se utilizar SVMs (WITTEN e FRANK, 2005; HAN et al., 2012).

## REFERÊNCIAS

- BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.
- BOUCKAERT, R. R.; FRANK, E.; HALL, M. A.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. WEKA - experiences with a Java open-source project. **Journal of Machine Learning Research**, v. 11, p. 2533–2541, 2010.
- CLOUDE, S. R.; POTTIER, E. A review of target decomposition theorems in radar polarimetry. **IEEE Transactions on Geoscience and Remote Sensing**, v. 34, n. 2, p. 498–518, 1996.
- CLOUDE, S. R.; POTTIER, E. An entropy based classification scheme for land applications of polarimetric SAR. **IEEE Transactions on Geoscience Remote Sensing**, v. 35, n. 1, p. 68–78, 1997.
- EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **As principais fitofisionomias do Cerrado. 2018.** Disponível em: <<https://www.embrapa.br/cerrados/colecao-entomologica/bioma-cerrado>>. Acesso em: 14 ago. 2018.
- FREEMAN, A.; DURDEN, S. L. A three-component scattering model for polarimetric SAR Data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 3, p. 963–973, 1998.
- HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3<sup>a</sup> ed., Whaltan: Elsevier, 2012. 744 p.
- HSU, W.; LEE, M. L.; ZHANG, J. Image mining: trends and developments. **Journal of Intelligent Information Systems**, v. 19, n. 1, p. 7–23, 2002.
- HELLMANN, M. P. SAR polarimetry tutorial. Rennes: University of Rennes, 2001. 109 p. Disponível em: <<http://epsilon.nought.de/>>. Acesso em: 01 fev. 2017.
- KASPERCZUK, A.; DARDZIŃSKA, A. Comparative evaluation of the different data mining techniques used for the medical database. **Acta Mechanica et Automatica**, v. 10, n. 3, p. 233–238, 2016.
- LEE, J.; GRUNES, M. R. Classification of multilook polarimetric SAR imagery based on the complex Wishart distribution. **International Journal of Remote Sensing**, v. 15, n.11, p. 2.299-2.311, 1994.
- LEE, J.; POTTIER, E. **Polarimetric Radar Imaging: from Basics to Applications**. Nova York: CRC Press, 2009, 398 p.
- LEWIS, A. J.; HENDERSON, F. M. Radar fundamentals: the geoscience perspective. In: HENDERSON, F. M.; LEWIS, A. J. **Manual of Remote Sensing: Principles and**



**Applications of Imaging Radars.** 3<sup>a</sup> ed., Nova York: John Wiley & Sons, p. 131–181, 1998.

OUCHI, K. Recent trend and advance of synthetic aperture radar with selected topics. **Remote Sensing**, v. 5, n. 2, p. 716–807, 2013.

REIGBER, A. **Synthetic Aperture Radar: Basic Concepts and Image.** Rennes: University of Rennes, 2001. Disponível em: <<http://epsilon.nought.de/>>. Acesso em: 05 nov. 2016.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fito-fisionomias do Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. (Eds.). **Cerrado: Ecologia e Flora**, Planaltina: Embrapa Cerrados, 2008, p. 151–199.

RICHARDS, J. A. **Remote Sensing with Imaging Radar.** Berlim: Springer, 2009. 361 p.

SANO, E. E.; ROSA, R.; BRITO, J. L. S.; FERREIRA, L. G. Land cover mapping of the tropical savanna region in Brazil. **Environmental Monitoring and Assessment**, v. 166, p. 116–124, 2010.

SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; SOUZA, R. C. M. Remote-sensing image mining: Detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, v. 29, n. 16, p. 4803–4822, 2008.

SHARMA, R.; GHOSH, A.; JOSHI, P. K. Decision tree approach for classification of remotely sensed satellite data using open source support. **Journal of Earth System Science**, v. 122, n. 5, p. 1237–1247, 2013.

VAN DER SANDEN, J. J. **Radar remote sensing to support tropical forest management** (tese de doutorado em Agronomia), 1997, Wageningen: Wageningen Agricultural University, 1997, 330 p.

VAN ZYL, J. J. Unsupervised classification of scattering behavior using radar polarimetry data. **IEEE Transactions on Geoscience Remote Sensing**, v. 27, n. 1, p. 36–45, 1989.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning. Tools and Techniques.** 2<sup>a</sup> ed., San Francisco: Morgan Kaufmann, 2005, 560 p.

YAMAGUCHI, Y.; MORIYAMA, T.; ISHIDO, M.; YAMADA, H. Four-component scattering model for polarimetric SAR image decomposition. **IEEE Transactions on Geoscience Remote Sensing**, v. 43, n. 8, p. 1699–1706, 2005.

**ARTIGO 1**

**TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS ÀS IMAGENS DO SATÉLITE ALOS-2/PALSAR-2 PARA CLASSIFICAÇÃO DE USO E COBERTURA DE TERRAS**

Flávio F. Camargo <sup>a,b,\*</sup>, Edson E. Sano <sup>a</sup>, Cláudia M. Almeida <sup>c</sup>, José C. Mura <sup>c</sup>

<sup>a</sup> Universidade de Brasília (UnB), Instituto de Geociências, Brasília, DF, Brasil

<sup>b</sup> Departamento Nacional de Infraestrutura de Transportes (DNIT), São Paulo, SP, Brasil

<sup>c</sup> Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP, Brasil

\* Autor correspondente:

Departamento Nacional de Infraestrutura de Transportes (DNIT)

Rua Ciro Soares de Almeida, 180 - Vila Maria, CEP: 02167-000, São Paulo, SP, Brasil

Tel.: +55 11 3240-7907

E-mail: flavio.fortes.camargo@gmail.com

(artigo submetido à revista Computers & Geosciences)

**RESUMO**

Este artigo propõe uma metodologia baseada em mineração de dados para a classificação de imagens de radar de abertura sintética (do acrônimo em inglês SAR - *synthetic aperture radar*) obtidas pelo satélite ALOS-2/PALSAR-2, visando ao mapeamento de uso e cobertura de terras. A área de estudo selecionada, com 356 km<sup>2</sup>, localiza-se na porção norte do Distrito Federal. A abordagem metodológica proposta combina segmentação multirresolução, atributos de segmentos e procedimentos interativos e iterativos de aprendizado da máquina. Foram gerados 397 atributos com base nas imagens em amplitude obtidas nas polarizações HH e HV. Esses atributos foram processados com o emprego dos algoritmos de classificação J48 (árvore de

decisão), *Random Forest* (Floresta Aleatória - RF) e Rede Neural Artificial *Multilayer Perceptron* (Perceptron Multicamada - MLP), no *software* WEKA 3.8. Todos os classificadores apresentaram índices de concordância Kappa superiores a 0,70, com destaque para o desempenho da Rede Neural Artificial, com índice Kappa de 0,87. O algoritmo RF apresentou desempenho inferior aos resultados da literatura, o que pode ser atribuído a limitações no tamanho e distribuição espacial da amostragem de treinamento realizada. A metodologia proposta é ágil e replicável, inclusive para imagens obtidas por sensores de radar que operam em comprimentos de onda diferentes daquele do satélite ALOS-2/PALSAR-2.

**Palavras-chave:** Mineração de dados, Aprendizado de máquina, Árvore de decisão, Floresta aleatória, Perceptron multicamadas

### ABSTRACT

This paper proposes a workflow for the classification of synthetic aperture radar (SAR) images obtained by the ALOS-2/PALSAR-2 satellite, aiming at the land use and land cover classification. The selected study area (356 km<sup>2</sup>) is located in the northern portion of Brasília, the Federal District of Brazil. The proposed methodological approach combines multiresolution segmentation, object attributes, and machine learning procedures. A set of 397 attributes was generated based on the amplitude, HH and HV polarization images. These attributes were processed in the WEKA 3.8 software using the J48 decision tree (J48 DT), Random Forest (RF) and Multilayer Perceptron Artificial Neural Network (MLP ANN) classifiers. Classification results attained Kappa indices higher than 0.70, especially the MLP ANN algorithm, with a Kappa index of 0.87. The RF algorithm presented low performance in comparison with the results presented in the literature, probably due to the reduced number and poor spatial distribution of the training samples. The proposed workflow demands low time processing and has the potential to be reproduced not only for ALOS-2/PALSAR-2 images from other study sites but also for further SAR images obtained at different wavelengths.

**Keywords:** Data mining, Machine learning, Decision tree, Random forest, Multilayer perceptron.

### 3.1 - Introdução

O sensoriamento remoto tem-se consolidado como um conjunto eficiente de tecnologias para mapeamento da superfície terrestre. A capacidade de aquisição sistemática de imagens de áreas extensas do terreno, em diferentes resoluções espaciais e em diferentes regiões do espectro eletromagnético, torna o sensoriamento remoto um importante instrumento para identificação, caracterização e quantificação dos diferentes tipos de uso e cobertura de terras (BARGIEL e HERRMANN, 2011; CONGALTON et al., 2014). Por sua vez, dados sobre uso e cobertura de terras são essenciais em diversas atividades, tais como no planejamento urbano e regional (WENTZ et al., 2008; GAMBA e ALDRIGHI, 2012; QI et al., 2012), inventário de recursos naturais (FREITAS et al., 2008; WESTRA et al., 2010; BIRO et al., 2013; EVANS e COSTA, 2013; REYNOLDS et al., 2016) e em modelagens de processos ambientais em escalas global e regional (BRAUN e HOCHSCHILD, 2017), notadamente nos casos que envolvem redução de emissões de gases de efeito estufa relacionadas com desmatamento e degradação florestal (ROSENQVIST et al., 2003; FEDDEMA et al., 2005; MILES e KAPOs, 2008; HAARPAINNER et al., 2015).

Embora pouco utilizados principalmente por causa da complexidade de interpretação e porque a maioria dos seus dados são comercializados, no âmbito do mapeamento de uso e cobertura de terras, destacam-se os dados de radar de abertura sintética (do acrônimo em inglês SAR – *synthetic aperture radar*), obtidos por sensores imageadores que operam na faixa das micro-ondas do espectro eletromagnético (entre 1 e 120 cm). Os sensores de radar possuem capacidade de operação diurna ou noturna e de gerar imagens independentemente da presença de nuvens, e com alguma limitação em presença de chuva, pois utilizam antenas que emitem sua própria radiação eletromagnética em comprimentos de onda relativamente longos (OUCHI, 2013).

Os sensores de radar apresentam ainda outras vantagens em relação aos sensores ópticos, isto é, sensores que operam nas faixas espectrais do visível e do infravermelho. Dentre essas outras vantagens, destaca-se a independência em relação às variações das condições de iluminação solar (elevação e azimute solar) durante o dia ou ano. Imagens de radar apresentam ainda, em relação às imagens ópticas, maior sensibilidade ao conteúdo de água no solo (RAHMAN et al., 2008), à rugosidade da superfície terrestre (DUARTE et al., 2008; TOLLERUD e FANTLE, 2014), e à estrutura da vegetação (BERGEN et al., 2009), indicando, no mínimo, uma importante complementariedade

em relação às imagens ópticas. Países como Alemanha, Canadá, Itália e Japão têm lançado satélites de radar que operam nas bandas X, C ou L (comprimentos de onda em torno de 3 cm, 5 cm e 23 cm, respectivamente) e em diferentes combinações de polarizações (HH, HV, VH e VV; H = polarização horizontal; V = polarização vertical; primeira letra indica polarização incidente e a segunda, polarização refletida do alvo) (OUCHI, 2013).

Concomitantemente ao avanço do sensoriamento remoto, a mineração de dados (MD) surgiu como uma técnica voltada ao tratamento, análise e descoberta de conhecimento em diferentes bases de dados (PIATETSKY-SHAPIRO e FAYYAD, 2012; TSAI, 2012, 2013). A MD corresponde a um campo interdisciplinar que combina inteligência artificial, gerenciamento e visualização de dados, aprendizado de máquina, algoritmos matemáticos e estatística. Ela oferece diferentes metodologias para tomadas de decisão, resoluções de problemas, análises, planejamentos, diagnósticos, detecções de padrões, integrações, prevenções, aprendizados e inovações. As publicações científicas empregando a MD têm crescido em áreas variadas de atuação, incluindo-se as áreas de Geografia e Ciência da Informação Geográfica (TSAI, 2012, 2013).

Dentro das atividades de mapeamento de uso e cobertura de terras com dados de sensoriamento remoto, podem ser destacados diversos estudos que utilizaram, com sucesso, as diferentes técnicas de MD, tais como a árvore de decisão (AD) (SILVA et al., 2008; BISWAL et al., 2013; KÖRTING et al., 2013; SHARMA et al., 2013; PASSO et al., 2014; ANJOS et al., 2017), o classificador *Random Forest* (Floresta Aleatória - RF) (OK et al., 2012; BELGIU e DRAGUT, 2016; ANJOS et al., 2017), as Redes Neurais Artificiais (RNA) (FOODY, 2017) e o *Support Vector Machine* (Máquina de Suporte Vetorial - SVM) (ZHANG et al., 2015).

Com relação ao uso da técnica de MD para classificar imagens de radar, destaca-se o estudo conduzido por Waske e Braun (2009), que avaliaram o potencial dos algoritmos *Boosted Decision Tree* (Árvore de Decisão Impulsionada) e RF para classificar séries temporais de dados SAR obtidos na banda C (comprimento de onda em torno de 5 cm) de diferentes áreas. Qi et al. (2012) propuseram uma nova abordagem de classificação que envolvia o uso de imagens polarimétricas e interferométricas de radar na banda C, processadas pelos algoritmos de segmentação de imagens e AD. Topouzelis e Psyllos (2012) desenvolveram um novo método para detectar manchas de óleo no mar por meio de classificação de imagens SAR pela técnica de RF. Shiraishi et al. (2014) investigaram a precisão e a velocidade de

processamento dos algoritmos *Naive Bayes*, *AdaBoost*, *Multilayer Perceptron* (Perceptron Multicamada - MLP), RF e SVM na classificação de mosaicos de imagens de radar do satélite japonês ALOS/PALSAR. Uhlmann e Kiranyaz (2014) empregaram SVM e RF na classificação e avaliação de atributos de cores extraídos da decomposição polarimétrica de dados SAR obtidos na banda C. Du et al. (2015) estudaram o potencial de atributos texturais e polarimétricos extraídos de imagens SAR obtidas na banda C para mapear diferentes classes de uso e cobertura de terras. Para isso, foram utilizados os algoritmos *Rotation Forest* (Floresta de Rotação), RF e SVM.

Dentro deste contexto, este artigo apresenta uma nova metodologia de classificação de imagens SAR obtidos na banda L pelo satélite japonês ALOS-2/PALSAR-2 para discriminar classes de uso e cobertura de terras. O procedimento envolve o uso combinado de técnicas de segmentação de imagens e classificadores AD, RF e MLP. Esta abordagem proposta é interativa, iterativa e replicável, possibilitando a obtenção de resultados rápidos e com altos índices de acerto.

Este estudo foi organizado conforme se segue. Na Seção 3.2, é apresentada a área de estudo, os três classificadores utilizados, informações sobre o sistema de imageamento do satélite ALOS-2/PALSAR-2 e a descrição da abordagem metodológica empregada. Os resultados e discussões são apresentados na Seção 3.3, enquanto na Seção 3.4 são destacadas as principais conclusões obtidas por este estudo.

## **3.2 - Material e Métodos**

### **3.2.1 - Material**

#### **3.2.1.1 - Área de estudo**

A área de estudo, com aproximadamente 356 km<sup>2</sup> de extensão, localiza-se na porção norte do Distrito Federal do Brasil, mais especificamente, entre 15° 39' 02" e 15° 53' 26" de latitude sul e entre 47° 54' 02" e 48° 01' 41" de longitude oeste (Figura 3.1). Esta área foi selecionada devido à diversidade relativamente grande de tipos de uso e cobertura de terras, com destaque para a presença de vegetação nativa de Cerrado na parte central da área, áreas urbanas consolidadas e em consolidação na parte sul e um reservatório (represa Santa Maria) para captação de água para abastecimento urbano na parte norte.

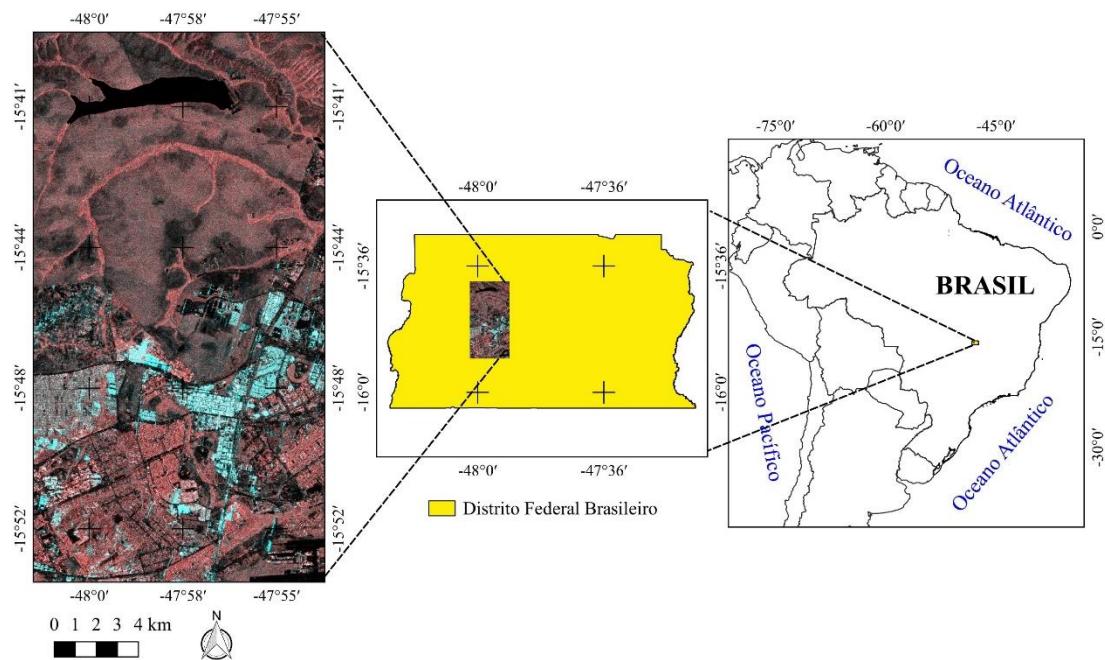


Figura 3.1. Localização da área de estudo no Distrito Federal. A imagem corresponde a uma composição colorida RGB (RGB/HV-HH-HH) da cena do satélite ALOS-2/PALSAR-2, obtida em passagem de 29 de abril de 2015. Datum horizontal: WGS84; Projeção: UTM 23S.

### 3.2.1.2 - O software WEKA e classificadores utilizados

O software WEKA (*Waikato Environment for Knowledge Analysis* – Waikato Ambiente para Análise do Conhecimento) consiste em um ambiente computacional de domínio público e composto de diversos algoritmos de aprendizado de máquina específicos para tarefas de MD. WEKA foi desenvolvido pela Universidade de Waikato, na Nova Zelândia. Os algoritmos disponíveis nesse ambiente foram implementados em linguagem Java, e os códigos são abertos (WITTEN e FRANK, 2005; BOUCKAERT et al., 2010). O WEKA apresenta uma interface simples e intuitiva, facilitando a sua operação. No WEKA *Explorer*, há uma sequência de abas com ferramentas de pré-processamento, seleção de atributos, visualização de dados e classificação, dentre outras funcionalidades.

Três algoritmos de classificação disponíveis nesse *software* foram utilizados neste artigo, a saber: a AD J48, o RF e o MLP. Detalhes sobre o funcionamento destes algoritmos são descritos em Witten e Frank (2005).

### 3.2.1.3 - A missão ALOS-2/PALSAR-2

O satélite japonês denominado de *Advanced Land Observing Satellite-2* (Satélite Avançado de Observação da Terra 2 - ALOS-2), foi lançado em 24 de maio de 2014 e opera com um sensor SAR denominado de *Phase Array L-band Synthetic Aperture Radar-2* (Radar de Abertura Sintética Banda L do tipo Matriz Faseada 2 - PALSAR-2). O ALOS-2 corresponde ao segundo satélite da série ALOS e foi lançado com a missão de garantir consistência e continuidade dos dados SAR na banda L (comprimento de onda de 23,5 cm) (ROSENQVIST et al., 2014). O ALOS-2 utiliza três modos distintos de imageamento denominados de *Spotlight*, *StripMap* e *ScanSAR*, com diferentes resoluções espaciais, faixas de imageamento e polarizações (Quadro 3.1).

Quadro 3.1. Detalhes dos diferentes modos de aquisição do satélite ALOS-2/PALSAR-2.

| Modo de observação   | <i>Spotlight</i> | <i>StripMap</i>  |                       |             | <i>ScanSAR</i> |             |
|----------------------|------------------|------------------|-----------------------|-------------|----------------|-------------|
|                      |                  | <i>Ultrafine</i> | <i>High Sensitive</i> | <i>Fine</i> | Normal         | <i>Wide</i> |
| Resolução espacial   | 3 m              | 3 m              | 6 m                   | 10 m        | 100 m          | 60 m        |
| Faixa de imageamento | 25 km            | 50 km            | 50 km                 | 70 km       | 350 km         | 480 km      |
| Polarização          | HH               | HH               | HH, HV, VH, VV        | HH, HV      | HH, HV         | HH, HV      |

Fonte: Adaptado de Rosenqvist et al. (2014).

As missões espaciais do ALOS são o cerne de um plano global de aquisições de dados de observação da Terra denominado de *Basic Observation Scenario* (Cenário Básico de Observação - BOS), que procura manter em funcionamento uma estratégia de atendimento simultâneo a diferentes comunidades de usuários e áreas de aplicação. Aqui, o objetivo é minimizar conflitos durante as aquisições por meio de um planejamento pré-estabelecido (SUZUKI et al., 2013, JAXA, 2014).

### 3.2.1.4 – Dados e chaves de interpretação utilizados

Foram utilizadas imagens em amplitude (16 bits) do satélite ALOS-2/PALSAR-2 de 29 de abril de 2015, obtidas no modo *StripMap High Sensitive*, produto *Fine Beam Dual* (FBD), ou seja, com 6 metros de resolução espacial e polarização dupla (HH, HV),

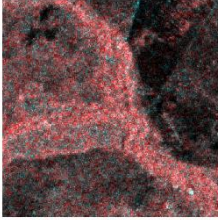
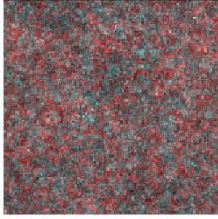
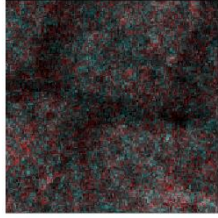
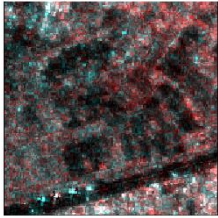
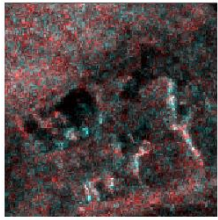


cedidas pela Agência Aeroespacial de Exploração do Japão (*Japan Aerospace Exploration Agency - JAXA*), no âmbito da iniciativa ALOS Kyoto e Carbono. O nível de processamento foi o 3.1, isto é, previamente filtradas para atenuação do ruído *speckle* e georreferenciadas para projeção Universal Transversa de Mercator (UTM), *datum* horizontal WGS84. A imagem selecionada foi obtida com ângulo de incidência, no centro da cena, de 32,9° e órbita descendente.

Para a identificação dos tipos de uso e cobertura de terras presentes na área de estudo, foram utilizados mosaicos de ortofotos aéreas, do ano de 2015, disponíveis *online* nos *sites* da Companhia de Planejamento do Distrito Federal (CODEPLAN, 2016) e da Secretaria de Estado de Gestão do Território e Habitação do Distrito Federal (SEGETH, 2016). Utilizou-se ainda, como referência, um mapa semidetalhado da cobertura vegetal do Parque Nacional de Brasília, elaborado por Ferreira et al. (2007).

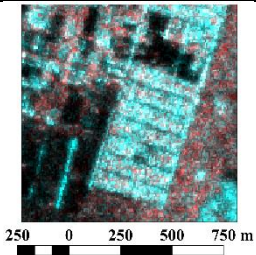
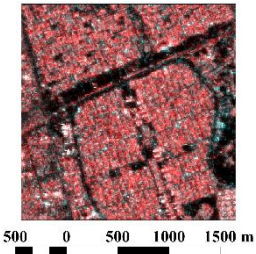
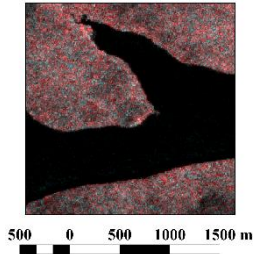
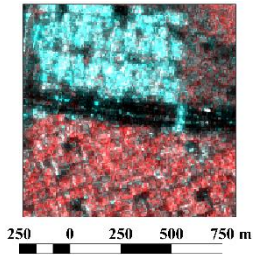
Com base nesse material, foram definidas as seguintes classes temáticas presentes na área de estudo (Quadro 3.2): Mata de Galeria, Cerrado Típico, Cerrado Ralo, Pastagem e Cerrado antropizado, Solo exposto, Área urbana com reflexão de canto, Área urbana sem reflexão de canto, Reservatório e Rodovia pavimentada de grande porte. Cerrado Típico corresponde a uma formação savânica composta por um mosaico de árvores, arbustos e gramíneas de variadas proporções, mas com predomínio de arbustos (RIBEIRO e WALTER, 2008). No Cerrado Ralo, a ocorrência de árvores é bem menor em relação ao Cerrado Típico. Reflexão de canto refere-se a uma dupla reflexão que ocorre com a radiação eletromagnética emitida por um sensor de radar. Nesse caso, a intensidade da radiação refletida é quase igual à emitida, causando saturação de *pixels* nas imagens de radar. Esse fenômeno é bastante comum em áreas urbanas e depende basicamente da orientação das edificações em relação ao ângulo de visada do sensor. Ainda com base nas inspeções visuais desse material, foram definidas as áreas potenciais para coleta de amostras de valores digitais (regiões de interesse) nas imagens do ALOS-2/PALSAR-2 para treinamento e validação dos classificadores empregados.

Quadro 3.2. Chave de interpretação visual de composições coloridas RGB das polarizações HH, HV (RGB/HV-HH-HH). Datum horizontal: WGS84; Projeção: UTM 23S.

| Classe                         | Elementos de interpretação  | Composição colorida RGB   |
|--------------------------------|---|---|
| Mata de Galeria                | Textura rugosa; forma geométrica alongada; coloração avermelhada.                                 |    |
| Cerrado Típico                 | Textura rugosa; sem forma geométrica definida; coloração vermelho-azulada.                        |    |
| Cerrado Ralo                   | Textura rugosa; sem forma geométrica definida; coloração escura com partes em ciano escuro.       |   |
| Pastagem e Cerrado antropizado | Textura lisa a levemente rugosa; forma geométrica predominantemente retangular; coloração escura. |  |
| Solo exposto                   | Textura lisa; sem forma geométrica definida; coloração escura.                                    |  |

(continua)

Quadro 3.2 - Conclusão

| Classe                              | Elementos de interpretação                                       | Composição colorida RGB   |
|-------------------------------------|--|---|
| Área urbana com reflexão de canto   | Textura rugosa; forma geométrica regular; coloração ciano.       |    |
| Área urbana sem reflexão de canto   | Textura rugosa; forma geométrica regular; coloração avermelhada. |    |
| Reservatório                        | Textura lisa; sem forma geométrica definida; coloração preta.    |   |
| Rodovia pavimentada de grande porte | Textura lisa; forma geométrica linear; coloração preta.          |  |

### 3.2.2 - Método

#### 3.2.2.1 - Fluxograma

As principais etapas do trabalho estão descritas no fluxograma mostrado na Figura 3.2, envolvendo desde a etapa de obtenção de imagens até a classificação e validação dos resultados.

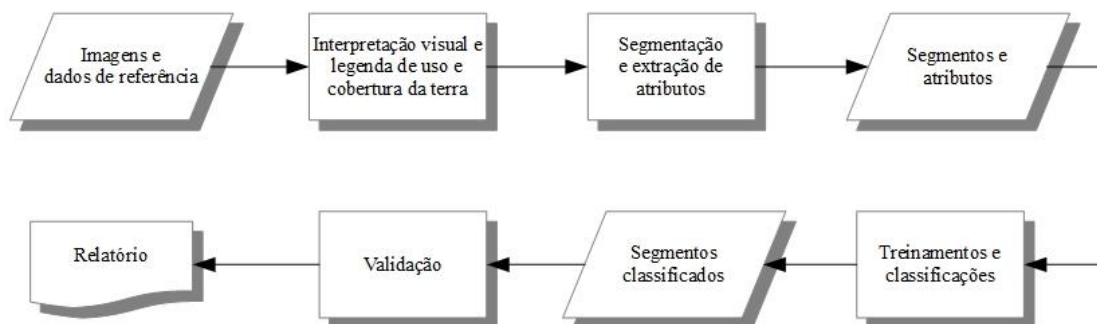


Figura 3.2. Fluxograma das principais etapas do trabalho.

### 3.2.2.2 - Segmentação de imagens

Para a segmentação das imagens do satélite ALOS-2/PALSAR-2 (polarizações HH e HV), foi utilizada a segmentação multiresolução por crescimento de regiões, disponível no *software eCognition 8.7* (TRIMBLE, 2011). Essa segmentação realiza um procedimento de otimização heurística que minimiza a heterogeneidade média dos segmentos de uma imagem em uma dada resolução ou escala (BAATZ e SCHÄPE, 2000; BENZ et al., 2004; MEINEL e NEUBERT 2004). As principais variáveis empregadas foram o fator de escala e a composição de homogeneidade. Esta última variável é dividida em cor e forma. A forma, por sua vez, está subdividida em compacidade e suavidade. A escala define o tamanho dos segmentos de uma imagem em cada nível de resolução, e a composição de homogeneidade testa a igualdade entre segmentos.

Nesta pesquisa, apenas um nível de segmentação foi gerado e os parâmetros do segmentador foram definidos de forma empírica, por meio de tentativa e erro. Assim, selecionou-se o parâmetro de escala 350, e foram atribuídos maiores pesos aos critérios de homogeneidade, forma e suavidade, em detrimento à cor e à compacidade. A Figura 3.3 apresenta um exemplo da segmentação obtida em uma região situada na porção oeste da área de estudo.

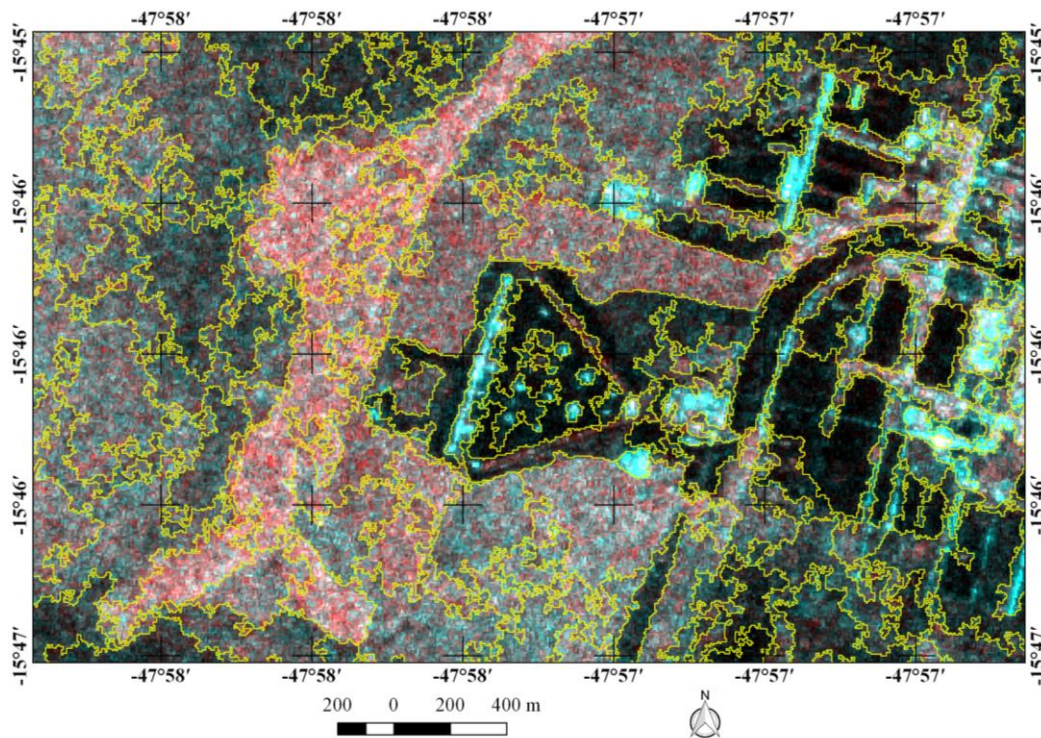


Figura 3.3. Exemplo de um resultado da segmentação de imagens do satélite ALOS-2/PALSAR-2 envolvendo as polarizações HH e HV. Os seguintes parâmetros de segmentação foram considerados: escala = 350; forma = 0,9; cor = 0,1; compactação = 0,1; suavização = 0,9. Os segmentos foram sobrepostos a uma composição colorida RGB (RGB/HV-HH-HH). Datum horizontal: WGS 84; Projeção: UTM 23S.

Após a segmentação, foram extraídos os atributos dos segmentos dentro do próprio *software eCognition 8.7*. Dentre as diversas categorias de métricas de atributos disponíveis, foram utilizadas as seguintes categorias: *layer values* (valores da camada), *textural* (textural), *pixel-based* (baseado em *pixel*), *to-neighbors* (relativo aos vizinhos), *geometric* (geométrico) e *scene* (da cena) (Quadro 3.3). Deste modo, para cada imagem disponível (amplitude HH, amplitude HV, intensidade HH e intensidade HV) foram extraídos atributos das seis categorias acima citadas. Este processo culminou em um conjunto de 397 *layers* de segmentos com diferentes atributos que, posteriormente, foram exportados para tratamento no *software WEKA 3.8*. A Figura 3.4 apresenta alguns exemplos de atributos extraídos. Os atributos texturais são obtidos por meio do método de Haralick atualizado, o qual emprega matrizes de co-ocorrência de níveis de cinza (*Gray Level Co-Occurrence Matrix – GLCM*) e também vetores de diferença de níveis de cinza (*Gray Level Difference Vector – GLDV*) (TRIMBLE, 2011).

Quadro 3.3. Categorias e correspondentes tipos de atributos que foram considerados neste estudo.

| Categoria  | Atributos  |
|--|--|
| <i>Layer values</i><br>(valores da camada)       | <i>Mean</i> (média), <i>standard deviation</i> (desvio padrão), <i>skewness</i> (assimetria), <i>brightness</i> (brilho), <i>maximum difference</i> (máxima diferença)   |
| <i>Textural (Haralick)</i><br>(textural)         | <i>GLCM homogeneity</i> (homogeneidade GLCM), <i>GLCM contrast</i> (contrast GLCM), <i>GLCM dissimilarity</i> (dissimilaridade GLCM), <i>GLCM entropy</i> (entropia GLCM), <i>GLCM angular 2<sup>nd</sup> moment</i> (2 <sup>nd</sup> momento angular GLCM), <i>GLCM mean</i> (média GLCM), <i>GLCM standard deviation</i> (desvio padrão GLCM), <i>GLCM correlation</i> (correlação GLCM), <i>GLDV angular 2<sup>nd</sup> moment</i> (2 <sup>nd</sup> momento angular GLDV), <i>GLDV entropy</i> (entropia GLDV), <i>GLDV mean</i> (média GLDV), <i>GLDV contrast</i> (constraste GLDV)   |
| <i>Pixel-based</i><br>(baseado em <i>pixel</i> ) | <i>Ratio</i> (razão), <i>minimum pixel value</i> (mínimo valor do <i>pixel</i> ), <i>maximum pixel value</i> (máximo valor do <i>pixel</i> ), <i>mean of inner border</i> (média dos valores dos <i>pixels</i> internos à borda do segmento), <i>mean of outer border</i> (média dos valores dos <i>pixels</i> externos à borda do segmento), <i>border contrast</i> (contraste de borda), <i>contrast to neighbor pixels</i> (contraste aos <i>pixels</i> do segmento vizinho), <i>edge contrast of neighbor pixels</i> (contraste aos <i>pixels</i> de borda do segment vizinho), <i>standard deviation to neighbor pixels</i> (desvio padrão aos <i>pixels</i> do segmento vizinho), <i>circular mean</i> (média circular), <i>circular standard deviation</i> (desvio padrão circular), <i>circular standard deviation/mean</i> (razão entre desvio padrão e média circulares) |
| <i>To neighbors</i><br>(relativo aos vizinhos)   | <i>Mean difference to neighbors</i> (diferença média aos vizinhos), <i>mean difference to darker neighbors</i> (diferença média aos vizinhos mais escuros), <i>mean difference to brighter neighbors</i> (diferença média aos vizinhos mais claros), <i>number of brighter objects</i> (número de objetos mais claros), <i>number of darker objects</i> (número de objetos mais escuros), <i>relative border to brighter objects</i> (borda relativa aos objetos mais claros)  |
| <i>Geometric</i><br>(geométrico)                 | <i>Area (pixel)</i> (área do segmento em <i>pixels</i> ), <i>border length (pixel)</i> (comprimento da borda em <i>pixels</i> ), <i>length (pixel)</i> (comprimento em <i>pixels</i> ), <i>width (pixel)</i> (largura em <i>pixels</i> ), <i>length/width</i> (razão largura/comprimento), <i>volume (pixel)</i> (volume em <i>pixels</i> )  |

(continua)

| <i>Scene</i><br>(relativo à cena) | <i>Ratio to scene</i> (razão em relação à cena toda) |
|-----------------------------------|--|
|-----------------------------------|--|

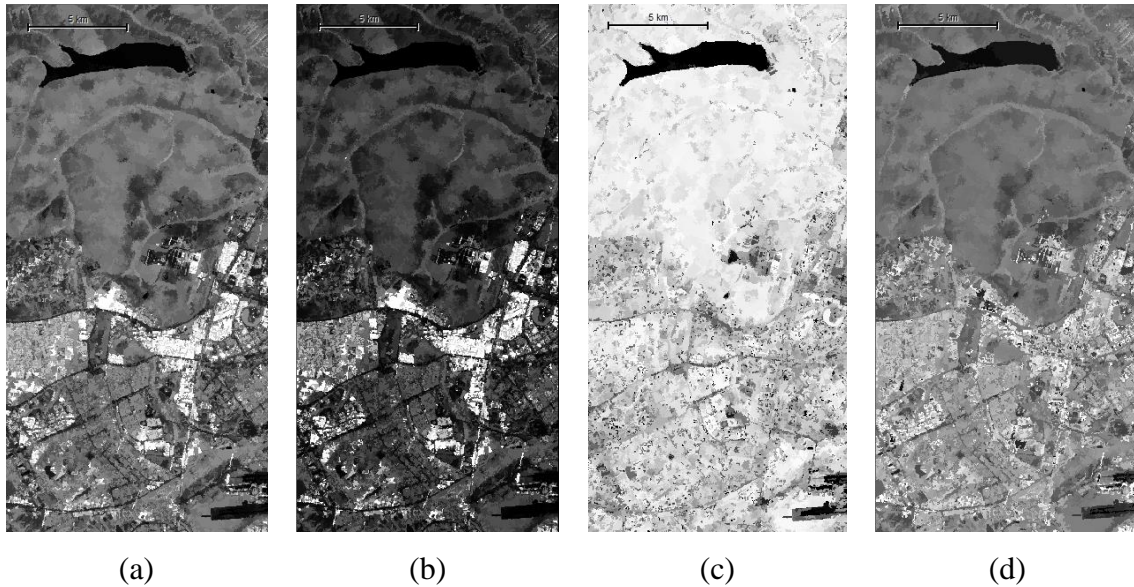


Figura 3.4. Exemplos de atributos extraídos por meio de segmentos gerados por segmentação multiresolução: média por objeto da polarização HH em amplitude (a); média por objeto da polarização HH em intensidade (b); GLCM entropia por objeto (todas as direções) (c); e desvio-padrão por objeto da polarização HH em amplitude (d). *Datum* horizontal: WGS84; Projeção: UTM 23S.

### 3.2.2.3 - Classificação e validação

Conforme mencionado anteriormente, neste estudo, foram selecionados os classificadores AD J48, RF e RNA MLP. Cabe destacar que estes classificadores são não-paramétricos, portanto, flexíveis aos mais variados tipos e distribuições de dados de entrada. Para todos os classificadores, foram realizadas as etapas de treinamento, classificação e validação.

Primeiramente, com base na legenda e nas chaves de interpretação visual, realizou-se a coleta de segmentos amostrais, tanto para o treinamento dos classificadores quanto para a validação das classificações. Foram coletadas ao menos 30 amostras para cada classe, com exceção das classes Reservatório e Pastagem/Cerrado antropizado, pois não se dispunham de muitos segmentos. As amostras foram separadas aleatoriamente no *software* QGIS em dois grupos, um para treinamento (mínimo de 10 segmentos por classe) e outro para validação (mínimo de 20 segmentos por classe) (Figura 3.4). Segundo Congalton (1988) e Congalton e Green (2009), o número mínimo ideal de

amostras de validação por classe seria 50, para áreas de estudo com extensão de até 4.047 km<sup>2</sup> e uma legenda composta por até 12 classes. Todavia, neste trabalho foram utilizados segmentos gerados por segmentação para validação. Esses segmentos possuem tamanhos variáveis e agregam muitos *pixels*. Dado o tamanho da área de estudo (356 km<sup>2</sup>), não foi viável a coleta de 50 segmentos por classe. Desse modo, considerou-se razoável o número de 20 segmentos por classe para validação, considerando uma acurácia temática mínima de 85%, conforme estudo apontado por Genderen e Lock (1977).

Deste modo, foram gerados dois arquivos em formato *shapefile*, um para treinamento dos classificadores e outro para validação das classificações (Figura 3.5). Ainda com relação às classificações, um terceiro arquivo em formato *shapefile* foi gerado com todos os segmentos e contendo todos os campos de atributos gerados anteriormente.

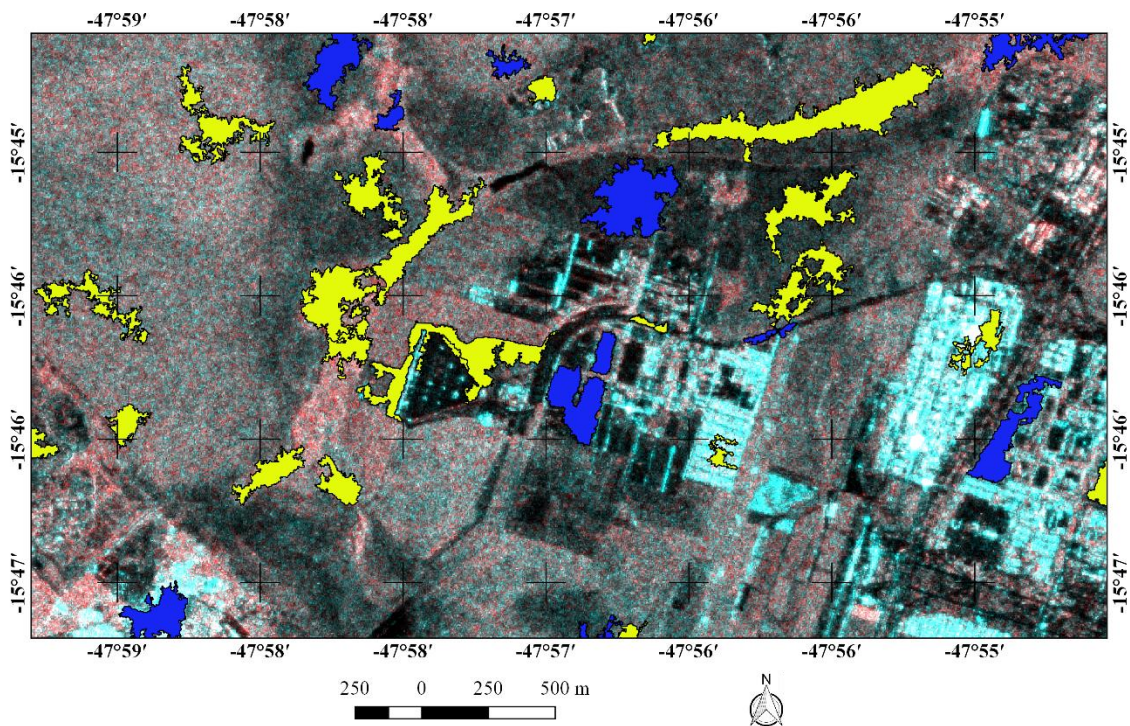


Figura 3.5. Localização das amostras de treinamento (em azul) e de validação (em amarelo) na porção sul da área de estudo. A imagem refere-se à composição colorida RGB das polarizações HH, HV. Datum horizontal: WGS84; Projeção: UTM 23S.

Para a validação dos resultados das classificações, foram construídas matrizes de confusão, comparando os segmentos classificados com os segmentos do conjunto amostral de validação. A acurácia da classificação foi analisada por meio de erros por



classe, acurácia global e índice Kappa. Os diferentes índices Kappa obtidos foram comparados entre si para verificar qual classificador apresentou o melhor desempenho. Para tal propósito, empregaram-se testes de hipóteses com base na distribuição normal padrão (teste Z) (CONGALTON e GREEN, 2009).

### 3.3 - Resultados e Discussão

Na Figura 3.6 são apresentadas, respectivamente, as classificações obtidas com os algoritmos AD J48, RF e RNA MLP. A análise visual da Figura 3.6 permite verificar que a classificação obtida pela RNA MLP apresentou melhor organização espacial dos padrões de uso e cobertura de terras. Verifica-se um delineamento adequado das rodovias de grande porte e também das quadras urbanas. O reservatório ao norte foi classificado corretamente, sem apresentar confusão com a classe Solo exposto. As áreas de Cerrado da classificação com a RNA MLP, por sua vez, apresentam distribuição espacial semelhante àquelas observadas nos mosaicos de ortofotos e também no mapa de semidetalhe.

Na Figura 3.7, é mostrado o resultado do gráfico da AD J48, em que foram gerados nove nós-folhas (*number of leaves*) que correspondem às classes de interesse. Houve ainda a formação de 17 ramos (*size of the tree*) que resultaram das partições da cena por meio do cálculo iterativo da entropia ou ganho de informação. Observa-se que os atributos texturais GLCM e GLDV, particularmente o segundo momento angular, e os atributos relativos ao contraste e brilho entre segmentos e a cena toda (brilho, borda relativa, razão em relação à cena toda e contraste aos *pixels* de borda) foram elencados como os mais hábeis para a separação das classes. Com relação às imagens de entrada, somente as polarizações HH e HV em amplitude apresentaram atributos selecionados. As imagens em intensidade não apresentaram atributos selecionados pelo processo *top-down* da AD J48.

A matriz de confusão da AD J48 (Tabela 3.1) indica que a acurácia global foi de 76%, enquanto o índice Kappa foi de 0,72. O índice Kappa foi considerado significativo, ou seja, diferente de zero (Quadro 3.4). A análise pormenorizada dos erros e acertos contidos na Tabela 3.1 possibilita verificar elevada confusão entre as classes Solo exposto e Rodovia de grande porte, o que já era esperado, haja vista a predominância de espalhamento quase especular nessas áreas devido à baixa rugosidade do terreno em relação ao comprimento de onda relativamente longo da banda L do ALOS-2/PALSAR-

2. A classe Solo exposto obteve um acerto de 50% e apresentou erros de comissão com as classes Rodovia de grande porte e Reservatório. As análises das classes Pastagem/Cerrado antropizado e Reservatório ficaram prejudicadas, dada a pequena quantidade de amostras disponíveis para validação.

De acordo com a literatura pertinente (SILVA et al., 2008; QI et al., 2012; BISWAL et al., 2013; SHARMA et al., 2013), a AD apresentou habilidade para tratar um massivo conjunto de dados de modo não paramétrico (397 atributos), selecionando um conjunto diminuto de atributos que apresentaram o melhor resultado (Figura 3.7). Todavia, conforme se verifica na Figura 3.6a e na Tabela 3.1, houve ainda muita confusão entre algumas classes.

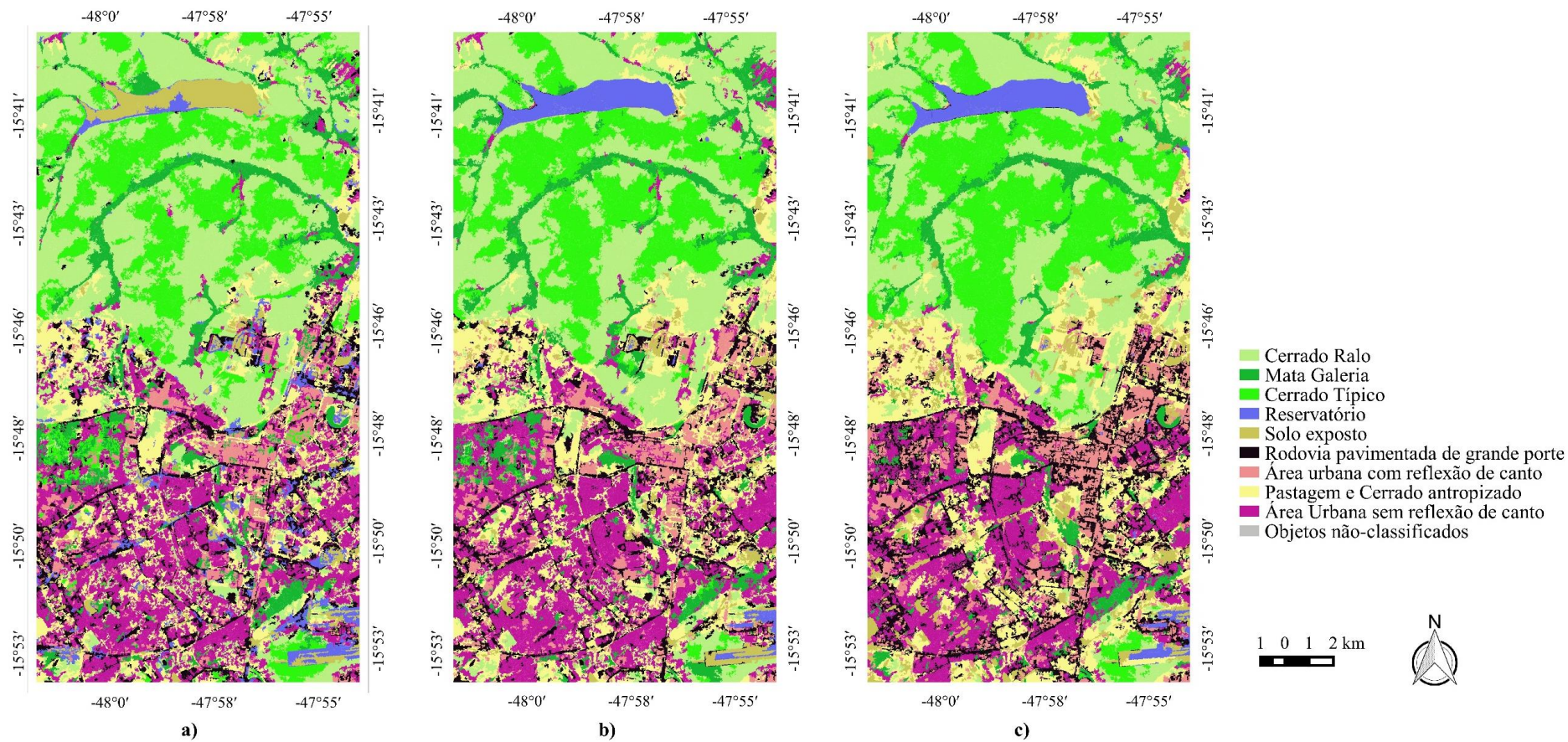


Figura 3.6. Resultados da classificação de imagens pelos algoritmos AD J48 (a), RF (b) e RNA MLP (c).  
*Datum* horizontal: WGS84; Projeção: UTM 23S.

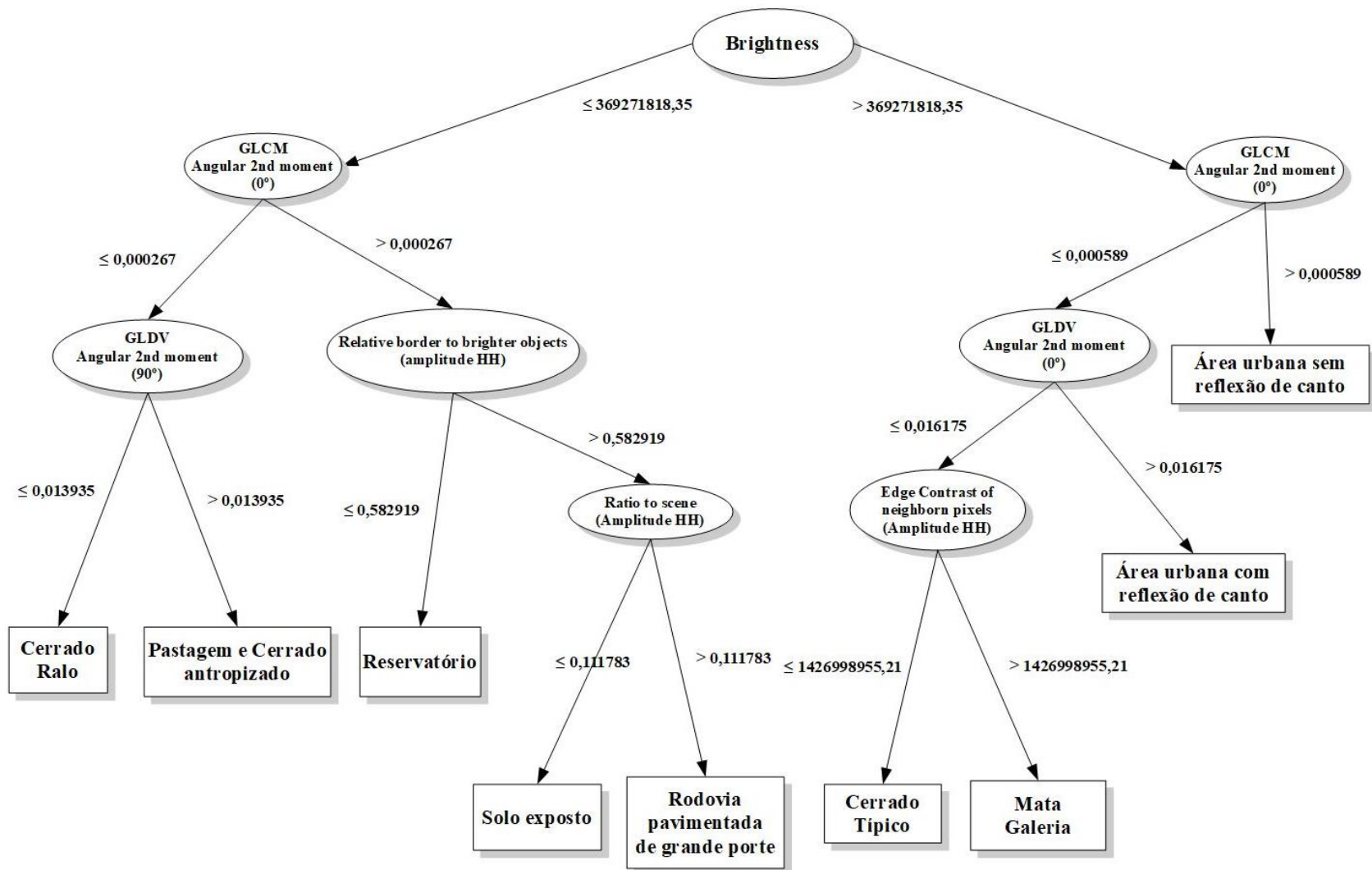


Figura 3.7. Árvore de decisão gerada pelo classificador J48.

Tabela 3.1. Matriz de confusão da classificação de imagens obtida pelo algoritmo AD J48.

|  | Verdade Terrestre |           |          |           |           |           |          |           |          |    | Total      |          |               |
|--|-------------------|-----------|----------|-----------|-----------|-----------|----------|-----------|----------|----|------------|----------|---------------|
|  | a                 | b         | c        | d         | e         | f         | g        | h         | i        | j  |            |          |               |
| <b>20</b>  | 0                 | 0         | 0        | 0         | 0         | 0         | 0        | 0         | 0        | 0  | 20         | <b>a</b> | Classificação |
| 0  | <b>22</b>         | 1         | 0        | 0         | 0         | 0         | 0        | 9         | 0        | 32 | <b>b</b>   |          |               |
| 0  | 2                 | <b>20</b> | 0        | 0         | 0         | 0         | 0        | 0         | 0        | 22 | <b>c</b>   |          |               |
| 0  | 0                 | 0         | <b>3</b> | 3         | 0         | 0         | 0        | 0         | 0        | 6  | <b>d</b>   |          |               |
| 0  | 0                 | 0         | 1        | <b>12</b> | 6         | 0         | 1        | 0         | 0        | 20 | <b>e</b>   |          |               |
| 0  | 0                 | 0         | 0        | 9         | <b>13</b> | 0         | 2        | 0         | 0        | 24 | <b>f</b>   |          |               |
| 0  | 0                 | 1         | 0        | 0         | 0         | <b>22</b> | 0        | 0         | 0        | 23 | <b>g</b>   |          |               |
| 3  | 0                 | 0         | 0        | 0         | 0         | 0         | <b>5</b> | 0         | 0        | 8  | <b>h</b>   |          |               |
| 1  | 2                 | 0         | 0        | 0         | 2         | 1         | 0        | <b>21</b> | 0        | 27 | <b>i</b>   |          |               |
| 0  | 0                 | 0         | 0        | 0         | 0         | 0         | 0        | 0         | <b>0</b> | 0  | <b>j</b>   |          |               |
| <b>Total</b>   | 24                | 26        | 22       | 4         | 24        | 21        | 23       | 8         | 30       | 0  | <b>182</b> |          |               |
| a = Cerrado Ralo; b = Mata de Galeria; c = Cerrado Típico; d = Reservatórios; e = Solo exposto; f = Rodovias de grande porte; g = AUCRC; h = Pastagem/Cerrado antropizado; i = AUSRC; j = Não-classificado |                   |           |          |           |           |           |          |           |          |    |            |          |               |
| Acurácia global: 76% Índice Kappa: 0,72  |                   |           |          |           |           |           |          |           |          |    |            |          |               |

Quadro 3.4. Teste de hipótese da classificação obtida pelo algoritmo AD J48.

|   |
|---|
| <p>H<sub>0</sub>: Kappa J48 = 0</p> <p>H<sub>1</sub>: Kappa J48 &gt; 0</p> <p>Z = 19,95; P-valor = 0,0000; <math>\alpha</math> = 0,05</p> <p>Conclusão: Rejeita H<sub>0</sub> (Kappa J48 é significativamente maior que zero)</p> |
|---|

O algoritmo RF apresentou índices de concordância maiores que os do AD J48: a acurácia global foi de 79% e o índice Kappa de 0,76 (Tabela 3.2). O índice Kappa foi considerado significativo, ou seja, diferente de zero (Quadro 3.5). Todavia, os resultados apresentados pelos classificadores AD J48 e RF não são estatisticamente diferentes, em um nível de significância de 5% (Quadro 3.6). Isto significa que, nesta pesquisa, esses classificadores apresentaram desempenhos iguais. Este resultado diverge de alguns estudos que indicaram que o RF apresenta desempenho superior ao da AD J48 (GHIMIRE et al., 2012; HAN et al., 2015; BELGIU e DRAGUT, 2016).

Pode-se atribuir o baixo desempenho do algoritmo RF ao tamanho e distribuição da amostragem utilizada. Como já salientado anteriormente, o RF é um classificador do tipo *bagging*, ou seja, as árvores de decisões são criadas a partir de subconjuntos do mesmo conjunto amostral com reposição. Deste modo, o RF é sensível às características do conjunto amostral utilizado (BELGIU e DRAGUT, 2016). A análise pormenorizada dos erros e acertos mostra que o classificador RF também apresentou elevada confusão entre as classes Solo exposto e Rodovias de grande porte.

Tabela 3.2. Matriz de confusão da classificação de imagens obtida pelo algoritmo RF.

|  | Verdade Terrestre |           |          |           |           |           |          |           |          |    | Total      |          |               |
|--|-------------------|-----------|----------|-----------|-----------|-----------|----------|-----------|----------|----|------------|----------|---------------|
|  | a                 | b         | c        | d         | e         | f         | g        | h         | i        | j  |            |          |               |
| <b>20</b>  | 0                 | 0         | 0        | 0         | 0         | 0         | 0        | 0         | 0        | 0  | 20         | <b>a</b> | Classificação |
| 1  | <b>21</b>         | 2         | 0        | 0         | 0         | 0         | 0        | 8         | 0        | 32 | <b>b</b>   |          |               |
| 0  | 1                 | <b>21</b> | 0        | 0         | 0         | 0         | 0        | 0         | 0        | 22 | <b>c</b>   |          |               |
| 0  | 0                 | 0         | <b>5</b> | 1         | 0         | 0         | 0        | 0         | 0        | 6  | <b>d</b>   |          |               |
| 0  | 0                 | 0         | 0        | <b>15</b> | 5         | 0         | 0        | 0         | 0        | 20 | <b>e</b>   |          |               |
| 0  | 0                 | 1         | 1        | 7         | <b>15</b> | 0         | 1        | 0         | 0        | 25 | <b>f</b>   |          |               |
| 0  | 0                 | 0         | 0        | 0         | 0         | <b>22</b> | 0        | 0         | 0        | 22 | <b>g</b>   |          |               |
| 3  | 0                 | 0         | 0        | 0         | 0         | 0         | <b>5</b> | 0         | 0        | 8  | <b>h</b>   |          |               |
| 1  | 4                 | 0         | 0        | 0         | 1         | 1         | 0        | <b>20</b> | 0        | 27 | <b>i</b>   |          |               |
| 0  | 0                 | 0         | 0        | 0         | 0         | 0         | 0        | 0         | <b>0</b> | 0  | <b>j</b>   |          |               |
| <b>Total</b>   | 25                | 26        | 24       | 6         | 23        | 21        | 23       | 6         | 28       | 0  | <b>182</b> |          |               |
| a = Cerrado Ralo; b = Mata de Galeria; c = Cerrado Típico; d = Reservatórios; e = Solo exposto; f = Rodovias de grande porte; g = AUCRC; h = Pastagem/Cerrado antropizado; i = AUSRC; j = Não-classificado |                   |           |          |           |           |           |          |           |          |    |            |          |               |
| Acerto Global: 79% Índice Kappa: 0,76  |                   |           |          |           |           |           |          |           |          |    |            |          |               |

Quadro 3.5. Teste de hipótese da classificação obtida pelo algoritmo RF.

|  |
|--|
| H0: Kappa RF = 0   |
| H1: Kappa RF > 0   |
| Z = 22,05; P-valor = 0,0000; $\alpha = 0,05$                                     |
| Conclusão: Rejeita H <sub>0</sub> (Kappa RF é significativamente maior que zero) |

Quadro 3.6. Teste de hipótese comparando o desempenho dos classificadores AD J48 e RF.

|  |
|--|
| H <sub>0</sub> : Kappa J48 – Kappa RF = 0  |
| H <sub>1</sub> : Kappa J48 – Kappa RF < 0  |
| Z = -0,76; P-valor = 0,2245; $\alpha$ = 0,05   |
| Conclusão: Aceita H <sub>0</sub> (Kappa J48 e RF não são significantemente diferentes) |

A RNA MLP apresentou valores relativamente elevados de acurácia global (89%), índice Kappa (0,87) e baixa confusão entre classes (Tabela 3.3). O índice Kappa foi considerado significativo, ou seja, diferente de zero (Quadro 3.7). Houve baixa confusão entre as classes, e todas obtiveram acertos iguais ou superiores a 80%. Mesmo classes de difícil separação, em termos de atributos, apresentaram baixa confusão (Reservatório, Solo exposto e Rodovias de grande porte, todos com tendência de apresentarem baixos índices de retroespalhamento do sinal emitido pelo radar). Observa-se que esse classificador conseguiu explorar melhor o conjunto de atributos disponíveis, comparativamente à AD J48 e ao RF (Quadro 3.8). Por se tratar de uma rede neural artificial, também é possível afirmar que o treinamento foi feito de modo adequado, haja vista o elevado nível de acerto obtido (WITTEN e FRANK, 2005; KASPERCZUK e DARDZIŃSKA, 2016).

O teste de hipótese apresentado no Quadro 3.8 demonstra ainda que os resultados apresentados pelos classificadores RF e RNA MLP são estatisticamente diferentes, em um nível de significância de 5%. Apesar do baixo desempenho do RF, possivelmente devido a problemas na amostragem de treinamento, este resultado é consoante com outras pesquisas que também destacaram o alto desempenho da RNA MLP relativamente a outros classificadores, incluindo o RF (SHIRAISHI et al., 2014).

Quadro 3.7. Teste de hipótese da classificação de imagens obtida pelo classificador RNA MLP.

|   |
|---|
| H <sub>0</sub> : Kappa MLP = 0  |
| H <sub>1</sub> : Kappa MLP > 0  |
| Z = 32,96; P-valor = 0,000; $\alpha$ = 0,05   |
| Conclusão: Rejeita-se H <sub>0</sub> (Kappa MLP é significantemente maior que zero) |

Quadro 3.8. Teste de hipótese comparando o desempenho dos classificadores RF e RNA MLP.

|  |
|--|
| $H_0$ : Kappa RF – Kappa MLP = 0   |
| $H_1$ : Kappa RF – Kappa MLP < 0   |
| $Z = -2,60$ ; P-valor = 0,0047; $\alpha = 0,05$                                |
| Conclusão: rejeita $H_0$ (Kappa RF é significativamente menor que o Kappa MLP) |

Tabela 3.3. Matriz de confusão obtida pelo classificador RNA MLP.

|  | Verdade Terrestre |    |    |    |    |    |    |    |    |   | Total      |   |
|--|-------------------|----|----|----|----|----|----|----|----|---|------------|---|
|  | a                 | b  | c  | d  | e  | f  | g  | h  | i  | j |            |   |
| 20   | 0                 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0 | 20         | a |
| 1  | 28                | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 3  | 0 | 32         | b |
| 0  | 0                 | 22 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0 | 22         | c |
| 0  | 0                 | 0  | 6  | 0  | 0  | 0  | 0  | 0  | 0  | 0 | 6          | d |
| 0  | 0                 | 0  | 0  | 18 | 1  | 0  | 1  | 0  | 0  | 0 | 20         | e |
| 0  | 0                 | 0  | 1  | 3  | 20 | 0  | 0  | 0  | 0  | 0 | 24         | f |
| 0  | 0                 | 0  | 0  | 0  | 1  | 22 | 0  | 0  | 0  | 0 | 23         | g |
| 2  | 0                 | 1  | 0  | 1  | 0  | 0  | 4  | 0  | 0  | 0 | 8          | h |
| 0  | 2                 | 0  | 0  | 0  | 2  | 1  | 0  | 22 | 0  | 0 | 27         | i |
| 0  | 0                 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0 | 0          | j |
| <b>Total</b>   | 23                | 30 | 23 | 7  | 22 | 24 | 23 | 5  | 25 | 0 | <b>182</b> |   |
| a = Cerrado Ralo; b = Mata de Galeria; c = Cerrado Típico; d = Reservatórios; e = Solo exposto; f = Rodovias de grande porte; g = AUCRC; h = Pastagem/Cerrado antropizado; i = AUSRC; j = Não-classificado |                   |    |    |    |    |    |    |    |    |   |            |   |
| Acurácia global: 89% Índice Kappa: 0,87  |                   |    |    |    |    |    |    |    |    |   |            |   |

### 3.4 - Conclusões

A utilização de objetos de segmentação torna bastante flexível a tarefa de classificação de imagens. Além da possibilidade de utilizar estatísticas de segmentos (grupos de *pixels*), forma, contexto e características geométricas, destaca-se a possibilidade de geração de arquivos ASCII contendo poucos registros relativos aos segmentos e seus atributos. Isto diminui sobremaneira a dimensionalidade dos dados de entrada, facilitando os processamentos posteriores e possibilitando o emprego de ferramentas de mineração de dados que não são exclusivas da comunidade de



sensoriamento remoto e disponíveis em *softwares* de domínio público, como é o caso do WEKA 3.8, utilizado neste estudo.

As classificações realizadas demandaram tempo relativamente curto de processamento (< 1 minuto), isto para um arquivo de treinamento com 79 registros, um arquivo a ser classificado com 4.868 registros e um arquivo independente de validação do modelo com 182 registros. Essa velocidade propicia agilidade na avaliação e escolha de classificadores, haja vista que há muitos classificadores disponíveis no *software* WEKA 3.8. Destacam-se ainda os relatórios amigáveis e rápidos que são gerados ao final das classificações.

A segmentação no *eCognition* 8.7 também apresenta muitas vantagens, tais como: segmentador acurado e rápido; diversas ferramentas de extração de atributos dos segmentos; e exportação de segmentos e atributos em diversos formatos (*shapefile*, ASCII etc.). Todavia, como o *eCognition* 8.7 é um *software* proprietário, outras alternativas devem ser avaliadas para compor este procedimento de classificação junto com o WEKA 3.8. Há opções gratuitas de segmentadores como, por exemplo, SPRING (INPE, 2018) e InterIMAGE (LVC, 2018).

Dos resultados obtidos, verifica-se que todos os classificadores apresentaram desempenhos elevados. Isto pode ser atribuído à habilidade não-paramétrica destes classificadores para processarem diversos tipos e distribuições de dados. Destaque deve ser dado ao desempenho do classificador RF, que ficou aquém dos desempenhos relatados na literatura.

Os procedimentos de classificação adotados mostraram ser eficazes na extração de informações das imagens SAR na banda L do satélite ALOS-2/PALSAR-2. Isto é um fato relevante, haja vista que, devido a condições meteorológicas, em alguns lugares da superfície terrestre, somente imagens SAR podem estar disponíveis para estudos e levantamentos. Acrescentam-se ainda as características das imagens SAR que frequentemente não são passíveis de classificação por métodos paramétricos, tornando ainda mais importante o uso de classificadores alternativos tais como os aqui estudados.

Por fim, destacam-se também que a interatividade e a iteratividade intrínsecas dos procedimentos de aprendizado de máquina propiciam maior objetividade na tarefa de classificação e, por conseguinte, tornam o processo passível de replicação em diferentes momentos e por diferentes indivíduos em tarefas de mapeamento do uso e cobertura de terras.

## Agradecimentos

Os autores agradecem à Agência Espacial Japonesa (JAXA), mais especificamente, ao *Kyoto & Carbon Protocol*, pela cessão das imagens ALOS-2/PALSAR-2.

## Referências

ANJOS, C. S.; ALMEIDA, C. M.; GALVÃO, L. S.; FILHO, C. R. S.; LACERDA, M. G.; PRATI, R. C. Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão C4.5 e Floresta randômica. **Boletim de Ciências Geodésicas**, v. 23, n. 2, p. 371- 388, 2017.

BAATZ, M.; SCHÄPE, A. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: ANGEWANDTE GEOGRAPHISCHE INFORMATIONSVERARBEITUNG, 12. Beiträge zum AGIT-Symposium, Salzburg, 2000. **Anais...** Karlsruhe: Herbert Wichmann Verlag, p. 12–23.

BARGIEL, D.; HERRMANN, S. Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight TerraSAR-X data. **Remote Sensing**, v. 3, n. 5, p. 859–877, 2011.

BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BENZ, U. C.; HOFMANN, P.; WILLHAUCK, G.; LINGENFELDER, I.; HEYNEN, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS ready information. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 58, n. 3–4, p. 239–258, 2004.

BERGEN, K. M.; GOETZ, S. J.; DUBAYAH, R. O.; HENEBRY, G. M.; HUNSAKER, C. T.; IMHOFF, M. L.; NELSON, R. F.; PARKER, G. G.; RADELOFF, V. C. Remote sensing of vegetation 3-D structure for biodiversity and habitat: review and implications for lidar and radar spaceborne missions. **Journal of Geophysical Research**, v. 114, G00E06, p. 1–13, 2009.

BIRO, K.; PRADHAN, B.; SULIEMAN, H.; BUCHROITHNER, M. Exploitation of TerraSAR-X data for land use/land cover analysis using object-oriented classification approach in the African Sahel area, Sudan. **Journal of the Indian Society of Remote Sensing**, v. 41, n. 3, p. 539–553, 2013.

BISWAL, S.; GHOSH, A.; SHARMA, R.; JOSHI, P. K. Satellite data classification using open source support. **Journal of the Indian Society of Remote Sensing**, v. 41, n. 3, p. 523–530, 2013.

BOUCKAERT, R. R.; FRANK, E.; HALL, M. A.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. WEKA - experiences with a Java open-source project. **Journal of Machine Learning Research**, v. 11, p. 2533–2541, 2010.

BRAUN, A.; HOCHSCHILD, V. A SAR-based index for landscape changes in African savannas. **Remote Sensing**, v. 9, n. 4, Paper n. 359, 23 p., 2017.

CODEPLAN. Companhia de Planejamento do Distrito Federal. **Geoserviço**. Disponível em: <<http://www.codeplan.df.gov.br/component/content/article/331-geo-servico/306-geo-servico.html>>. Acesso em: 17 set. 2016.

CONGALTON, R. G. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, v. 54, n. 5, p. 593–600, 1988.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Boca Raton: CRC Press, 2009, 200 p.

CONGALTON, R. G.; GU, J.; YADAV, K.; THENKABAIL, P.; OZDOGAN, M. Global land cover mapping: a review and uncertainty analysis. **Remote Sensing**, v. 6, n. 12, p. 12070–12093, 2014.

DU, P.; SAMAT, A.; WASKE, B.; LIU, S.; LI, Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 105, p. 38–53, 2015.

DUARTE, R. M.; WOZNIAK, E.; RECONDO, C.; CABO, C.; MARQUÍNEZ, J.; FERNÁNDEZ, S. Estimation of surface roughness and stone cover in burnt soils using SAR images. **Catena**, v. 74, p. 264–272, 2008.

EVANS, T. L.; COSTA, M. Landcover classification of the lower Nhecolândia subregion of the Brazilian Pantanal wetlands using ALOS/PALSAR, RADARSAT-2 and ENVISAT/ASAR imagery. **Remote Sensing of Environment**, v. 128, p. 118–137, 2013.

FEDDEMA, J. J.; OLESON, K. W.; BONAN, G. B.; MEARNNS, L. O.; BUJA, L. E.; MEEHL, G. A.; WASHINGTON, W. M. The importance of land-cover change in simulating future climates, **Science**, v. 310, n. 5754, p. 1674–1678, 2005.

FERREIRA, M. E.; FERREIRA, L. G.; SANO, E. E.; SHIMABUKURO, Y. E. Spectral linear mixture modelling approaches for land cover mapping of tropical savanna areas in Brazil. **International Journal of Remote Sensing**, v. 28, n. 2, p. 413–429, 2007.

FOODY, G. M. Impacts of sample design for validation data on the accuracy of feedforward neural network classification. **Applied Sciences**, v. 7, n. 888, 15 p., 2017.

FREITAS, C. C.; SOLER, L. S.; SANT'ANNA, S. J. S.; DUTRA, L. V.; SANTOS, J. R.; MURA, J. C.; CORREIA, A. H. Land use and land cover mapping in the Brazilian

Amazon using polarimetric airborne P-band SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, n. 10, p. 2956–2970, 2008.

GAMBA, P.; ALDRIGHI, M. SAR data classification of urban areas by means of segmentation techniques and ancillary optical data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 5, n. 4, p. 1140–1148, 2012.

GENDEREN, J. L.; LOCK, B. F. Testing land-use map accuracy. **Photogrammetric Engineering & Remote Sensing**, v. 43, n. 9, p. 1135–1137, 1977.

GHIMIRE, B.; ROGAN, J.; GALIANO, V. R.; PANDAY, P.; NEETI, N. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. **GIScience & Remote Sensing**, v. 49, p. 623–643, 2012.

HAARPAINTNER, J.; BLANCO, D. F.; ENBLE, F.; DATTA, P.; MAZINGA, A.; SINGA, C.; MANE, L. Tropical forest remote sensing services for the Democratic Republic of Congo inside the EU FP7 ‘Recover’ Project (Final Results 2000-2012). In: THE INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES, v. XL-7/W3, 36<sup>a</sup> International Symposium on Remote Sensing of Environment. Berlim. **Anais...** p. 397–402, 2015.

HAN, H.; LEE, S.; IM, J.; KIM, M.; LEE, M. I.; AHN, M.; CHUNG, S. R. Detection of convective initiation using meteorological imager onboard communication, ocean, and meteorological satellite based on machine learning approaches. **Remote Sensing**, v. 7, n. 7, p. 9184–9204, 2015.

INPE. Instituto Nacional de Pesquisas Espaciais. **Sistema de Processamento de Informações Georreferenciadas (SPRING)**. São José dos Campos: INPE, 2018.

JAXA. Japan Aerospace Exploration Agency. ALOS-2 basic observation scenario (1<sup>a</sup> ed.). 2014. Disponível em: <[www.eorc.jaxa.jp /ALOS-2/en/obs/scenario/ALOS-2\\_Basic\\_Observation\\_Scenario\\_First-Ed\\_E\\_v00.pdf](http://www.eorc.jaxa.jp/ALOS-2/en/obs/scenario/ALOS-2_Basic_Observation_Scenario_First-Ed_E_v00.pdf)>. Acesso em: 10 ago. 2015.

KASPERCZUK, A.; DARDZIŃSKA, A. Comparative evaluation of the different data mining techniques used for the medical database. **Acta Mechanica et Automatica**, v. 10, n. 3, p. 233–238, 2016.

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA – geographic data mining analyst. **Computers & Geosciences**, v. 57, p. 133–145, 2013.

LVC. Laboratório de Visão Computacional da Pontifícia Universidade Católica do Rio de Janeiro. **InterImage: interpreting images freely**. Rio de Janeiro: LVC, 2018. Disponível em: <<http://www.lvc.ele.puc-rio.br/projects/interimage/pt-br/download/>>. Acesso em: 15 dez. 2018.

MEINEL, G.; NEUBERT, M. A. A comparison of segmentation programs for high resolution remote sensing data. **International Archives of Photogrammetry and Remote Sensing**, v. 23, Part B, p. 1097–1105, 2004.

MILES, L.; KAPOS, V. Reducing greenhouse gas emissions from deforestation and forest degradation: global land-use implications. **Science**, v. 320, n. 5882, p. 1454-1455, 2008.

OK, A. O.; AKAR, O.; GUNGOR, O. Evaluation of random forest method for agricultural crop classification. **European Journal of Remote Sensing**, v. 45, p. 421–432, 2012.

OUCHI, K. Recent trend and advance of synthetic aperture radar with selected topics. **Remote Sensing**, v. 5, n. 2, p. 716–807, 2013.

PASSO, D.; BIAS, E.; BRITES, R.; COSTA, G. Comparison of the accuracy of classifications generated by interimage and by INTERIMAGE integrated with data mining. **South-Eastern European Journal of Earth Observation and Geomatics**, v. 3, n. 2S, p. 93–98, 2014.

PIATETSKY-SHAPIRO, G.; FAYYAD, U. An introduction to SIGKDD and a reflection on the term ‘data mining’. **SIGKDD Explorations**, v. 13, n. 2, p. 103–104, 2012.

QI, Z.; YEH, A. G-O.; LI, X.; LIN, Z. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. **Remote Sensing of Environment**, v. 118, p. 21–39, 2012.

RAHMAN, M. M.; MORAN, M. S.; THOMA, D. P.; BRYANT, R.; COLLINS, C. D. H.; JACKSON, T.; ORR, B. J.; TISCHLER, M. Mapping surface roughness and soil moisture using multi-angle radar imagery without ancillary data. **Remote Sensing of Environment**, v. 112, p. 391–402, 2008.

REYNOLDS, J.; WESSON, K.; DESBIEZ, A. L. J.; OCHOA-QUINTERO, J. M.; LEIMGRUBER, P. Using remote sensing and random forest to assess the conservation status of critical Cerrado habitats in Mato Grosso do Sul, Brazil. **Land**, v. 5, Paper 5020012, 12 p., 2016.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. (Eds.). **Cerrado: Ecologia e Flora**, Planaltina: Embrapa Cerrados, 2008. p. 151-199.

ROSENQVIST, A.; MILNE, A.; LUCAS, R.; IMHOFF, M.; DOBSON, C. A review of remote sensing technology in support of the Kyoto Protocol. **Environmental Science & Policy**, v. 6, p. 441–455, 2003.

ROSENQVIST, A.; SHIMADA, M.; SUZUKI, S.; OHGUSHI, F.; TADONO, T.; WATANABE, M.; TSUZUKU, K.; WATANABE, T.; KAMIJO, S.; AOKI, E. Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. **Remote Sensing of Environment**, v. 155, p. 3–12, 2014.

SEGETH. Secretaria de Estado de Gestão do Território e Habitação do Distrito Federal. **Geoportal SEGETH DF**. Disponível em: <<http://geoportal.segeth.df.gov.br/mapa/>>. Acesso em: 17 set. 2016.

SHARMA, R.; GHOSH, A.; JOSHI, P. K. Decision tree approach for classification of remotely sensed satellite data using open source support. **Journal of Earth System Science**, v. 122, n. 5, p. 1237–1247, 2013.

SHIRAISHI, T.; MOTOHKA, T.; THAPA, R. B.; WATANABE, M.; SHIMADA, M. Comparative assessment of supervised classifiers for land use–land cover classification in a tropical region using time-series PALSAR mosaic data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 4, p. 1186–1199, 2014.

SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; SOUZA, R. C. M. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, v. 29, n. 16, p. 4803–4822, 2008.

SUZUKI, S.; KANKAKU, Y.; SHIMADA, M. ALOS-2 acquisition strategy. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS 2013), Melbourne, Australia. **Anais...** p. 2412-2415, 2013.

TOLLERUD, H. J.; FANTLE, M. S. The temporal variability of centimeter-scale surface roughness in a playa dust source: synthetic aperture radar investigation of playa surface dynamics. **Remote Sensing of Environment**, v. 154, p. 285–297, 2014.

TOPOUZELIS, K.; PSYLLOS, A. Oil spill feature selection and classification using decision tree forest on SAR image data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 68, p. 135–143, 2012.

TRIMBLE. **eCognition Developer 8.7: Reference Book**. Munique: Trimble, 2011.

TSAI, H. Global data mining: an empirical study of current trends, future forecasts and technology diffusions. **Expert Systems with Applications**, v. 39, p. 8172–8181, 2012.

TSAI, H. Knowledge management vs. data mining: Research trend, forecast and citation approach. **Expert Systems with Applications**, v. 40, p. 3160–3173, 2013.

UHLMANN, S.; KIRANYAZ, S. Integrating color features in polarimetric SAR image classification. **IEEE Transactions on Geosciences and Remote Sensing**, v. 52, p. 2197–2216, 2014.

WASKE, B.; BRAUN, M. Classifier ensembles for land cover mapping using multitemporal SAR imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, p. 450–457, 2009.

WENTZ, E. A.; NELSON, D.; RAHMAN, A.; STEFANOV, W. L.; ROY, S. S. Expert system classification of urban land use/cover for Delhi, India. **International Journal of Remote Sensing**, v. 29, n. 15–16, p. 4405–4427, 2008.

WESTRA, T.; WULF, R.; COILLIE, F.; CRABBE, S. Optimal ENVISAT advanced synthetic aperture radar image parameters for mapping and monitoring Sahelian floodplains. **Journal of Applied Remote Sensing**, v. 4, n. 1, Paper 043511, 17 p., 2010.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning. Tools and techniques**. (2<sup>a</sup> ed.). San Francisco: Morgan Kaufmann, 2005, 560 p.

ZHANG, H.; LIN, H.; LI, Y. Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 5, p. 1061–1065, 2015.

**ARTIGO 2**

**AVALIAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS E DE ATRIBUTOS SAR POLARIMÉTRICOS ALOS-2/PALSAR-2 NO MAPEAMENTO DO USO E COBERTURA DE TERRAS EM ÁREA DE CERRADO**

Flávio F. Camargo <sup>a,b,\*</sup>, Edson E. Sano <sup>a</sup>

<sup>a</sup> Universidade de Brasília (UnB), Instituto de Geociências, Brasília, DF, Brasil

<sup>b</sup> Departamento Nacional de Infraestrutura de Transportes (DNIT), São Paulo, SP, Brasil

\* Autor correspondente:

Departamento Nacional de Infraestrutura de Transportes (DNIT)

Rua Ciro Soares de Almeida, 180 - Vila Maria, CEP: 02167-000, São Paulo, SP, Brasil

Tel.: +55 11 3240-7907

E-mail: flavio.fortes.camargo@gmail.com

**RESUMO**

Este artigo propõe uma metodologia baseada em mineração de dados para a classificação de imagens polarimétricas de radar de abertura sintética (do acrônimo em inglês SAR – *synthetic aperture radar*) obtidas pelo satélite ALOS-2/PALSAR-2, visando ao mapeamento de uso e cobertura de terras. A área de estudo, com 3.660 km<sup>2</sup>, está localizada no centro-leste do estado de Goiás e em setores a norte e nordeste do Distrito Federal. A abordagem metodológica proposta combina atributos polarimétricos, segmentação multirresolução, atributos de segmentos e procedimentos iterativos e iterativos de aprendizado da máquina. Também foi utilizado, para fins de comparação e controle, um classificador *pixel-a-pixel* polarimétrico baseado em medidas de distâncias obtidas da distribuição Wishart (Polarimétrico Wishart – PW). Foram gerados 125



atributos utilizando imagens multipolarimétricas: componentes de decomposição de alvos (algoritmos de van Zyl, Freeman-Durden, Yamaguchi e Cloude-Pottier), parâmetros polarimétricos incoerentes (índices de biomassa e razões de polarização) e imagens polarizadas em amplitude (HH, HV, VH e VV). Esses atributos foram classificados utilizando os algoritmos *Naive Bayes*, J48 (árvore de decisão), *Random Forest* (Floresta Aleatória - RF), Rede Neural Artificial *Multilayer Perceptron* (Perceptron Multicamada - MLP) e *Support Vector Machine* (Máquina de Suporte Vetorial - SVM), no ambiente do *software* WEKA 3.8. Os classificadores RF, MLP e SVM apresentaram os melhores desempenhos, tendo sido considerados estatisticamente iguais (Índices Kappa de 0,66 a 0,68) em ambos os cenários propostos (legendas de nove e cinco classes). Os classificadores NB e J48 também apresentaram resultados estatisticamente iguais (Índices Kappa de 0,55 a 0,59) em ambos os cenários. O classificador J48 foi mais adequado para identificação de áreas urbanas e de cobertura vegetal natural, enquanto o classificador PW apresentou o menor desempenho entre todos os demais, porém, com potencial para boa identificação de cobertura vegetal arbórea com os dados em banda L. O fluxo de trabalho proposto é ágil e replicável inclusive para imagens obtidas por sensores de radar que operam em comprimentos de onda diferentes do satélite ALOS-2/PALSAR-2.

**Palavras-chave:** ALOS-2/PALSAR-2, polarimetria SAR, mineração de dados, uso e cobertura de terras, Cerrado.

## ABSTRACT

This paper proposes a workflow for the classification of polarimetric synthetic aperture radar (SAR) images obtained by means of the satellite ALOS-2/PALSAR-2, aiming at the mapping of land use and land cover. The study area (3,660 km<sup>2</sup>) is located in the mid-east of the Goiás State and in the northern and northeastern sectors of the Federal District. The proposed methodological approach combines polarimetric attributes, multiresolution segmentation, segment attributes and iterative machine learning procedures. A polarimetric pixel-to-pixel classifier based on distance measures obtained from the Wishart distribution (Polarimetric Wishart - PW) was also used for comparison and control purposes. A set of 125 attributes was generated using multipolarimetric images: target decomposition components (van Zyl, Freeman-Durden, Yamaguchi, and

Cloude-Pottier algorithms), incoherent polarimetric parameters (biomass indices and polarization ratios), and amplitude polarized images (HH, HV, VH, and VV). These attributes were classified using the Naive Bayes (NB), J48 (decision tree), Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM) algorithms, in the WEKA 3.8 software. The RF, MLP, and SVM classifiers presented the best performances and they were classified as statistically equal, considering scenarios of legend level (nine and five thematic classes). NB and J48 classifiers also presented statistically equal results in both scenarios. J48 classifier was the most suitable algorithm for identifying urban areas and natural vegetation cover, while the PW classifier presented the lowest performance among all the algorithms used, however, with good potential for the identification of tree coverage using L-band data. The proposed workflow is agile and also replicable for images obtained by means of SAR sensors that operate at different wavelengths from that employed by the ALOS-2/PALSAR-2 satellite.

**Keywords:** ALOS-2/PALSAR-2, SAR polarimetry, data mining, land-use and land-cover, Brazilian savanna.

#### **4.1 - Introdução**

O sensoriamento remoto representa um conjunto eficiente de tecnologias para mapeamento da superfície terrestre. A capacidade de aquisição sistemática de imagens de áreas extensas do terreno, em diferentes resoluções espaciais e em diferentes regiões do espectro eletromagnético, torna o sensoriamento remoto um importante instrumento para identificação, caracterização e quantificação dos diferentes tipos de uso e cobertura de terras (BARGIEL e HERRMANN, 2011; CONGALTON et al., 2014).

Por sua vez, dados sobre uso e cobertura de terras são essenciais em diversas atividades, tais como no planejamento urbano e regional (WENTZ et al., 2008; GAMBA e ALDRIGHI, 2012; QI et al., 2012), inventário de recursos naturais (FREITAS et al., 2008; WESTRA et al., 2010; BIRO et al., 2013; EVANS e COSTA, 2013; REYNOLDS et al., 2016) e em modelagens de processos ambientais em escalas global e regional (BRAUN e HOCHSCHILD, 2017), notadamente nos casos que envolvem redução de emissões de gases de efeito estufa relacionadas com

desmatamento e degradação florestal (ROSENQVIST et al., 2003; FEDDEMA et al., 2005; MILES e KAPOS, 2008; HAARPAINTNER et al., 2015).

Embora pouco utilizados principalmente por causa da complexidade de interpretação e porque a maioria dos seus dados são comercializados, no âmbito do mapeamento de uso e cobertura de terras, destacam-se os dados de radar de abertura sintética (do acrônimo em inglês SAR – *synthetic aperture radar*), obtidos por sensores imageadores que operam na faixa das micro-ondas do espectro eletromagnético (entre 1 e 120 cm). Os sensores de radar possuem capacidade de operação diurna ou noturna e de gerar imagens independentemente da presença de nuvens, com alguma limitação nos casos de nuvens de chuva, pois utilizam antenas que emitem sua própria radiação eletromagnética em comprimentos de onda relativamente longos (LEE e POTTIER, 2009; RICHARDS, 2009; OUCHI, 2013).

Os sensores de radar apresentam ainda outras vantagens em relação aos sensores ópticos, isto é, sensores que operam nas faixas espectrais do visível e do infravermelho. Dentre essas outras vantagens, destaca-se a independência em relação às variações das condições de iluminação solar (elevação e azimute solar) durante o dia ou ano. Imagens de radar apresentam ainda, em relação às imagens ópticas, maior sensibilidade ao conteúdo de água no solo (RAHMAN et al., 2008), à rugosidade da superfície terrestre (DUARTE et al., 2008; TOLLERUD e FANTLE, 2014) e à estrutura da vegetação (BERGEN et al., 2009), indicando, no mínimo, uma importante complementariedade em relação às imagens ópticas. Países como Alemanha, Canadá, Itália e Japão têm lançado satélites de radar que operam nas bandas X, C ou L (comprimentos de onda em torno de 3 cm, 5 cm e 23 cm, respectivamente) e em diferentes combinações de polarizações (HH, HV, VH e VV; H = polarização horizontal; V = polarização vertical; primeira letra indica polarização incidente e a segunda, polarização refletida do alvo) (OUCHI, 2013).

No caso específico do Cerrado, diversos autores têm avaliado o potencial das imagens SAR na identificação das diferentes fisionomias vegetais e dos usos e coberturas de terras. Sano et al. (2001) avaliaram o potencial de dados JERS-1 em amplitude (banda L, polarização HH) na separação de fitofisionomias de Cerrado em área do Parque Nacional de Brasília (PNB), em Brasília/DF. Estes mesmos autores também avaliaram o potencial dos dados JERS-1 na estimativa de índice de área foliar (IAF).

Os resultados produzidos por Sano et al. (2001) foram promissores, possibilitando um seguinte trabalho no PNB explorando dados multitemporais JERS-1 e LANDSAT-5/TM, abarcando as estações seca e chuvosa, de modo a avaliar a sinergia entre dados SAR e ópticos e também avaliar os potenciais e limitações de cada conjunto de dados. Os dados SAR na banda L apresentaram maior discriminação de tipos de vegetação, dada a capacidade destes dados em penetrar nos dosséis e, assim, possibilitar uma melhor caracterização da estrutura das fitofisionomias (SANO et al., 2005).

Bitencourt et al. (2007) também avaliaram o potencial de imagens dos sensores JERS-1 (banda L, polarização HH) e LANDSAT-5/TM na estimativa de parâmetros da vegetação de Cerrado em áreas no estado de São Paulo. Os dados JERS-1 foram especificamente utilizados para modelagem e estimativa de biomassa lenhosa acima do solo devido à capacidade dos dados SAR de interagir com a estrutura (troncos e galhos) da vegetação. Os modelos ajustados atingiram valores de coeficientes de determinação ( $R^2$ ) de até 0,87, entre biomassa lenhosa acima do solo e  $\sigma^0$  (dB).

No que tange à produção de mapas temáticos, Sano et al. (2009) analisaram o potencial de dados ALOS/PALSAR e algoritmos de segmentação e classificação visando à interpretação visual e posterior classificação de padrões de uso e cobertura de terras em área de Cerrado no Distrito Federal. Estes autores obtiveram um índice de acerto Tau ( $\tau$ ) de 0,70, para uma legenda com 10 classes.

Considerando a vegetação de savana de forma mais abrangente, diversos autores avaliaram o potencial de dados ALOS/PALSAR no estudo de áreas deste domínio de vegetação localizadas no continente Africano. Braun e Hochschild (2017) desenvolveram índices baseados em séries temporais ALOS/PALSAR para identificar e avaliar mudanças na paisagem de savana em área no entorno do campo de refugiados de Djabal, no leste de Chade. O algoritmo RF foi empregado para gerar as classificações de uso e cobertura de terras utilizadas nas métricas de paisagem a serem empregadas em tomadas de decisões humanitárias e ambientais.

Por sua vez, Bouvet et al. (2018) utilizaram mosaico ALOS/PALSAR para desenvolver estimativas de biomassa acima do solo para todo o continente Africano. Do mesmo modo, Odipo et al. (2016), utilizando imagens ALOS/PALSAR, também desenvolveram um modelo de estimativa de biomassa acima do solo, calibrado e validado com dados LIDAR, para área localizada no *Kruger National Park*, nordeste da África do Sul.

Com relação ao mapeamento temático, Symeonakis et al. (2018) avaliaram o classificador RF aplicados a imagens ALOS/PALSAR e LANDSAT visando ao aprimoramento dos mapeamentos de uso e cobertura de terras em área entre África do Sul e Botsuana. Considerando as classificações que utilizaram somente imagens SAR, a acurácia global atingiu 82% com uma legenda de quatro classes. Urbazev et al. (2015) empregaram imagens ALOS/PALSAR para o mapeamento de cobertura vegetal lenhosa em área de savana na África do Sul. Diferente dos demais trabalhos apresentados, esses autores testaram atributos polarimétricos gerados por meio dos algoritmos de van Zyl e Freeman-Durden. Empregou-se o classificador RF, sendo treinado e validado com conjuntos independentes extraídos de levantamento LIDAR. As decomposições foram utilizadas com o intuito de explorar as componentes referentes ao espalhamento volumétrico para identificação de biomassa lenhosa (troncos e galhos). Segundo os autores, as polarizações cruzadas (HV e VH) contribuíram mais ao modelo.

Concomitantemente ao avanço do sensoriamento remoto, a mineração de dados (MD) surgiu como uma técnica voltada ao tratamento, análise e descoberta de conhecimento em diferentes bases de dados (PIATETSKY-SHAPIRO e FAYYAD, 2012; TSAI, 2012; 2013). A MD corresponde a um campo interdisciplinar que combina inteligência artificial, gerenciamento e visualização de dados, aprendizado de máquina, algoritmos matemáticos e estatística. Ela oferece diferentes metodologias para tomadas de decisão, resoluções de problemas, análises, planejamentos, diagnósticos, detecções de padrões, integrações, prevenções, aprendizados e inovações. As publicações científicas empregando MD têm crescido em áreas variadas de atuação, incluindo-se as áreas de Geografia e Ciência da Informação Geográfica (TSAI, 2012; 2013).

Dentro das atividades de mapeamento de uso e cobertura de terras com dados de sensoriamento remoto, podem ser destacados diversos estudos que utilizaram, com sucesso, as diferentes técnicas de MD, tais como *Naive Bayes* (NB) (EILANDER et al., 2014), árvores de decisão (ADs) (SILVA et al., 2008; BISWAL et al., 2013; KÖRTING et al., 2013; SHARMA et al., 2013; PASSO et al., 2014; ANJOS et al., 2017), o classificador *Random Forest* (Floresta Aleatória - RF) (OK et al., 2012; BELGIU e DRAGUT, 2016; ANJOS et al., 2017), as Redes Neurais Artificiais (RNA) (FOODY, 2017) e o *Support Vector Machine* (Máquina de Suporte Vetorial - SVM) (ZHANG et al., 2015).

Com relação ao uso da técnica de MD para classificar imagens de radar, destaca-se o estudo conduzido por Waske e Braun (2009) que avaliaram o potencial dos algoritmos

*Boosted Decision Tree* (Árvore de Decisão Reforçada) e RF para classificar séries temporais de dados SAR obtidos na banda C (comprimento de onda em torno de 5 cm) de diferentes áreas. Qi et al. (2012) propuseram uma nova abordagem de classificação que envolvia o uso de imagens polarimétricas e interferométricas de radar na banda C, processadas pelos algoritmos de segmentação de imagens e AD. Topouzelis e Psyllos (2012) desenvolveram um novo método para detectar manchas de óleo no mar por meio de classificação de imagens SAR pela técnica de RF. Shiraishi et al. (2014) investigaram a precisão e a velocidade de processamento dos algoritmos *Naive Bayes*, *AdaBoost*, *Multilayer Perceptron* (Perceptron Multicamada - MLP), RF e SVM na classificação de mosaicos de imagens de radar do satélite ALOS/PALSAR. Uhlmann e Kiranyaz (2014) empregaram SVM e RF na classificação e avaliação de atributos de cores extraídos da decomposição polarimétrica de dados SAR obtidos na banda C. Du et al. (2015) estudaram o potencial de atributos texturais e polarimétricos extraídos de imagens SAR obtidas na banda C para mapear diferentes classes de uso e cobertura de terras. Para isso, foram utilizados os algoritmos *Rotation Forest* (Floresta de Rotação), RF e SVM.

Dentro desse contexto, este artigo apresenta um novo procedimento de classificação de imagens SAR polarimétricas obtidas na banda L pelo satélite ALOS-2/PALSAR-2 para discriminar classes de uso e cobertura de terras em área de Cerrado. O procedimento envolve o uso combinado de parâmetros polarimétricos, técnicas de segmentação de imagens e algoritmos de aprendizado de máquina (NB, AD, RF, MLP e SVM). Ainda, classificações com um classificador polarimétrico específico são realizadas de modo a servirem de controle e comparação dos resultados obtidos pela nova abordagem proposta. Os procedimentos são interativos, iterativos e replicáveis, possibilitando a obtenção de resultados rápidos e com altos índices de acerto.

Este artigo está organizado conforme se segue: na Seção 4.2, são descritas características do sistema de imageamento do satélite ALOS-2/PALSAR-2. Na Seção 4.3, é apresentada a área de estudo. Na Seção 4.4, o material e os métodos utilizados são apresentados e descritos. Os resultados e discussões são apresentados na Seção 4.5, enquanto na Seção 4.6 são destacadas as principais conclusões obtidas por este estudo.

## 4.2 - A missão ALOS-2/PALSAR-2

O satélite japonês denominado de *Advanced Land Observing Satellite-2* (Satélite Avançado de Observação da Terra 2 - ALOS-2), foi lançado em 24 de maio de 2014 e opera com um sensor SAR denominado de *Phase Array L-band Synthetic Aperture Radar-2* (Radar de Abertura Sintética Banda L do tipo Matriz Faseada 2 - PALSAR-2). O ALOS-2 corresponde ao segundo satélite da série ALOS e foi lançado com a missão de garantir consistência e continuidade dos dados SAR na banda L (comprimento de onda de 23,5 cm). O ALOS-2/PALSAR-2 utiliza três modos distintos de imageamento denominados de *Spotlight*, *StripMap* e *ScanSAR*, com diferentes resoluções espaciais, faixas de imageamento e polarizações (ROSENQVIST et al., 2014) (Quadro 4.1).

Quadro 4.1. Detalhes dos diferentes modos de aquisição do satélite ALOS-2/PALSAR-2.

| Modo de observação   | <i>Spotlight</i> | <i>StripMap</i>  |                       |             | <i>ScanSAR</i> |             |
|----------------------|------------------|------------------|-----------------------|-------------|----------------|-------------|
|                      |                  | <i>Ultrafine</i> | <i>High Sensitive</i> | <i>Fine</i> | Normal         | <i>Wide</i> |
| Resolução espacial   | 3 m              | 3 m              | 6 m                   | 10 m        | 100 m          | 60 m        |
| Faixa de imageamento | 25 km            | 50 km            | 50 km                 | 70 km       | 350 km         | 480 km      |
| Polarização          | HH               | HH               | HH, HV, VH, VV        | HH, HV      | HH, HV         | HH, HV      |

Fonte: Adaptado de Rosenqvist et al. (2014).

As missões espaciais do ALOS são o cerne de um plano global de aquisições de dados de observação da Terra denominado de *Basic Observation Scenario* (Cenário Básico de Observação - BOS) que procura manter em funcionamento uma estratégia de atendimento simultâneo a diferentes comunidades de usuários e áreas de aplicação. Aqui, o objetivo é minimizar conflitos durante as aquisições por meio de um planejamento pré-estabelecido (SUZUKI et al., 2013; JAXA, 2014).

## 4.3 - Área de estudo

A área de estudo, com aproximadamente 3.660 km<sup>2</sup> de abrangência, localiza-se na porção centro-leste do estado de Goiás, abrangendo também áreas a norte e nordeste do Distrito Federal (Figura 4.1). Esse recorte espacial consiste em uma cena ALOS-

2/PALSAR-2 obtida no modo *Stripmap/High Sensitive*, órbita ascendente (de sul para norte com apontamento da antena para leste).

Esta área foi selecionada devido à diversidade relativamente grande de tipos de uso e cobertura de terras, com destaque para a presença de vegetação nativa de Cerrado, áreas agrícolas predominantemente no centro da cena e também áreas urbanas consolidadas na porção centro-sul (INPE, 2017; MAPBIOMAS, 2017). Com relação às atividades agropecuárias, ocorrem, em grande extensão, pastagens e culturas agrícolas anuais (INPE, 2017; MAPBIOMAS, 2017), especialmente milho e soja (IBGE, 2017). Por outro lado, no que tange às áreas naturais de Cerrado, verifica-se a presença de formações florestais, savânicas e campestres (MAPBIOMAS, 2017).

A área de estudo encontra-se em relevo predominantemente de gênese denudacional, sendo uma superfície regional de aplainamento. Destaque é dado à Serra Geral do Paranã, que atravessa a cena ALOS-2/PALSAR-2. Essa serra é formada por metassedimentos dobrados (anticlinais e sinclinais), afetados posteriormente por processos de falhamentos inversos (GOIÁS, 2006).

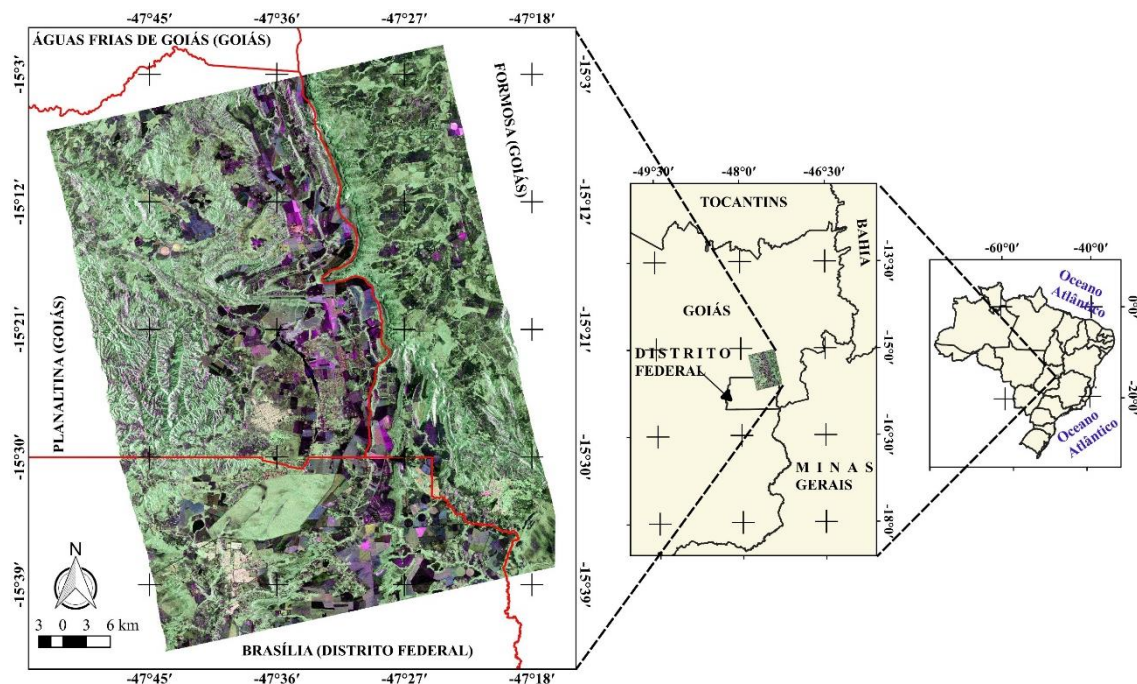


Figura 4.1. Localização da área de estudo, entre Goiás e Distrito Federal. A imagem corresponde a uma composição colorida RGB (RGB/HH-HV-VV) da cena do satélite ALOS-2/PALSAR-2, obtida em 14 de maio de 2016. *Datum* horizontal: WGS84; Projeção: UTM 23S.



## 4.4 - Material e Métodos

### 4.4.1 - Material

Neste experimento, foram empregadas imagens ALOS-2/PALSAR-2 polarimétricas (modo SM2 *polarimetry – High Sensitive*), nas quatro polarizações disponíveis (HH, HV, VH e VV) e no nível de processamento 1.1, ou seja, imagens em formato complexo (*Single Look Complex - SLC*), com um *look* em azimute, comprimidas em alcance (*slant range*) e apresentando as informações sobre as intensidades e as fases dos sinais registrados (Quadro 4.2).

Quadro 4.2. Relação de imagens utilizadas na pesquisa.

| Imagens utilizadas | Características principais   | Resolução espacial | Datas das passagens       | Fontes |
|--------------------|--|--------------------|---------------------------|--------|
| ALOS-2/PALSAR-2    | <ul style="list-style-type: none"> <li>- 01 cena <i>Single Look Complex</i> (formato CEOS; nível L1.1);</li> <li>- Modo <i>Stripmap/High Sensitive</i>;</li> <li>- Banda L (<math>\lambda = 23,5</math> cm)</li> <li>- <i>Full polarization/Full polarimetric</i>;</li> <li>- Ângulo de incidência no centro da cena = 27,8°;</li> <li>- Órbita ascendente.</li> </ul> | ~ 6 m              | 14/05/2016                | JAXA   |
| LANDSAT-8/OLI      | - 04 bandas multiespectrais (visível ao infravermelho próximo)   | 30 m               | 02/05/2016;<br>03/06/2016 | USGS   |
|                    | - 01 uma banda pancromática  | 15 m               |                           |        |

(continua)

Quadro 4.2 - Conclusão

| Imagens utilizadas                         | Características principais  | Resolução espacial               | Datas das passagens | Fontes   |
|--|---|----------------------------------|---------------------|--|
| - Imagens Google©<br>- Imagens Bing©/HERE© | Imagens de alta resolução espacial de diversos sensores orbitais (no espectro do visível) | Diversas (incluindo submétricas) | Diversas datas      | Acesso <i>on-line</i> via <i>QuickMapServices</i> (NEXTGIS, 2018) e QGIS 3.0 |

Foram utilizadas ainda as seguintes imagens auxiliares para a interpretação visual das amostras de validação: imagens multiespectrais de média resolução espacial do satélite LANDSAT-8/OLI; e imagens na região do espectro visível e em alta resolução espacial de diversos sensores disponíveis nas bases de imagens *Google©* e *Bing©/HERE©*, e acessadas por meio da extensão *QuickMapServices* do QGIS 3.0 (NEXTGIS, 2018).

Com relação a outros tipos de dados auxiliares (Quadro 4.3), foram analisados os relatórios da Produção Agrícola Municipal (PAM) do ano de 2015, do Instituto Brasileiro de Geografia e Estatística (IBGE) (IBGE, 2017), referentes aos municípios inseridos na área de estudo. Foram ainda consultadas e analisadas as bases cartográficas de uso e cobertura do solo MapBiomias (MAPBIOMAS, 2017) e TerraClass (INPE, 2017), além de dados sobre precipitação pluviométrica na área de estudo (INMET, 2018).

A respeito dos aplicativos computacionais, foram utilizados o SNAP 6.0 (pré-processamento dos dados L1.1), *eCognition Developer* 8.7 (segmentação e extração de atributos por segmento), WEKA 3.8 (mineração de dados e classificação) e PCI Geomatica 2017 (pré-processamentos dos dados no nível 1.1 e classificação polarimétrica).

Quadro 4.3. Relação de dados e de material auxiliar utilizados na pesquisa.

| Dados auxiliares            | Características principais  | Data de produção | Fonte |
|-----------------------------|---|------------------|-------|
| Produção Agrícola Municipal | - Inventário por município da área total plantada referente a culturas temporárias e permanentes. | 2015             | IBGE  |

(continua)

Quadro 4.3 - Conclusão

| Dados auxiliares   | Características principais   | Data de produção    | Fonte      |
|--|--|---------------------|------------|
| Base cartográfica<br>MapBiomias de uso e<br>cobertura do solo            | - Mapas e dados<br>históricos de uso e<br>cobertura do solo<br>disponíveis na <i>internet</i> .  | 2016                | MapBiomias |
| Base cartográfica<br>TerraClass Cerrado<br>de uso e cobertura do<br>solo | - Mapa de uso e<br>cobertura do solo do<br>bioma Cerrado (GO e<br>DF).   | 2013                | INPE       |
| Precipitação<br>pluviométrica  | - Dados diários de<br>precipitação<br>pluviométrica - estação<br>meteorológica<br>automática da Estação<br>Ecológica de Águas<br>Emendadas<br>(Brasília/DF). | Abril-<br>Maio/2016 | INMET      |

#### 4.4.2 - Fluxo metodológico

O fluxo metodológico é apresentado na Figura 4.2. Os principais grupos de atividades foram: geração da referência para validação temática; classificação *pixel-a-pixel* utilizando o classificador polarimétrico Wishart; geração de imagens polarizadas em amplitude, parâmetros polarimétricos, decomposições polarimétricas, ortorretificação etc.; mineração de dados e classificações usando aprendizado de máquina; e validações temáticas. As metodologias e procedimentos utilizados no fluxo da Figura 4.2 são detalhados na seção seguinte.

#### 4.4.3 - Métodos

##### 4.4.3.1 - Legenda e chaves de interpretação

Para a definição da legenda utilizada neste experimento (Quadro 4.4), foram consultados os tipos de atividades agropecuárias predominantes nos municípios que compõem a área de estudo por meio dos relatórios de 2015 da PAM. Com relação às culturas agrícolas, predominavam os cultivos temporários, totalizando em torno de

85.000 hectares de milho e 132.000 hectares de soja (IBGE, 2017). Esses números representam uma considerável ordem de grandeza haja vista que são estimativas relativas ao total da área de todos os municípios e não somente ao recorte espacial utilizado neste experimento. Todavia, esses números forneceram uma noção prévia de quais tipos de cultivos poderiam ser verificados nas visitas de campo realizadas no mês de setembro de 2017.

Posteriormente, também foram verificados os mapas de uso e cobertura da terra do projeto MapBiomas (MAPBIOMAS, 2017), ano de 2016, e do programa de monitoramento do Cerrado TerraClass (INPE, 2017), este referente ao ano de 2013. Utilizando essas bases cartográficas, foi possível verificar a distribuição das áreas naturais de Cerrado e também as áreas referentes a pastagens.

Também foi necessária a geração de mosaicos fusionados (15 m de resolução espacial) LANDSAT-8/OLI de 02 de maio de 2016 e 03 de junho de 2016 por meio de imagens disponíveis gratuitamente para *download* no *Global Visualization Viewer* (Visualizador de Visualização Global - GloVis) (USGS, 2017). Esses mosaicos abrangem a faixa espectral do visível ao infravermelho próximo. Para a fusão, foi empregado o algoritmo *pansharpening* do aplicativo PCI Geomatica.

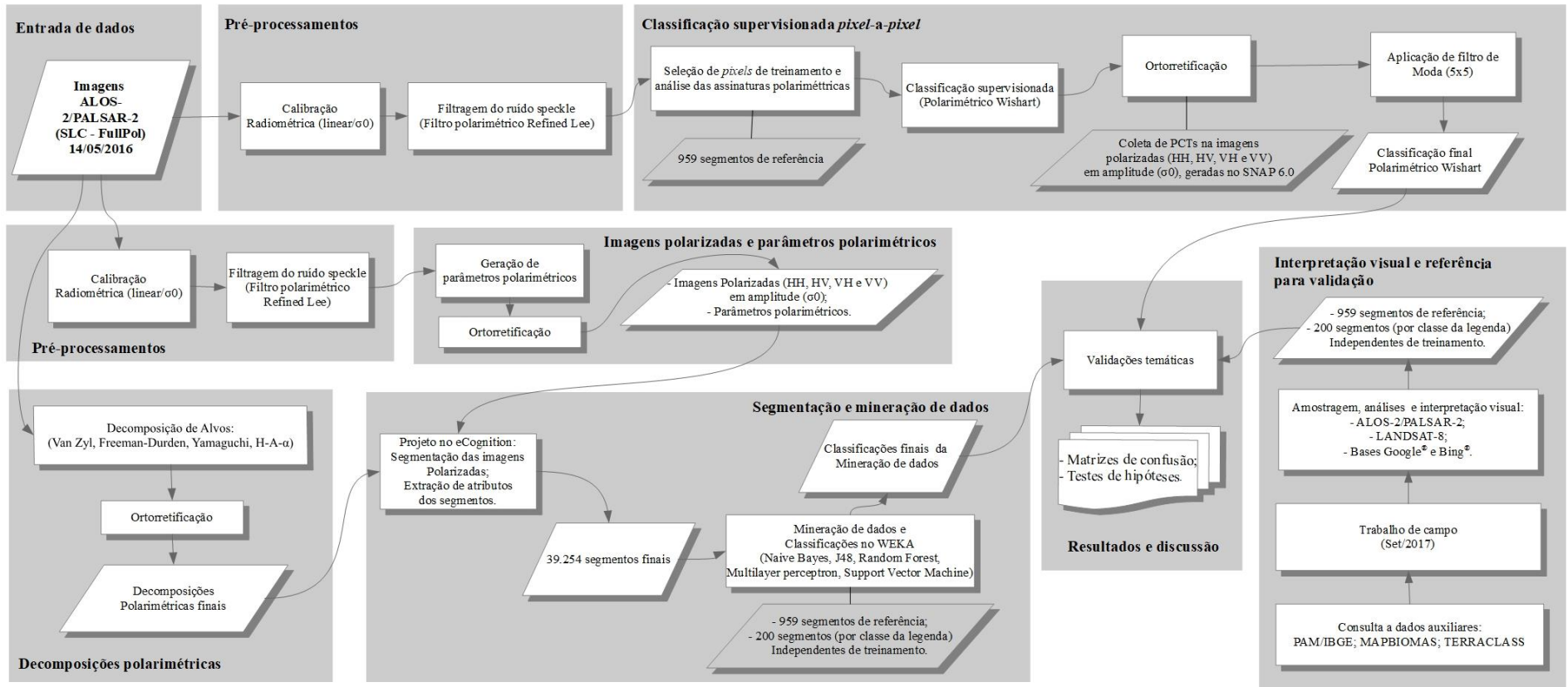


Figura 4.2. Fluxograma das principais etapas do trabalho.

De posse destes dados, atividades de campo foram realizadas nos dias 10 e 11 de setembro de 2017, percorrendo as rodovias BR-010 e GO-118, para verificação *in loco* dos tipos de uso e cobertura da terra. A Figura 4.3 apresenta a área de estudo, com o trajeto percorrido e alguns pontos visitados. Também são apresentados os pontos amostrais aleatórios e independentes posteriormente interpretados e utilizados na validação temática das classificações.

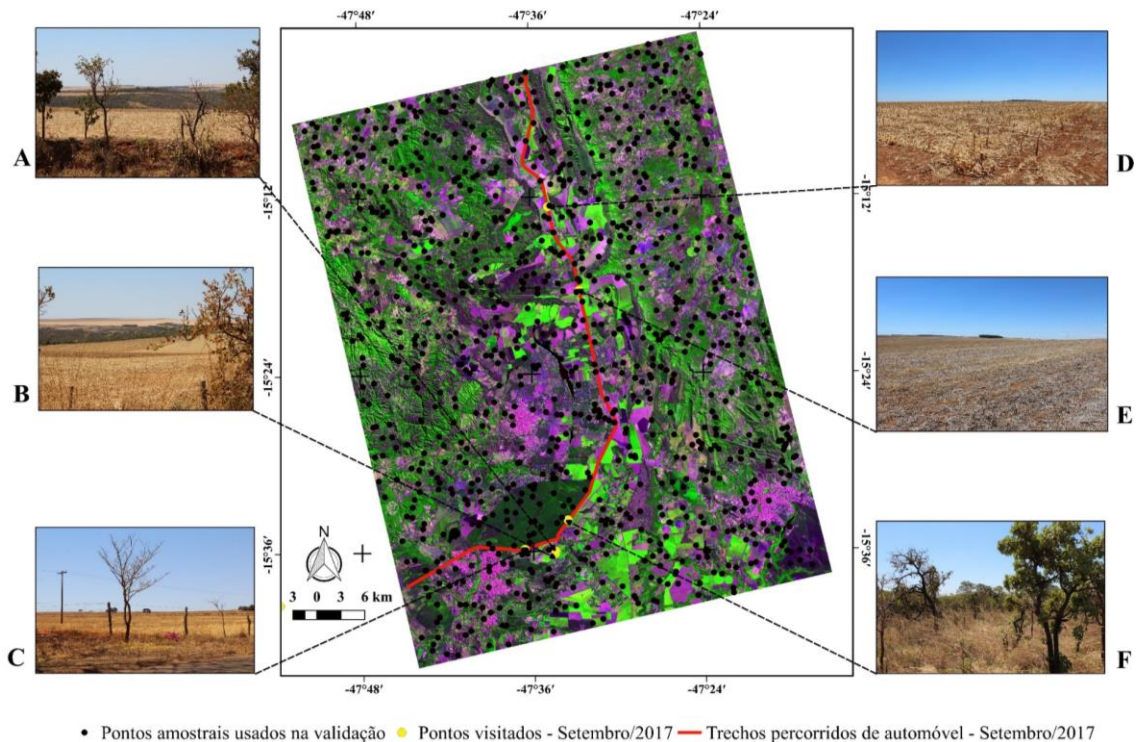



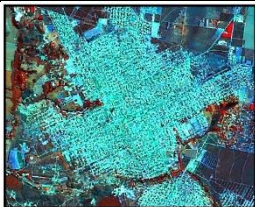
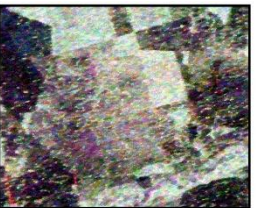
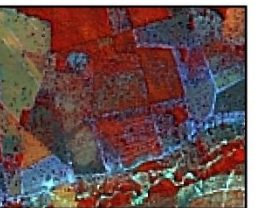
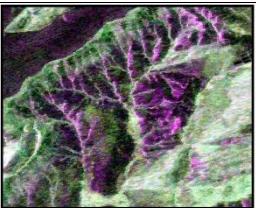
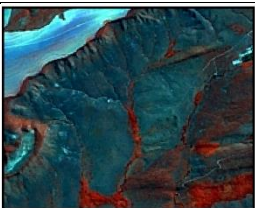
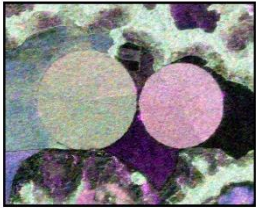

Figura 4.3. Composição colorida RGB LANDSAT-8/OLI (RGB/B4-B5-B3): Distribuição dos pontos amostrais utilizados na validação, pontos visitados em campo e trechos percorridos de automóvel (trabalho de campo realizado em setembro de 2017). Datum horizontal: WGS84; Projção: UTM 23S.

O tamanho da amostra para validação foi determinado de acordo com estudos de Congalton e Green (2009). Segundo esses autores, para áreas de estudo complexas e com extensões superiores a 4.047 km<sup>2</sup>, devem-se utilizar no mínimo 100 pontos amostrais por classe da legenda. Neste trabalho, foram empregados 1.000 pontos aleatórios e sem estratificação para uma legenda de nove classes. O Quadro 4.4 apresenta as classes da legenda e suas respectivas chaves de interpretação visual.

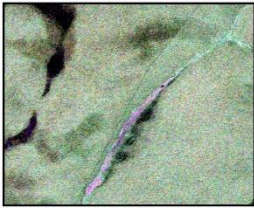
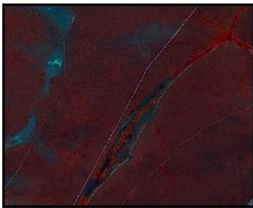
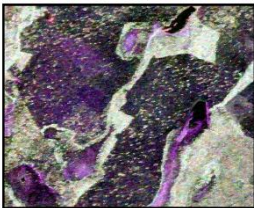
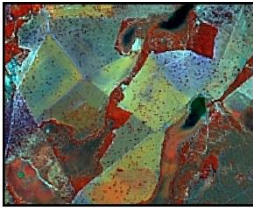

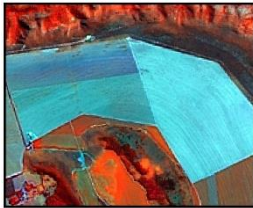

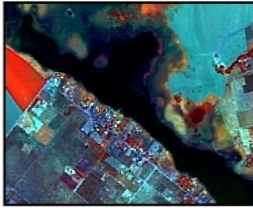
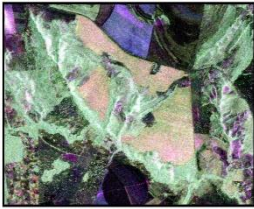
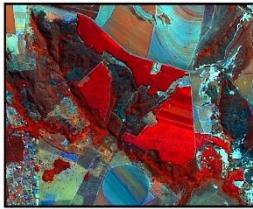
Também foram solicitados ao Instituto Nacional de Meteorologia (INMET), dados da estação meteorológica automática localizada na Estação Ecológica Águas Emendadas, com o intuito de se verificar se houve ou não precipitação pluviométrica

significativa no dia da passagem do satélite ALOS-2 (14 de maio de 2016), ou nos dias que antecederam à passagem. Os dados disponibilizados (ANEXO B) demonstraram que não houve precipitação pluviométrica, portanto, pode-se assumir que houve pouca alteração no sinal retroespalhado por conta do conteúdo de umidade do solo.

Quadro 4.4. Classes da legenda e respectivas chaves de interpretação visual utilizando composições coloridas RGB ALOS-2/PALSAR-2 (RGB/HH-HV-VV) e LANDSAT-8/OLI (RGB/B5-B3-B4). *Datum* horizontal: WGS84; Projeção: UTM 23S.

| Classe   | Elementos de interpretação   | Composição colorida ALOS-2/PALSAR-2   | Composição colorida LANDSAT-8/OLI   |
|--|--|---|---|
| Áreas com Edificações                                      | <ul style="list-style-type: none"> <li>- Formas regulares;</li> <li>- Predomínio de espalhamento de canto;</li> <li>- Brilho elevado.</li> </ul>   |  <p>0 1 2 km</p>   |  <p>0 1 2 km</p>   |
| Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo | <ul style="list-style-type: none"> <li>- Formas irregulares;</li> <li>- Adjacentes a áreas de cobertura arbórea;</li> <li>- Predomínio de espalhamento volumétrico;</li> <li>- Brilho moderado.</li> </ul>   |  <p>0 1 2 km</p>  |  <p>0 1 2 km</p>  |
| Campo Limpo  | <ul style="list-style-type: none"> <li>- Formas irregulares;</li> <li>- Ocorrência em áreas de relevo ondulado;</li> <li>- Espalhamento de superfície;</li> <li>- Tons escuros.</li> </ul>   |  <p>0 1 2 km</p> |  <p>0 1 2 km</p> |
| Culturas temporárias                                       | <ul style="list-style-type: none"> <li>- Formas geométricas regulares (círculos e/ou retângulos);</li> <li>- Predomínio de espalhamento volumétrico;</li> <li>- Brilho moderado a forte;</li> <li>- Ocorrem em terrenos planos/pouco ondulados.</li> </ul> |  <p>0 1 2 km</p> |  <p>0 1 2 km</p> |

Quadro 4.4 - Conclusão

| Classe                   | Elementos de interpretação   | Composição colorida ALOS-2/PALSAR-2   | Composição colorida LANDSAT-8/OLI   |
|--------------------------|--|---|---|
| Formações florestais     | <ul style="list-style-type: none"> <li>- Formas geométricas regulares e irregulares;</li> <li>- Predomínio de espalhamento volumétrico;</li> <li>- Brilho forte.</li> </ul>  |  <p>0 1 2 km</p>   |  <p>0 1 2 km</p>   |
| Pastagens                | <ul style="list-style-type: none"> <li>- Formas geométricas regulares e irregulares;</li> <li>- Predomínio de espalhamento de superfície com pequenas áreas de espalhamentos volumétricos e de canto;</li> <li>- Tons escuros salpicados de pontos brilhosos.</li> </ul> |  <p>0 1 2 km</p>   |  <p>0 1 2 km</p>   |
| Solo exposto/palhada     | <ul style="list-style-type: none"> <li>- Formas geométricas regulares;</li> <li>- Predomínio de espalhamento de superfície;</li> <li>- Geralmente adjacentes a áreas de culturas temporárias;</li> <li>- Tons escuros.</li> </ul>  |  <p>0 1 2 km</p> |  <p>0 1 2 km</p> |
| Corpo d'água continental | <ul style="list-style-type: none"> <li>- Formas geométricas irregulares;</li> <li>- Espalhamento de superfície (quase-especular);</li> <li>- Pouca ocorrência;</li> <li>- Tons escuros.</li> </ul>   |  <p>0 1 2 km</p> |  <p>0 1 2 km</p> |
| Silvicultura             | <ul style="list-style-type: none"> <li>- Formas regulares (retângulos);</li> <li>- Predomínio de espalhamento volumétrico;</li> <li>- Pouca ocorrência;</li> <li>- Brilho moderado.</li> </ul>   |  <p>0 1 2 km</p> |  <p>0 1 2 km</p> |



Segundo Ribeiro e Walter (2008), a composição fisionômica da vegetação no bioma Cerrado pode ser dividida em três grandes formações: florestais, savânicas e campestres. As formações florestais são constituídas de Mata Ciliar, Mata de Galeria, Mata Seca e Cerradão. Por sua vez, a formação savânica é o próprio Cerrado em sentido estrito (áreas com árvores e arbustos espalhados sobre um estrato gramíneo e sem um dossel contínuo) que pode ser subdividido em Cerrado Denso, Cerrado Típico, Cerrado Ralo e Cerrado Rupestre, além de Vereda, Parque de Cerrado e Palmeiral. Por fim, as formações campestres são representadas pelo Campo Sujo, Campo Limpo e Campo Rupestre.

Na legenda proposta no Quadro 4.4, há três classes que representam cobertura vegetal natural: Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; Campo Limpo; e Formações florestais. Observa-se que, devido a limitações inerentes às imagens utilizadas, houve a necessidade de se realizar adaptações e agregações em algumas classes apresentadas por Ribeiro e Walter (2008).

#### 4.4.3.2 - Pré-processamentos

##### Calibração radiométrica

O procedimento de calibração para coeficiente de retroespalhamento ( $\sigma^0$ ) foi realizado nas imagens com nível de processamento 1.1, utilizando o modelo definido por Shimada et al. (2013) e JAXA (2018). A seguinte formulação matemática foi utilizada (Eq. 4.1):

$$\sigma_{slc}^0 = 10 \cdot \log_{10} \langle I^2 + Q^2 \rangle + CF_1 - A \quad (4.1)$$

Em que  $I$  e  $Q$  são as partes real e imaginária respectivamente das imagens SLC.  $CF$  corresponde ao fator de calibração radiométrica (-83 dB), e  $A$  é o fator de conversão (32 dB) (SHIMADA et al., 2013; JAXA, 2018).

Nesta pesquisa, para a realização das filtragens e da classificação polarimétrica, os dados calibrados utilizaram uma escala linear para, após a obtenção dos produtos, serem convertidos de escala linear para uma escala logarítmica, em decibéis (dB).

## Filtragem do ruído *speckle*

A filtragem do ruído *speckle* empregou o filtro polarimétrico *Refined Lee*. Esse filtro foi selecionado com base em resultados verificados na literatura científica (LEE e POTTIER, 2009; YOMMY et al., 2015), os quais apontam desempenho adequado desse filtro na remoção de ruídos, preservando tanto as estatísticas (de modo similar à técnica *multilook*) quanto as feições pontuais e lineares das imagens (LEE e POTTIER, 2009; YOMMY et al., 2015). O filtro *Refined Lee* emprega uma janela adaptativa, que se alinha às feições de borda da imagem e realiza a filtragem na matriz de covariância, empregando pesos extraídos da imagem *span* (médias das intensidades de HH, HV+VH e VV) (LEE e POTTIER, 2009).

Nesta pesquisa, foram filtradas, para atenuação do ruído *speckle*, duas matrizes de covariância, a saber: uma gerada no aplicativo PCI Geomatica 2017, para a classificação com o algoritmo PW, e outra gerada no aplicativo SNAP 6.0, que foi empregada na geração das imagens polarizadas ( $\sigma^0$ ) (HH, HV, VH e VV) e também na geração dos parâmetros polarimétricos incoerentes. As matrizes de covariância e de coerência utilizadas nas decomposições polarimétricas não foram submetidas a filtrações para atenuação do ruído *speckle*. Cabe destacar que as decomposições polarimétricas foram geradas inteiramente no aplicativo SNAP 6.0.

No aplicativo PCI Geomatica 2017, utilizou-se o algoritmo *SARINGEST* para importação das imagens SLC. Esse algoritmo realiza a leitura do *orbit* das imagens e gera a matriz adequada para cada sistema sensor. No caso dos dados utilizados nesta pesquisa, gerou-se uma matriz de covariância 4 x 4. Não há qualquer menção, na documentação do algoritmo e do aplicativo, sobre o tamanho de janela empregada na construção da referida matriz de covariância.

No aplicativo SNAP 6.0, o algoritmo que aplica o filtro *Refined Lee* emprega diretamente como entrada as imagens SLC dispostas na matriz de espalhamento de Jones e, gera, na sequência, a matriz de covariância sem solicitar a especificação do tamanho de janela para o cálculo dessa matriz. Não há também qualquer menção, na documentação do algoritmo e do aplicativo SNAP 6.0, sobre o tamanho de janela empregada na construção da referida matriz de covariância.

Para fins de mineração de dados, aplicou-se uma janela 7 x 7 (Figura 4.4), dado que os efeitos do ruído *speckle* são diluídos nos atributos extraídos por segmentos. Todavia,

para a classificação polarimétrica *pixel-a-pixel*, verificou-se a necessidade da aplicação de uma janela maior (9 x 9) (Figura 4.5) visando homogeneização dos padrões de uso e cobertura da terra.

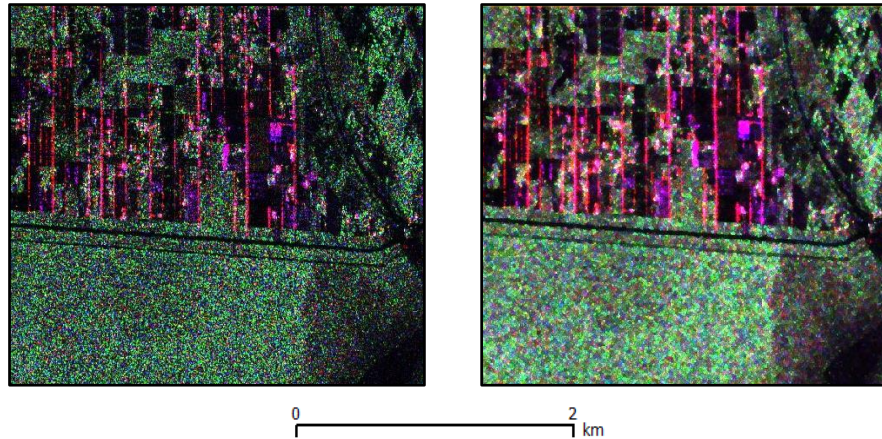


Figura 4.4. Filtragem do ruído *speckle* utilizando o filtro *Refined Lee* e janela 7 x 7.

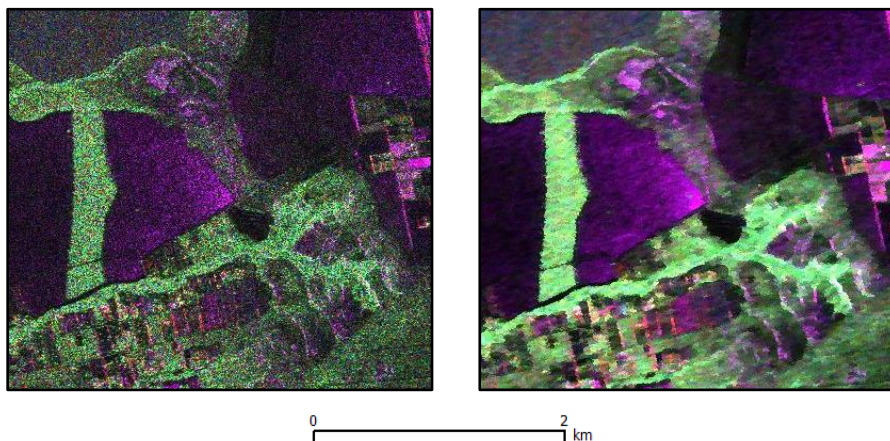


Figura 4.5. Filtragem do ruído *speckle* utilizando o filtro *Refined Lee* e janela 9 x 9.

#### 4.4.3.3 - Parâmetros polarimétricos incoerentes

Os parâmetros polarimétricos incoerentes consistem naqueles derivados diretamente das medidas de potência em  $\sigma^0$ , em diferentes polarizações, sem a utilização da informação das fases relativas dos elementos da matriz de espalhamento (BOERNER et al., 1998). Nesta pesquisa, esses parâmetros foram gerados para comporem o conjunto de atributos empregados na fase de mineração de dados. Os seguintes índices foram gerados: *Radar Vegetation Index* (Índice de Vegetação por Radar - RVI) (KIM e VAN ZYL, 2009); *Radar Forest Degradation Index* (Índice de Degradação Florestal por

Radar - RFDI) (MITCHARD et al., 2012); *Canopy Structure Index* (Índice de Estrutura de Dossel - CSI); *Volume Scattering Index* (Índice de Espalhamento Volumétrico - VSI); e *Biomass Index* (Índice de Biomassa - BMI) (POPE et al., 1994). Além desses índices, também foram geradas as razões de polarização paralela (*co-pol*) e cruzada (*cross-pol*) (BOERNER et al., 1998). O Quadro 4.5 apresenta um resumo dos parâmetros polarimétricos incoerentes utilizados nesta pesquisa.

O RVI é um índice sensível ao volume de biomassa, apresentando alta correlação com o conteúdo de água na vegetação e também com a área foliar (KIM e VAN ZYL, 2009). Por sua vez, o RFDI é um índice definido para destacar a ocorrência do espalhamento *double bounce* (dupla batida) e é aplicado para separação entre áreas florestadas e não-florestadas. A polarização HH é igualmente sensível aos espalhamentos *double bounce* e volumétrico. Por outro lado, na polarização HV, prevalece a sensibilidade ao espalhamento volumétrico (MITCHARD et al., 2012).

De acordo com Pope et al. (1994), CSI, VSI e BMI são índices biofísicos relacionados às características da vegetação. CSI é uma medida relativa da importância das estruturas vegetais verticais (troncos) em relação às horizontais. Em áreas dominadas por troncos e galhos e/ou ramos verticais, CSI vai apresentar altos valores, sendo assim um indicador útil na caracterização da cobertura vegetal. Por sua vez, VSI é uma medida do grau de despolarização do sinal retroespalhado, sendo utilizado como um indicador da densidade e espessura de dosséis. Por fim, BMI é um indicador relativo à quantidade e presença de material lenhoso em relação ao material foliar da cobertura vegetal. Por não ser normalizado, sofre interferência do relevo da área de estudo (POPE et al., 1994).

Com relação às razões de polarizações, a razão de polarização paralela tem o potencial de destacar as diferenças entre espalhamentos sensíveis à polarização horizontal (HH) e os sensíveis à polarização vertical (VV). As razões de polarização cruzada destacam a dominância dos espalhamentos nas polarizações paralelas HH e VV, respectivamente às polarizações cruzadas HV e VH (BOERNER et al., 1998; ESA, 2018).

Quadro 4.5. Relação de parâmetros polarimétricos incoerentes empregados nesta pesquisa.

| Parâmetro polarimétrico  | Formulação matemática   | Fonte                  |
|--|---|------------------------|
| <i>Radar Vegetation Index</i><br>(Índice de Vegetação por Radar - RVI)                     | $RVI = \frac{8\sigma_{hv}^0}{\sigma_{hh}^0 \sigma_{vv}^0 2\sigma_{hv}^0}$   | Kim e Van Zyl (2009)   |
| <i>Radar Forest Degradation Index</i><br>(Índice de Degradação Florestal por Radar - RFDI) | $RFDI = \frac{\sigma_{hh}^0 - \sigma_{hv}^0}{\sigma_{hh}^0 + \sigma_{hv}^0}$  | Mitchard et al. (2012) |
| <i>Canopy Structure Index</i><br>(Índice de Estrutura de Dossel - CSI)                     | $CSI = \frac{\sigma_{vv}^0}{\sigma_{vv}^0 + \sigma_{hh}^0}$   | Pope et al. (1994)     |
| <i>Volume Scattering Index</i><br>(Índice de Espalhamento Volumétrico - VSI)               | $VSI = \frac{(\sigma_{hv}^0 + \sigma_{vh}^0)/2}{((\sigma_{hv}^0 + \sigma_{vh}^0)/2) + ((\sigma_{vv}^0 + \sigma_{hh}^0)/2)}$ |                        |
| <i>Biomass Index</i><br>(Índice de Biomassa - BMI)   | $BMI = \frac{\sigma_{vv}^0 + \sigma_{hh}^0}{2}$   |                        |
| Razão de polarização paralela (HH/VV)  | $R_{co-pol} = \frac{\sigma_{hh}^0}{\sigma_{vv}^0}$  | Boerner et al. (1998)  |
| Razão de polarização cruzada (HH/HV)   | $R_{cross-pol} = \frac{\sigma_{hh}^0}{\sigma_{hv}^0}$   |                        |
| Razão de polarização cruzada (VV/VH)   | $R_{cross-pol} = \frac{\sigma_{vv}^0}{\sigma_{vh}^0}$   |                        |

#### 4.4.3.4 - Decomposições polarimétricas

A técnica de decomposição de alvos tem por objetivo representar o mecanismo médio de espalhamento como uma soma de elementos independentes, associando, a cada um destes elementos, um mecanismo físico de espalhamento (Cloude e Pottier, 1996). De acordo com Cloude e Pottier (1996) e Hellmann (2001), os algoritmos de decomposição de alvos podem ser classificados em dois grandes grupos: decomposições do tipo coerentes e decomposições do tipo incoerentes.

As decomposições coerentes assumem a ocorrência de espalhadores determinísticos e, por conseguinte, que a onda retroespalhada é polarizada. Geralmente, esse tipo de decomposição de alvos utiliza a matriz de espalhamento  $S$  para representar os estados de polarização da onda retroespalhada. Por outro lado, as decomposições incoerentes

consideram *a priori* que o espalhamento não é do tipo determinístico, portanto, a onda retroespalhada é parcialmente polarizada. Nessa situação, utilizam-se as matrizes de covariância e de coerência para caracterizar a onda retroespalhada (LEE e POTTIER, 2009; RICHARDS, 2009).

Geralmente, nas aplicações de sensoriamento remoto, por tratarem de áreas e alvos distribuídos, a premissa da ocorrência de alvos determinísticos puros não é válida (RICHARDS, 2009). Desse modo, utilizam-se, com frequência, as matrizes de reflexão de potência já destacadas.

Com base nesse entendimento, neste artigo, utilizaram-se somente métodos incoerentes de decomposição de alvos, sendo empregados assim os seguintes algoritmos de decomposição de alvos: van Zyl (com três componentes) (VAN ZYL, 1989); Freeman-Durden (com três componentes) (FREEMAN e DURDEN, 1998); Yamaguchi (com quatro componentes) (YAMAGUCHI et al., 2005); e entropia (H)-anisotropia (A)-ângulo  $\alpha$  (CLOUDE e POTTIER, 1996).

As decomposições foram geradas diretamente das imagens SLC, nível de processamento 1.1, por meio do aplicativo SNAP 6.0, utilizando uma janela de tamanho  $5 \times 5$  *pixels*. Não houve nenhum tipo de filtragem nas matrizes de covariância e coerência utilizadas nas decomposições polarimétricas.

O Quadro 4.6 apresenta um resumo dos teoremas de decomposição empregados nesta pesquisa. A Figura 4.6, por sua vez, apresenta algumas composições coloridas (RGB) geradas a partir das componentes obtidas.

Quadro 4.6. Relação de teoremas de decomposição de alvos utilizados nesta pesquisa.

| Método/Teorema        | Componente  | Descrição  |
|-----------------------|---|--|
| Cloude-Pottier (1996) | H (Entropia)  | Seus valores indicam a quantidade de mecanismos de espalhamento atuantes.  |
|                       | A (Anisotropia)   | Indicador da importância relativa dos mecanismos de espalhamento secundários.  |
|                       | $\alpha$ (ângulo alfa)  | Indica o tipo de mecanismo de espalhamento dominante.  |
| Van Zyl (1989)        | <i>even number of reflections</i><br>(número par de reflexões)  | Não há uma relação física com o mecanismo de espalhamento. Número de reflexões, sendo sempre um número inteiro divisível por 2.                |
|                       | <i>odd number of reflections</i><br>(número ímpar de reflexões) | Não há uma relação física com o mecanismo de espalhamento. Número de reflexões, sendo sempre um número inteiro sem divisão exata por 2.        |
|                       | <i>difuse scattering</i><br>(espalhamento difuso)               | Descreve alvos com espalhamento difuso.  |
| Freeman-Durden (1998) | Espalhamento <i>double bounce</i><br>(dupla batida)             | Modela o espalhamento <i>double-bounce</i> utilizando duas superfícies ortogonais.   |
|                       | Espalhamento volumétrico  | Modela o espalhamento volumétrico utilizando uma nuvem de dispersores cilíndricos e finos com distribuição aleatória.                          |
|                       | Espalhamento superficial  | Modela o espalhamento superficial utilizando espalhadores do tipo <i>bragg</i> , que ocorrem em superfícies ligeiramente rugosas ou onduladas. |
| Yamaguchi (2005)      | Espalhamento <i>double bounce</i><br>(dupla batida)             | Modela o espalhamento <i>double-bounce</i> utilizando duas superfícies ortogonais.   |
|                       | Espalhamento volumétrico  | Modela o espalhamento volumétrico utilizando uma nuvem de dispersores cilíndricos e finos com distribuição aleatória.                          |
|                       | Espalhamento superficial  | Modela o espalhamento superficial utilizando espalhadores tipo <i>bragg</i> , que ocorrem em superfícies ligeiramente rugosas/onduladas.       |

(continua)

Quadro 4.6 - Conclusão

| Método/Teorema | Componente               | Descrição   |
|----------------|--------------------------|---|
|                | Espalhamento tipo hélice | Modela um refletor do tipo hélice, que gera polarizações circulares para a esquerda e para a direita. Mais adequado para descrição de alvos não-naturais. |

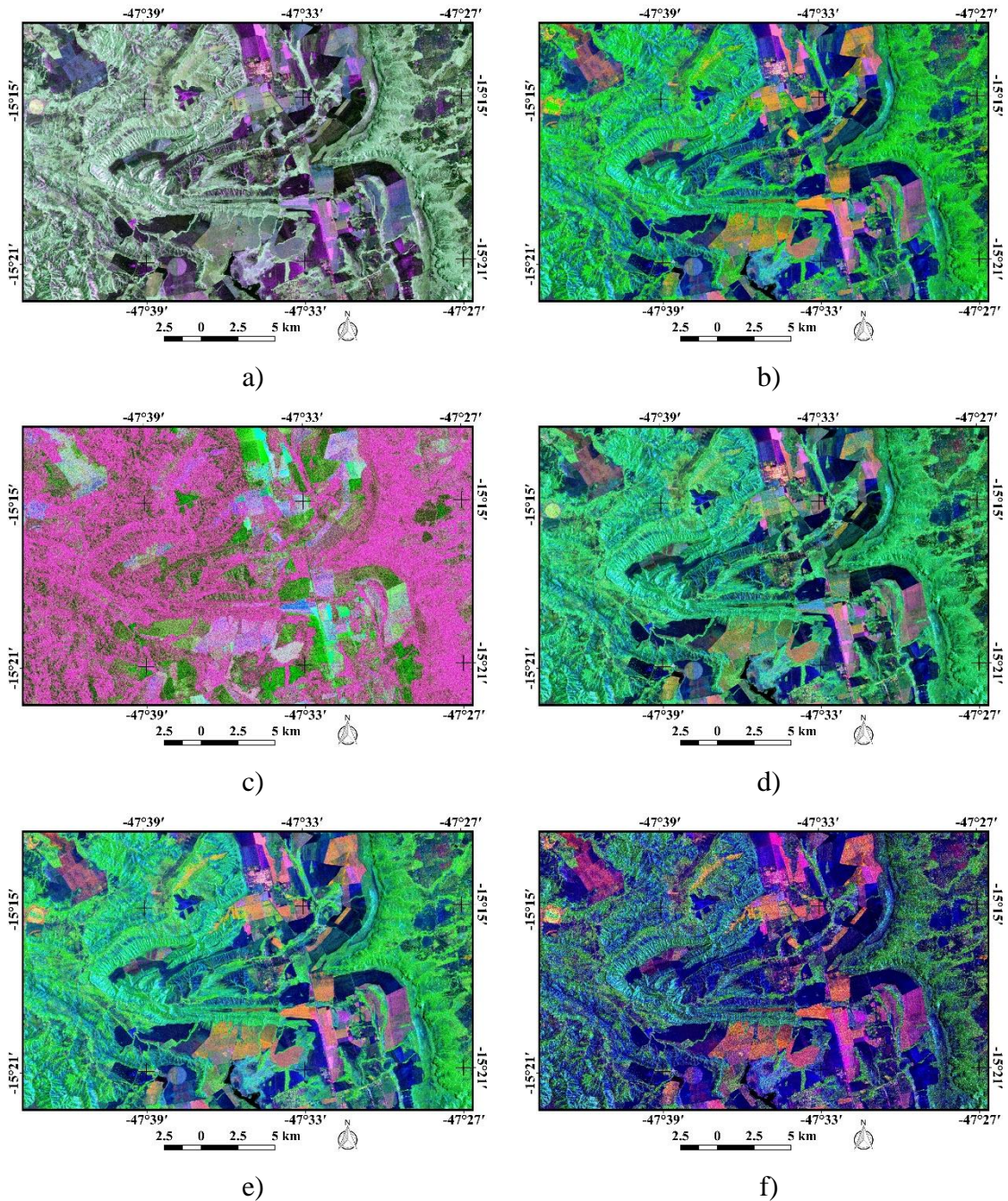


Figura 4.6. Composições coloridas RGB apresentando: a) imagens polarizadas em amplitude (HH/HV/VV); b) teorema de Freeman-Durden (*double/volume/surface*); c) teorema de Cloude-Pottier (H/A/ $\alpha$ ); d) teorema de van Zyl (*double/volume/surface*); e) Teorema de Yamaguchi sem a componente *Helix* (*double/volume/surface*) e, f) Teorema de Yamaguchi com a componente *Helix* (*double/helix/surface*). Datum horizontal: WGS84; Projeção: UTM 23S.



#### 4.4.3.5 - Ortorretificação

Neste trabalho, todos os pré-processamentos necessários foram realizados nas imagens comprimidas em alcance (*slant range*), dado que as informações das fases relativas tinham que estar preservadas para fins de decomposição polarimétrica, por exemplo. Esse mesmo procedimento também foi adotado para a classificação polarimétrica no aplicativo PCI Geomatica 2017. Assim, para fins da mineração de dados, as imagens polarizadas em  $\sigma^0$ , os parâmetros polarimétricos incoerentes e as decomposições polarimétricas foram posteriormente ortorretificadas utilizando o modelo *range doppler* específico para sensores SAR disponível no aplicativo SNAP 6.0.

O modelo *range doppler* utiliza informações do arquivo de metadados das imagens para se obter o vetor de estado da plataforma e para identificar os parâmetros de conversão de distância inclinada para a posição no terreno. Desse modo, esse modelo realiza a ortorretificação sem a necessidade de pontos de controle no terreno (PCTs). O algoritmo de ortorretificação do SNAP 6.0 realiza o *download* automático de modelos digitais de elevação (MDE) para a área da imagem (ESA, 2018). Nesta pesquisa, foram utilizados MDEs do projeto *Shuttle Radar Topography Mission* (Missão Topográfica do Radar Shuttle - SRTM), com 30 m de resolução espacial.

Para ortorretificar a classificação realizada no aplicativo PCI Geomatica, foi utilizado o modelo rigoroso Toutin (PCI GEOMATICS, 2018), que utiliza técnicas de radargrametria e também é específico para sensores SAR. Diferente do modelo *range doppler* do SNAP 6.0, o modelo rigoroso Toutin necessita de PCTs para seu ajuste. Assim, realizou-se a coleta de 10 PCTs nas ortoimagens obtidas utilizando o SNAP 6.0. Também foi empregado MDE SRTM de 30 m de resolução espacial. O modelo ajustado apresentou um erro médio quadrático (EMQ) de 3,94 m na resultante. As Figuras Figura 4.7 e Figura 4.8 apresentam, respectivamente, uma parte da cena ainda comprimida em alcance e, posteriormente, ortorretificada.

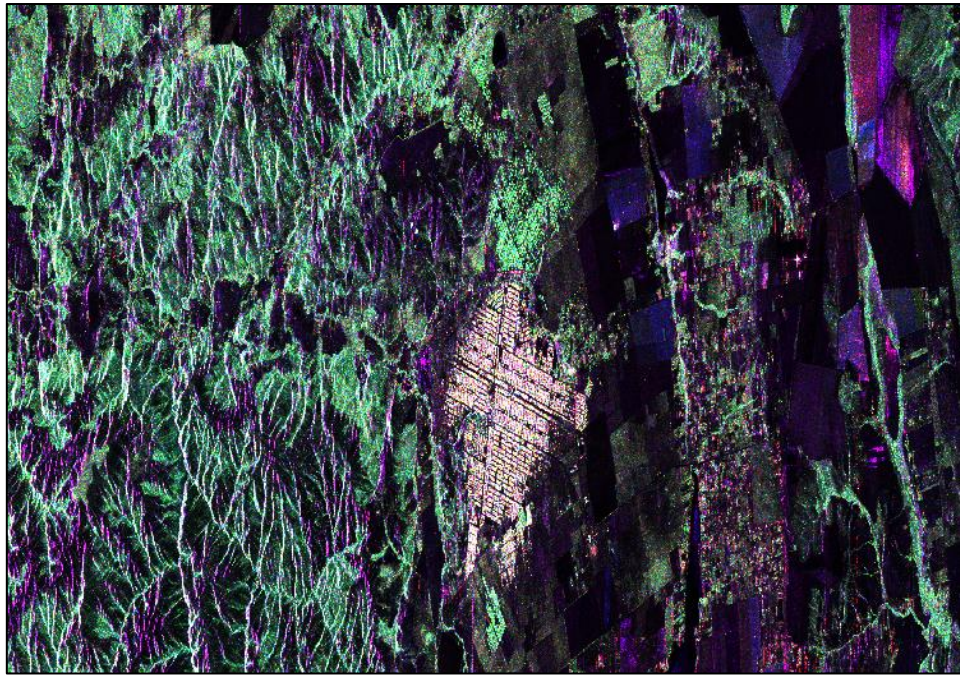


Figura 4.7. Parte da cena ALOS-2/PALSAR-2 SLC 1.1 original e comprimida em alcance.

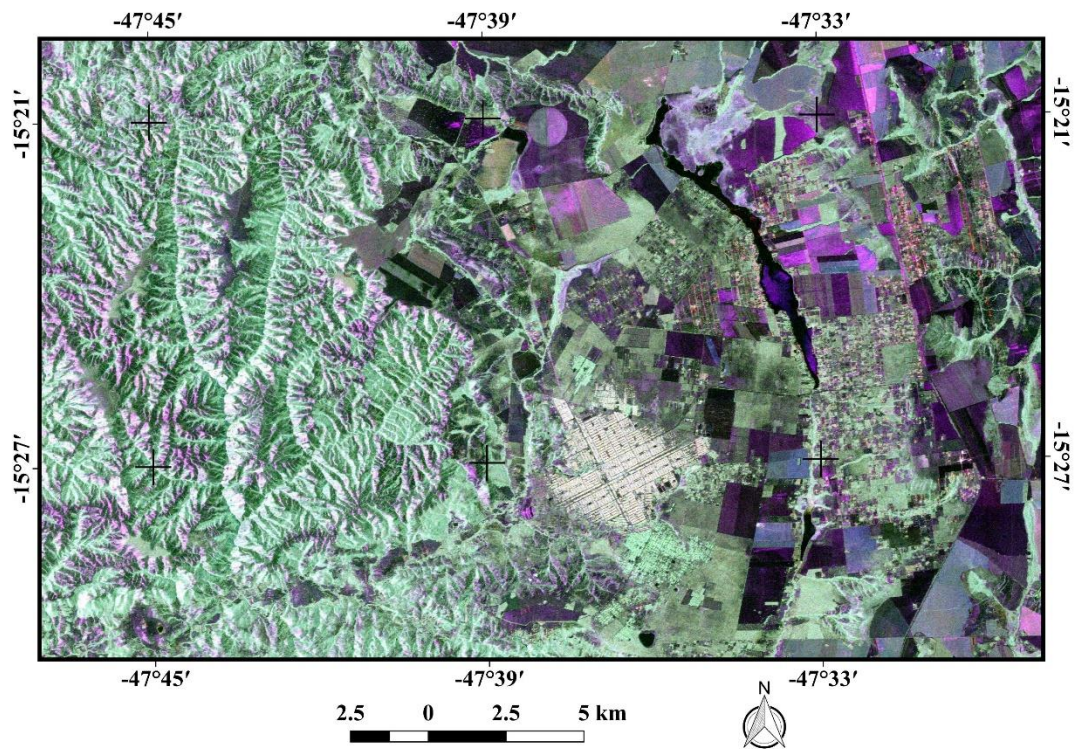


Figura 4.8. Parte da cena ALOS-2/PALSAR-2 SLC 1.1 ortoretificada, usando o modelo *range doppler* do SNAP 6.0. *Datum* horizontal: WGS84; Projeção: UTM 23S.

#### 4.4.3.6 - Segmentação e extração de atributos

Para a segmentação das imagens do satélite ALOS-2/PALSAR-2 (polarizações HH, HV, VH e VV em amplitude e calibradas em  $\sigma^0$ ), utilizou-se o algoritmo multiresolução por crescimento de regiões (TRIMBLE, 2011), o qual realiza um procedimento de otimização heurística, que minimiza a heterogeneidade média dos segmentos de uma imagem em dada resolução. As principais variáveis empregadas foram o fator de escala e a composição de homogeneidade. Essa última variável é dividida em cor e forma. A forma, por sua vez, está subdividida em compacidade e suavidade. A escala define o tamanho dos segmentos de uma imagem em cada nível de resolução, e a composição de homogeneidade testa a igualdade entre segmentos (BAATZ e SCHÄPE, 2000; BENZ et al., 2004; MEINEL e NEUBERT 2004).

Nesta pesquisa, apenas um nível de segmentação foi gerado, e os parâmetros do segmentador foram definidos de forma empírica, por meio de tentativa e erro (Quadro 4.7). Assim, selecionou-se o parâmetro de escala 50, e foram atribuídos pesos aos seguintes critérios de homogeneidade: forma, cor, suavidade e compacidade. A Figura 4.9 apresenta um exemplo da segmentação obtida em uma região situada na porção central da área de estudo.

Após a segmentação, foram extraídos os atributos dos segmentos dentro do próprio aplicativo *eCognition* 8.7. Dentre as diversas categorias de métricas de atributos disponíveis, foram utilizadas as seguintes categorias: *layer values* (valores da camada) (média, desvio-padrão e assimetria) e *pixel-based* (baseado em *pixel*) (valores mínimo e máximo do *pixel*) (Quadro 4.8). Desse modo, para cada uma das 25 imagens (oito parâmetros polarimétricos, 13 componentes de decomposições e quatro polarizações) disponíveis no projeto, foram extraídos os atributos das duas categorias acima citadas. Esse processo culminou em um conjunto de 125 *layers* de segmentos (cinco atributos vezes 25 imagens) (Apêndice A) com diferentes atributos que, posteriormente, foram exportados para classificação no aplicativo WEKA 3.8.

Quadro 4.7. Imagens e parâmetros utilizados na segmentação.

| Imagens e parâmetros utilizados na segmentação |   |
|--|---|
| Imagens  | Polarizações HH, HV, VH e VV ( $\sigma_0$ ) |
| Escala   | 50  |
| Forma  | 0,1   |
| Cor  | 0,9   |
| Suavidade                                      | 0,5   |
| Compacidade                                    | 0,5   |

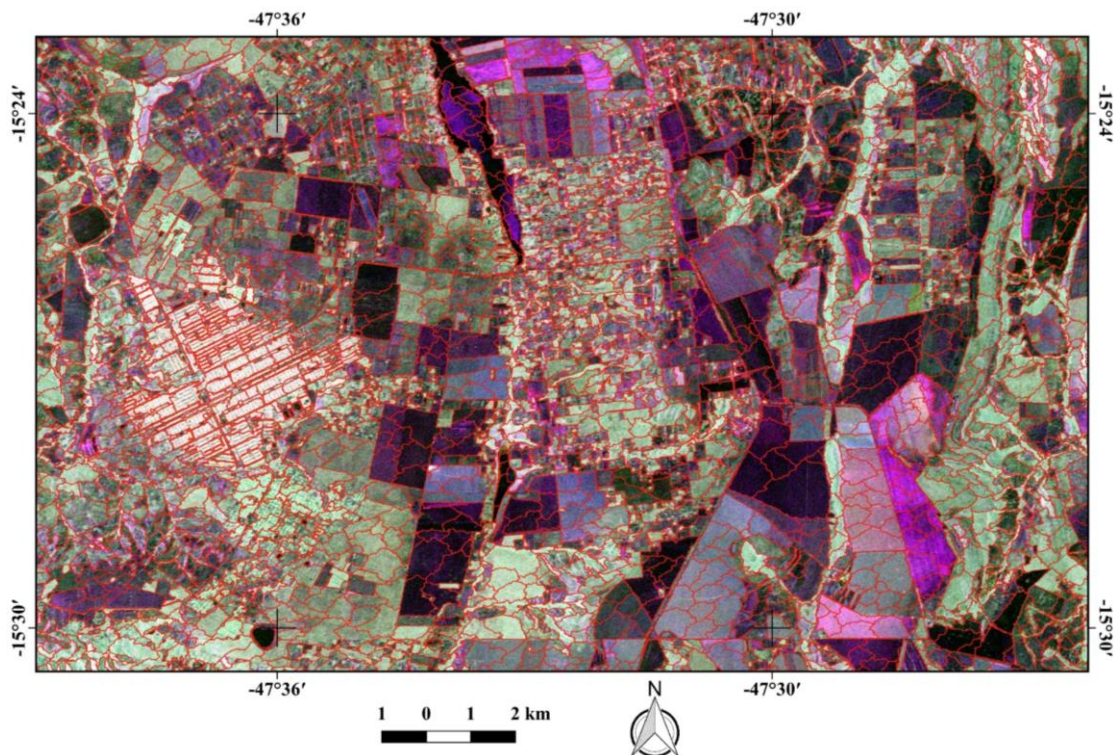


Figura 4.9. Exemplo da segmentação obtida em uma região situada na porção central da área de estudo. Datum horizontal: WGS84; Projeção: UTM 23S.

Quadro 4.8. Categorias e correspondentes tipos de atributos que foram considerados neste estudo.

| Categoria  | Atributos  |
|--|--|
| <i>Layer values</i><br>(valores da camada)       | <i>Mean</i> (média), <i>standard deviation</i> (desvio padrão) e <i>skewness</i> (assimetria).                           |
| <i>Pixel-based</i><br>(baseado em <i>pixel</i> ) | <i>minimum pixel value</i> (mínimo valor do <i>pixel</i> ) e <i>maximum pixel value</i> (máximo valor do <i>pixel</i> ). |

#### 4.4.3.7 - Classificações

##### Classificadores do aplicativo WEKA

Com relação aos algoritmos de aprendizado de máquina, as etapas de treinamento, classificação e validação foram realizadas dentro do próprio ambiente do aplicativo WEKA 3.8. Conforme já destacado, foram utilizados os algoritmos de classificação NB, AD J48, RF, MLP e SVM, os quais são descritos em detalhes por Witten e Frank (2005) e Han et al. (2012).

Primeiramente, com base na legenda e nas chaves de interpretação visual, realizou-se a coleta de segmentos amostrais, tanto para o treinamento dos classificadores quanto para a validação das classificações.

Foram coletadas 200 amostras de treinamento para cada classe, com exceção das classes Corpo d'água continental, Campo Limpo e Silvicultura, pois estas não dispunham de muitos segmentos (para cada uma destas foram coletadas 25 amostras de treinamento). A Figura 4.10 apresenta um gráfico *box-plot* gerado com o conjunto amostral total de treinamento contendo o atributo média de  $\sigma^0$  (dB) na polarização HH. Apesar da ocorrência de *outliers* e valores extremos, verificou-se adequada separação entre as classes da legenda com exceção das confusões entre as classes “Formações Florestais” e “Silvicultura”, “Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo” e “Culturas temporárias”, além de “Pastagens” e “Campo Limpo”.

Também foram coletados 959 segmentos independentes para fins de validação, utilizando, como orientação, os 1.000 pontos aleatórios e sem estratificação sorteados na etapa de campo (Seção 4.4.3.1). Dos 1.000 pontos, 41 tiveram que ser desprezados pois estavam localizados em feições de sombra, *layover* ou *foreshortening*. Deste modo, foram gerados sete arquivos em formato *shapefile*, cinco para treinamento dos classificadores (estratificados em 5, 25, 50, 100 e 200 amostras por classe), um arquivo para validação das classificações (com 959 segmentos) e outro para a classificação contendo todos os 39.254 segmentos gerados na etapa de segmentação. Portanto, cada algoritmo de classificação foi treinado cinco vezes e foram realizadas as respectivas validações com o mesmo conjunto de 959 segmentos.

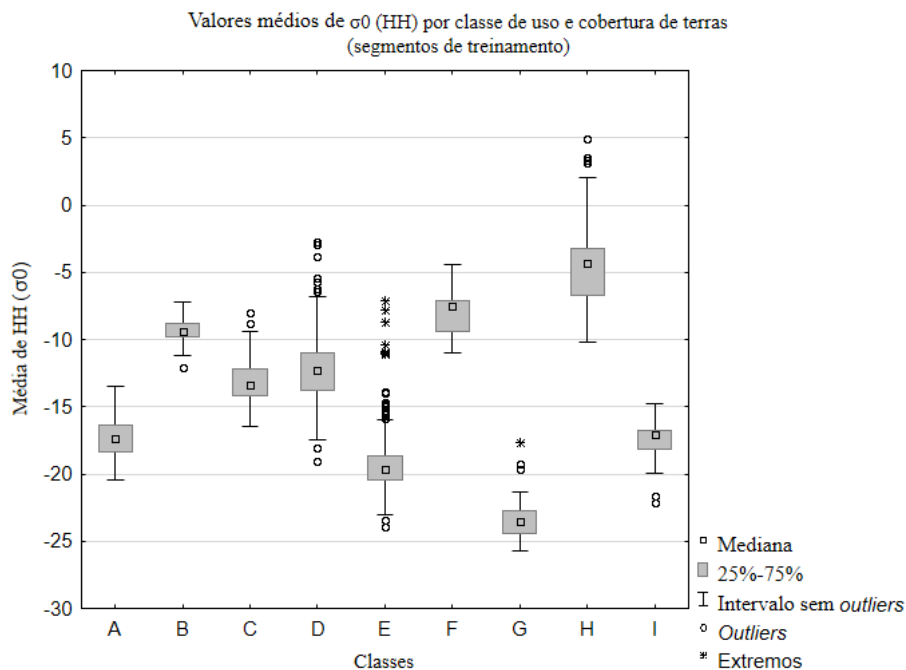


Figura 4.10. Gráfico *box-plot* gerado com o conjunto amostral total de treinamento contendo o atributo média de  $\sigma^0$  (dB) na polarização HH. A = Pastagens; B= Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

Posteriormente, com o intuito de se verificar a possibilidade de incremento na acurácia dos classificadores, um sexto conjunto de treinamento com 200 segmentos foi definido com base no procedimento de seleção de atributos *InfoGain* (WITTEN e FRANK, 2005). Assim, atributos com valor de ganho inferiores a 1,0 foram descartados e o novo conjunto de treinamento se limitou a 50 atributos (Apêndice D).

O Quadro 4.9 descreve os tipos de conjuntos de segmentos utilizados nesta pesquisa. A Figura 4.11, por sua vez, apresenta um exemplo da distribuição das amostras de treinamento e de validação.

Quadro 4.9. Conjuntos de segmentos utilizados nos classificadores do *software* WEKA 3.8.

| Conjuntos de segmentos utilizados | Quantidade | Descrição   |
|-----------------------------------|------------|---|
| Totalidade                        | 39.254     | Totalidade de segmentos obtidos para a cena ALOS-2/PALSAR-2. Utilizados nas classificações finais no WEKA.        |
| Amostras de treinamento           | 1.275      | Amostras utilizadas nos treinamentos de algoritmos de classificação do WEKA. Os treinamentos foram divididos em 5 |

Quadro 4.9 - Conclusão

| Conjuntos de segmentos utilizados | Quantidade | Descrição  |
|-----------------------------------|------------|--|
|                                   |            | conjuntos amostrais: 5, 25, 50, 100 e 200 segmentos por classe da legenda.   |
| Amostras de validação             | 959        | Amostras utilizadas na validação das classificações. Foram sorteadas 1.000 amostras para a cena ALOS-2/PALSAR-2. Todavia 41 amostras foram desconsideradas por coincidirem com áreas de <i>layover</i> , sombra ou <i>foreshortening</i> . |

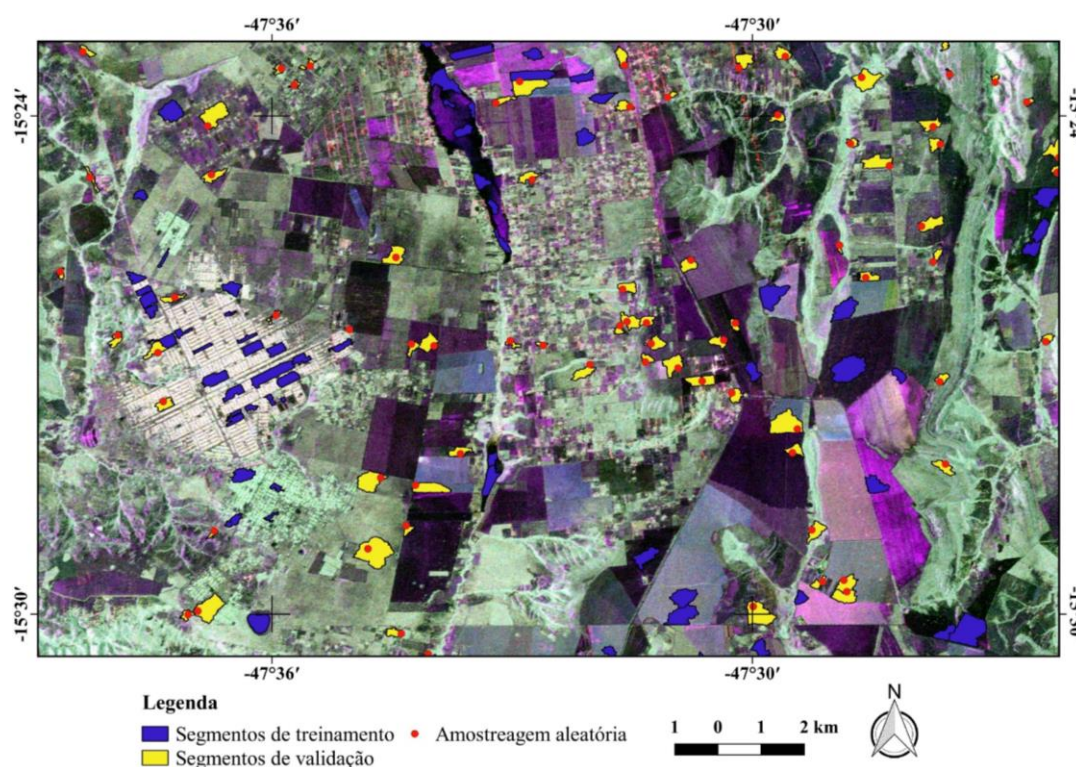


Figura 4.11. Exemplo da distribuição das amostras de treinamento e de validação. *Datum* horizontal: WGS84; Projeção: UTM 23S.

### Classificador supervisionado polarimétrico Wishart

Nesta pesquisa, utilizou-se o classificador PW disponível no aplicativo PCI Geomatica 2017. A partir da imagem polarimétrica do sensor ALOS-2/PALSAR-2 no nível de processamento SLC 1.1, gerou-se a matriz de covariância, sendo essa matriz posteriormente calibrada em  $\sigma^0$  (dB) e submetida ao filtro *Refined Lee*, de janela 9 x 9.

Na sequência, foram selecionadas regiões de interesse (ROIs) contendo *pixels* de treinamento homogêneos, em quantidade adequada e em áreas diferentes daquelas dos

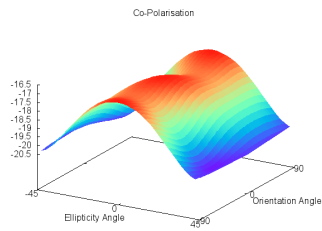
959 segmentos para validação. O Quadro 4.10 apresenta os conjuntos de *pixels* dos ROIs utilizados na classificação e validação.

Quadro 4.10. Conjuntos de *pixels* utilizados na classificação e validação.

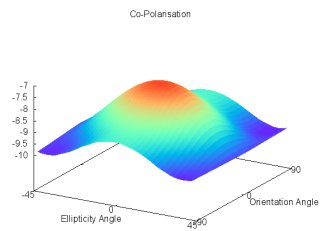
| Conjuntos de <i>pixels</i> utilizados | Quantidade  | Descrição  |
|---------------------------------------|-------------|--|
| Totalidade                            | 221.161.559 | Totalidade de <i>pixels</i> da cena ALOS-2/PALSAR-2 ( <i>slant range</i> ).  |
| Amostras de treinamento               | 241.621     | Amostras utilizadas no treinamento do algoritmo de classificação Polarimétrico Wishart. Foram coletados ROIs contendo <i>pixels</i> independentes daqueles do conjunto de validação. Imagens em <i>slant range</i> .   |
| Amostras de validação                 | 2.106.129   | Amostras utilizadas na validação da classificação do algoritmo Polarimétrico Wishart (amostras distintas e independentes das do conjunto de treinamento). Consistem em <i>pixels</i> contidos nas áreas dos 959 segmentos de validação apontados no Quadro 4.9. Imagens ortoretificadas. |

Na Figura 4.12 estão dipostas as assinaturas polarimétricas de cada classe, obtidas com as amostras de treinamento. Essas assinaturas polarimétricas orientaram de forma qualitativa o treinamento do classificador PW. Ao final da classificação, aplicou-se um filtro de moda 5 x 5 no *raster* temático visando à homogeneização dos padrões de uso e cobertura da terra e eliminação de *pixels* isolados em divergência com o padrão dominante do entorno.

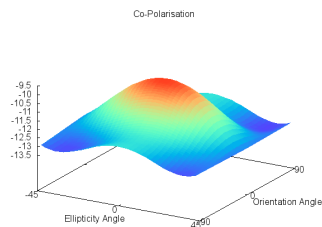




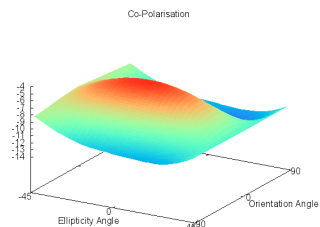
a) Pastagens



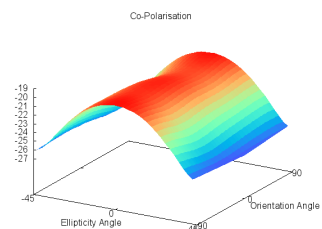
b) Formações florestais



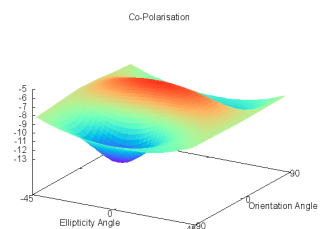
c) Cerrado Típico antropizado/Cerrado Ralo/Campo sujo



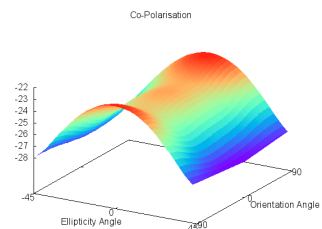
d) Culturas temporárias



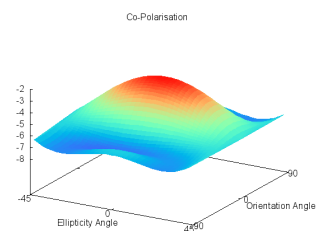
e) Solo exposto



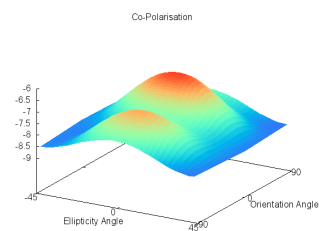
f) Silvicultura



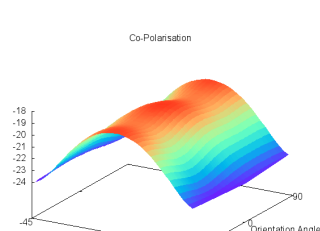
g) Corpo d'água continental



h) Áreas com edificações (com reflexão de canto)



i) Áreas com edificações (com espalhamento volumétrico)



j) Campo limpo

Figura 4.12. Assinaturas polarimétricas (co-polarização) de cada classe, obtidas com as amostras de treinamento: a) Pastagens; b) Formações florestais; c) Cerrado Típico antropizado/Cerrado Ralo/Campo Sujo; d) Culturas temporárias; e) Solo exposto; f) Silvicultura; g) Corpo d'água continental; h) Áreas com edificações (com reflexão de canto); i) Áreas com edificações (com espalhamento volumétrico); e j) Campo Limpo.

#### 4.4.3.8 - Validações

As validações foram realizadas com base nas matrizes de erros geradas para cada classificação. A matriz de erros é um recurso adequado para se avaliar classificações, pois permite a extração de uma série de métricas de acurácia. Nessa matriz, as colunas usualmente representam os dados de referência, e as linhas, a classificação obtida de dados de sensoriamento remoto. A partir da matriz de erros, podem ser avaliadas as acurácias individuais, de cada categoria, por meio dos erros de inclusão (erros de comissão) e de exclusão (erros de omissão) (CONGALTON e GREEN, 2009). O Quadro 4.11 apresenta e descreve as métricas de validação temática utilizadas nesta pesquisa. Também foram empregados testes de hipóteses com base na distribuição normal padrão para comparação dos desempenhos dos diferentes classificadores.

As classificações realizadas por meio de mineração de dados e algoritmos de aprendizado de máquina foram validadas com base em segmentos. Por outro lado, a classificação obtida utilizando o classificador PW foi validada com base nos *pixels* presentes nos 959 segmentos do conjunto amostral de validação.

Quadro 4.11. Métricas de validação temática utilizadas nesta pesquisa.

| Métricas de validação utilizadas | Descrição  | Fonte                    |
|----------------------------------|--|--------------------------|
| Acurácia Global                  | Consiste na razão simples entre o total de acertos da matriz (diagonal principal) e o total de amostras utilizadas na matriz.  | Congalton e Green (2009) |
| Índice Kappa                     | Índice de concordância entre classificações obtido do estimador KHAT ( $\hat{K}$ ) que se baseia na concordância real (indicada pela diagonal principal da matriz) e na chance de concordância (totais marginais em linhas e colunas). |                          |
| Acurácia do produtor             | Total de acertos por classe pelo total de amostras referentes à classe analisada e na perspectiva do produtor.   |                          |
| Acurácia do usuário              | Total de acertos por classe pelo total   |                          |

(continua)

Quadro 4.11 - Conclusão

| Métricas de validação utilizadas | Descrição   | Fonte                    |
|----------------------------------|---|--------------------------|
|                                  | de amostras referentes à classe analisada e na perspectiva do usuário.            | Congalton e Green (2009) |
| Kappa condicional do produtor    | Índice que emprega o estimador $\hat{K}$ por classe e na perspectiva do produtor. |                          |
| Kappa condicional do usuário     | Índice que emprega o estimador $\hat{K}$ por classe e na perspectiva do usuário.  |                          |

#### 4.5 - Resultados e Discussão

As Tabelas 4.1 e 4.2 apresentam respectivamente os valores de acurácia global e de índice Kappa das classificações (legenda com nove classes) realizadas com diferentes conjuntos de treinamento e distintos classificadores utilizados nesta pesquisa. As matrizes de erros empregadas referentes aos resultados das Tabelas 4.1 e 4.2 estão dispostas no Apêndice E deste documento. Nas Figuras 4.13 e 4.14 são apresentados respectivamente os gráficos com as variações de resultados para acurácia global e índice Kappa também de acordo com diferentes conjuntos de treinamento e distintos classificadores do aplicativo já mencionado.

O item “200 INFOGAIN”, nas Tabelas 4.1 e 4.2 e Figuras 4.13 e 4.14, representa a amostra de treinamento com 200 segmentos por classe e com quantidade reduzida de atributos (50 atributos) selecionada por meio do algoritmo *Infogain*. Adicionalmente, também estão apresentados nas tabelas e gráficos, os valores de acurácia global e índice Kappa do classificador PW simplesmente para efeitos de comparação, haja vista que esse classificador foi treinado uma única vez e com um único conjunto de treinamento, conforme detalhes descritos no Quadro 4.10.

Observa-se na Figura 4.14 que, para todos os classificadores baseados em aprendizado de máquina, o índice de concordância Kappa aumentou gradativamente de acordo com o aumento no tamanho da amostra de treinamento. O valor máximo atingido foi 0,67 para o algoritmo SVM. Verificou-se também que os classificadores MLP, RF e SVM já partem de índices Kappa elevados (0,51), mesmo com poucas amostras de treinamento. Shiraishi et al. (2014) também verificaram esse comportamento em experimento com diferentes conjuntos amostrais de treinamento.

Belgiu e Dragut (2016) destacaram que o RF pode apresentar bom desempenho quando a amostra é pequena, mas a dimensão do espaço de atributos é grande.

Outro fato importante, observado na Figura 4.14, refere-se ao desempenho do classificador NB. Esse classificador rivaliza em desempenho com a AD J48, sendo superior no início com poucas amostras de treinamento e caindo em desempenho após o treinamento com 200 amostras. O desempenho do NB recebe destaque na literatura pertinente (WITTEN e FRANK, 2005; HAN et al., 2012), sendo recomendado até sua utilização preliminarmente a métodos mais sofisticados.

Este resultado do NB merece destaque, dado que se trata de um classificador de simples uso e custo computacional baixo. Dependendo da complexidade do uso e cobertura da terra da área a ser classificada e do tamanho desta área, o emprego do classificador NB passa a ser vantajoso para fins de projeto de mapeamento. Em áreas extensas e uma legenda com poucas classes e padrões simples e homogêneos de uso e cobertura da terra, o classificador NB pode ser suficiente para realização de um inventário preliminar.

Com relação aos classificadores que utilizam árvores de decisão (AD J48 e RF), observou-se que o emprego massivo de diversas ADs (caso do RF) possibilita índices de concordância Kappa mais elevados, comparativamente ao AD J48, em todos os cenários de treinamento. Esse resultado do RF corrobora o que se verifica na literatura científica pertinente (BELGIU e DRAGUT, 2016). O custo computacional não é drasticamente elevado, o que torna viável o uso do RF em detrimento do AD J48, dada a grande diferença de desempenho entre estes classificadores. Ainda na Figura 4.14, verifica-se que RF apresentou desempenho semelhante aos classificadores mais complexos, baseados em funções, ou seja, MLP e SVM. Isso ocorreu em todas as etapas de treinamento e está coerente com resultados verificados em outros trabalhos (SHIRAISHI et al., 2014).

As funções MLP e SVM apresentaram bom desempenho em conformidade com o que se verifica na literatura científica pertinente (SHIRAISHI et al., 2014). Cabe destacar que os classificadores MLP e SVM podem ser muito semelhantes em termos de implementação e desempenho, dependendo do tipo de função utilizada. No presente trabalho, os classificadores utilizam funções diferentes, sendo funções *kernel* sigmoidais para o MLP e funções *kernel* gaussianas de base radial para o SVM (WITTEN e FRANK, 2005; HAN et al., 2012).

A aplicação da técnica de seleção de atributos *Infogain* não possibilitou nenhum aumento do índice de concordância Kappa. Destaca-se que nesta pesquisa, a utilização do algoritmo *Infogain* se restringiu ao conjunto amostral com 200 segmentos por classe da legenda. Todavia, essa técnica pode ser útil para análise dos atributos verificando quais são mais importantes, como, por exemplo, qual tipo de teorema de decomposição polarimétrica contribui mais para dada tarefa de classificação.

No que tange ao classificador PW, constatou-se um desempenho muito reduzido se comparado aos resultados dos demais classificadores e, também, aos observados na literatura científica pertinente para este tipo de classificador em particular (LEE e POTTIER, 2009; SINGH et al., 2014) (Figura 4.14). Acredita-se que o tamanho e a complexidade da área de estudo afetaram o desempenho do classificador PW, haja vista que esse dispunha de menos recursos e atributos para realizar sua tarefa. Mais adiante, os resultados do classificador PW serão analisados detalhadamente considerando o seu desempenho por classes da legenda (nove classes) e também uma legenda agregada, com cinco classes.

Tabela 4.1. Valores de acurácia global por conjunto de treinamento (legenda com nove classes).

|                 |                               | Número de amostras de treinamento<br>(por classe da legenda) |        |        |        |        |                 |
|-----------------|-------------------------------|--|--------|--------|--------|--------|-----------------|
|                 |                               | 5  | 25     | 50     | 100    | 200    | 200<br>INFOGAIN |
| Acurácia Global | <i>Naive Bayes</i>            | 54,43%   | 61,84% | 64,23% | 65,80% | 63,50% | 63,40%          |
|                 | <b>J48</b>                    | 45,88%   | 57,24% | 65,48% | 65,38% | 67,57% | 69,13%          |
|                 | <i>Random Forest</i>          | 61,52%   | 66,84% | 69,97% | 72,88% | 73,20% | 73,20%          |
|                 | <i>Multilayer Perceptron</i>  | 60,10%   | 65,90% | 68,51% | 70,17% | 72,99% | 72,58%          |
|                 | <i>Support Vector Machine</i> | 64,03%   | 68,93% | 71,32% | 74,35% | 74,18% | 74,56%          |
|                 | Polarimétrico Wishart*        | 48,00%   |        |        |        |        |                 |

Tabela 4.2. Índice Kappa por conjunto de treinamento (legenda com nove classes).

|                         |                               | Número de amostras de treinamento<br>(por classe da legenda) |      |      |      |      |                 |
|-------------------------|-------------------------------|--|------|------|------|------|-----------------|
|                         |                               | 5  | 25   | 50   | 100  | 200  | 200<br>INFOGAIN |
| Kappa por Classificador | <i>Naive Bayes</i>            | 0,45   | 0,53 | 0,56 | 0,58 | 0,55 | 0,55            |
|                         | <b>J48</b>                    | 0,36   | 0,47 | 0,57 | 0,56 | 0,59 | 0,61            |
|                         | <i>Random Forest</i>          | 0,53   | 0,59 | 0,62 | 0,65 | 0,66 | 0,65            |
|                         | <i>Multilayer Perceptron</i>  | 0,51   | 0,57 | 0,60 | 0,62 | 0,66 | 0,65            |
|                         | <i>Support Vector Machine</i> | 0,55   | 0,61 | 0,64 | 0,67 | 0,68 | 0,67            |
|                         | <b>Polarimétrico Wishart</b>  | 0,35   |      |      |      |      |                 |

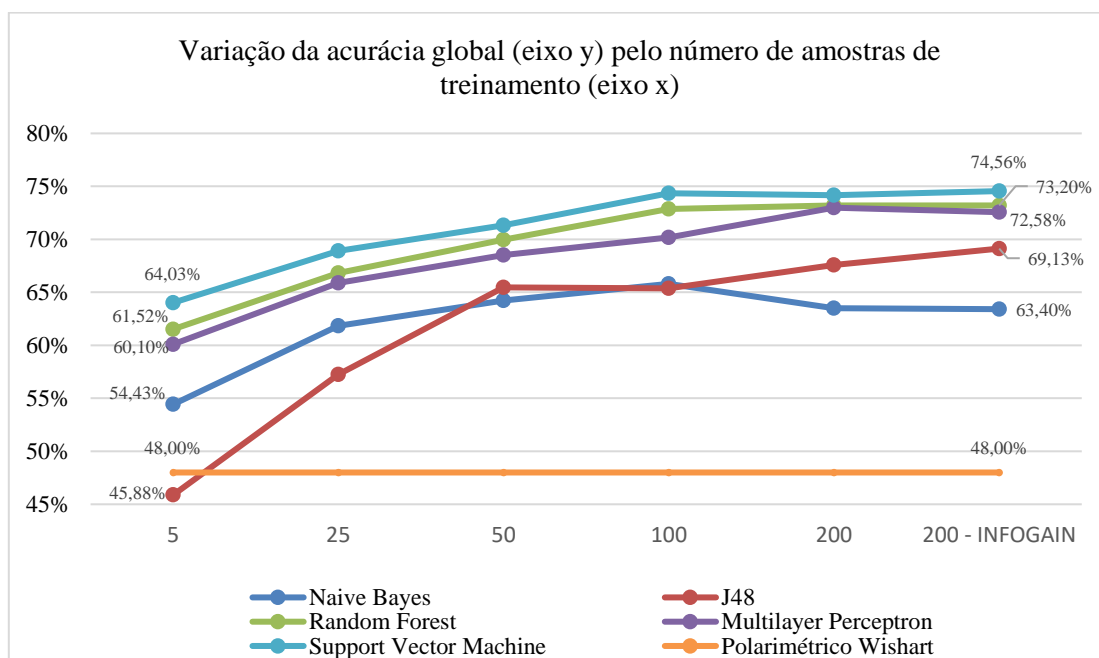


Figura 4.13. Variação dos valores de acurácia global por conjunto de treinamento (legenda com nove classes).

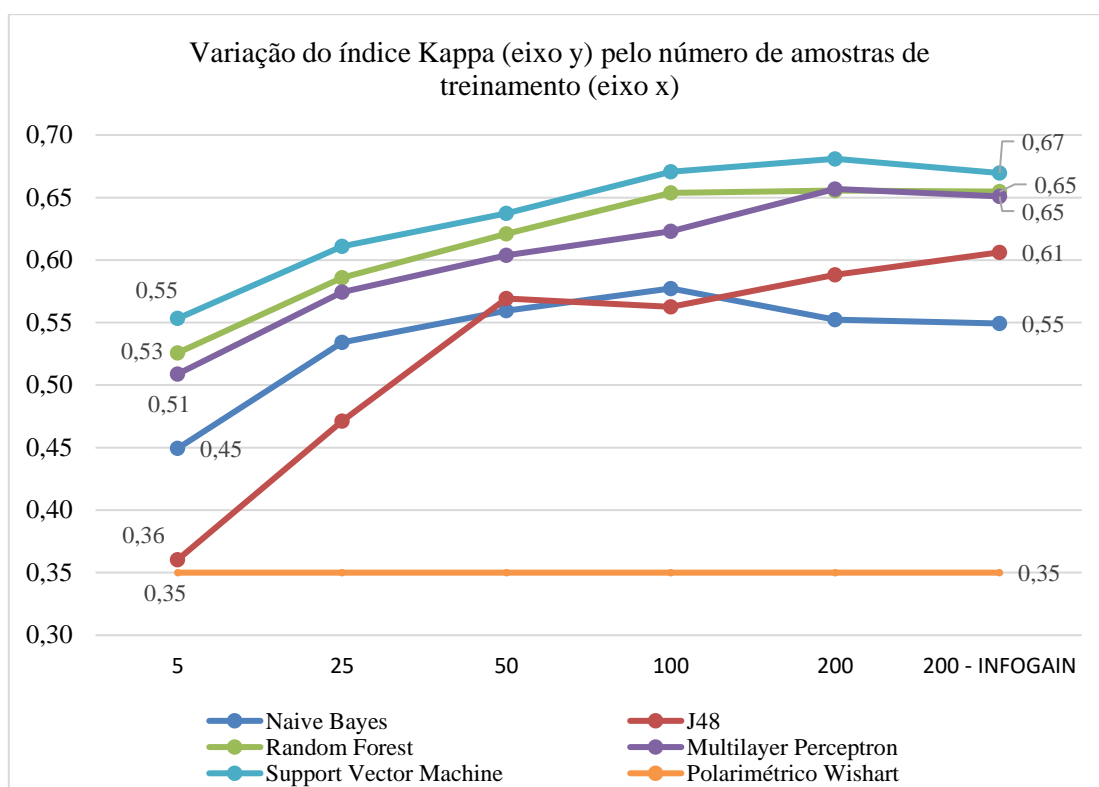


Figura 4.14. Variação do índice Kappa por conjunto de treinamento (legenda com nove classes).

Com relação ainda à legenda com nove classes, procedeu-se à análise detalhada das exatidões por usuário e produtor (Tabela 4.3; Figura 4.15) e também dos índices Kappa condicionais por usuário e por produtor (Tabela 4.4; Figura 4.16). Observa-se nas Tabelas 4.3 e 4.4 que a classe “Corpo d’água continental” não apresentou resultados. Isto se deve pela falta de amostras nesta classe que apresenta pouca abrangência em área na cena ALOS-2/PALSAR-2 utilizada (a amostragem foi realizada sem estratificação).

Considerando a perspectiva do usuário do mapa, observa-se que o classificador PW apresentou altos índices de concordância Kappa para as classes “Formações florestais” (KCU = 0,82), “Culturas temporárias” (KCU = 0,68) e “Solo exposto/palhada” (KCU = 0,74) (Tabela 4.4). As demais classes para o PW apresentaram índices muito baixos. Esses resultados indicam que o PW pode ser mais adequado para a identificação e separação de áreas com formações florestais e não florestais de modo amplo. Isto também pode ter sido condicionado pelas características da banda L e seu comprimento de onda que interage no interior do dossel e representa melhor a estrutura da vegetação (URBAZAEV et al., 2015; ODIPO et al., 2016; BOUVET et al., 2018).

Analisando a matriz de erros do PW (Apêndice E), também da perspectiva do usuário, verifica-se como o classificador gerou muitos erros de comissão. Destaque deve ser dado às áreas da classe “Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo” que foi incluída erroneamente e em grandes proporções tanto na classe “Formações florestais” quanto na classe “Culturas temporárias”.

Uma alternativa a ser testada usando o PW em área de Cerrado poderia ser o emprego de dados multifrequência, inserindo dados na banda C (aliado à banda L), por exemplo, para tentar melhorar os índices de acurácia. Lee e Pottier (2009) empregaram com sucesso o classificador PW supervisionado e dados multifrequência para a discriminação de diferentes tipos de blocos de gelo em área do mar de Beaufort, ao norte do Alasca. Outra alternativa seria o emprego de dados multitemporais, mesmo que em uma única frequência.

Considerando os classificadores utilizados na mineração de dados, verifica-se nas Tabelas 4.3 e 4.4 que há desempenhos semelhantes, por classe, entre os algoritmos RF, MLP e SVM. Isso ocorre tanto na perspectiva do usuário quanto na do produtor. Do mesmo modo, verificam-se semelhanças nos desempenhos dos algoritmos NB e AD J48, sendo este último um pouco superior ao NB na maioria das classes.

Destaca-se aqui o desempenho do classificador NB para as classes “Silvicultura” e “Campo Limpo”, nas quais o Kappa condicional por usuário chegou a atingir, respectivamente, 0,65 e 0,70, resultados esses superiores a todos os demais classificadores nas referidas classes e também na perspectiva do usuário.

As Figuras 4.15 e 4.16 destacam características relativas aos desempenhos dos classificadores. Por exemplo, observa-se que, na perspectiva do usuário, todos os classificadores de mineração de dados apresentaram os maiores índices de acerto na classe “Áreas com edificações”. Por outro lado, na perspectiva do produtor, a maioria dos classificadores de mineração de dados apresentaram maiores índices de acerto na classe “Formações florestais”. Ainda com relação a essas duas últimas figuras, para o classificador PW, observou-se que os maiores acertos foram na perspectiva do usuário e nas classes “Formações florestais”, “Culturas temporárias” e “Solo exposto/palhada”, conforme já observado nas Tabelas 4.3 e 4.4, destacado anteriormente.

Observando as matrizes de erro (Apêndice E) dos classificadores utilizados na mineração de dados, pode-se verificar que a classe “Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo” confundiu-se em grau elevado com as



classes “Pastagens” e “Formações florestais”. Esse tipo de confusão entre essas classes pode até ser esperado, dado que o Cerrado Ralo pode ter atributos e características semelhantes às pastagens nas imagens em banda L. Por outro lado, dependendo das condições de preservação ambiental do Cerrado Ralo/antropizado, confusão com formações arbóreas podem ocorrer devido ao espalhamento volumétrico predominante (URBAZAEV et al., 2015; ODIPO et al., 2016; BOUVET et al., 2018). Essas confusões mencionadas ocorreram em todos os algoritmos de aprendizado de máquina, tendo sido menores no classificador SVM.

Ainda com relação à mineração de dados, confusões acentuadas também ocorreram entre as classes “Pastagens” e “Solo exposto/palhada” (Apêndice E). Aqui também os erros podem ser atribuídos a atributos e características semelhantes entre as classes mencionadas nas imagens em banda L. Essas confusões mencionadas também ocorreram em todos os algoritmos de aprendizado de máquina, tendo sido menores no classificador SVM.

Na Figura 4.19, relativa ao treinamento com 200 amostras e nove classes, são apresentados os valores de acurácia global e índice Kappa obtidos. Nas Figuras 4.21 a 4.23, estão apresentadas as classificações referentes à legenda com nove classes. Na sequência, uma agregação de classes foi realizada gerando assim uma legenda com cinco classes. Basicamente, as classes de vegetação natural e silvicultura foram agrupadas na nova classe “Cobertura vegetal arbórea-arbustiva” e, as classes “Pastagens” e “Culturas temporárias” foram agrupadas na nova classe “Cobertura vegetal herbácea”. Além disto, as classes “Solo exposto” e “Campo Limpo” foram agrupadas na nova classe “Solo exposto”. Na Figura 4.20, estão dispostos os índices Kappa e as exatidões globais por classificador e referente a essa nova legenda. As Figuras 4.24 a 4.26 apresentam as classificações utilizando a legenda com cinco classes.

Na nova legenda, houve incremento nos índices de acurácia para todos os classificadores, como se pode depreender da comparação entre as Figuras 4.19 e 4.20. O maior salto foi verificado nos índices do classificador PW (índice Kappa e acurácia global), o que corrobora ainda mais a aptidão deste classificador para separar formações florestais de formações não-florestais. Nos algoritmos de aprendizado de máquina, houve saltos consideráveis nas exatidões globais, todavia, os índices Kappa tiveram pouco incremento.

Considerando os acertos por classe (Tabelas 4.5 e 4.6; Figuras 4.17 e 4.18), na legenda com cinco classes, pode-se verificar que o melhor desempenho do classificador PW fica restrito à identificação de cobertura arbórea, dado que seu melhor resultado foi um índice Kappa condicional do produtor de 0,67. Observou-se também que todos os algoritmos de aprendizado de máquina tiveram bom desempenho na identificação da classe “Áreas com edificações”, com destaque para a AD J48, que obteve alto grau de acerto para a referida classe tanto na perspectiva do usuário quanto na do produtor.

No que se refere à classe “Cobertura vegetal arbórea-arbustiva”, os classificadores RF, MLP e SVM apresentaram os melhores resultados em ambas as perspectivas do usuário e do produtor (Figura 4.18). A partir da Figura 4.18, verifica-se ainda que esses três classificadores mencionados obtiveram os melhores desempenhos para as classes “Solo exposto” e “Cobertura vegetal herbácea”. Novamente, destaque deve ser dado ao classificador NB, que apresentou moderados a altos índices de acerto (Kappa condicional do produtor) para as classes “Cobertura vegetal arbórea-arbustiva” e “Cobertura vegetal herbácea” (Tabela 4.6; Figura 4.18). O classificador NB também apresentou um bom desempenho na identificação da classe “Solo exposto”, na perspectiva do usuário (Tabela 4.6; Figura 4.18).

A análise pormenorizada das matrizes de erros referentes à legenda com cinco classes (Apêndice F) possibilita a identificação de padrões de confusão entre classes. De modo geral, em maior ou menor grau, todos os classificadores apresentam confusões entre “Cobertura vegetal herbácea” e “Solo exposto”, “Cobertura vegetal arbórea-arbustiva” e “Cobertura vegetal herbácea” e entre “Áreas com edificações” e “Cobertura vegetal arbórea-arbustiva”.

As confusões entre “Cobertura vegetal herbácea” e “Solo exposto” são de alguma forma previsíveis, haja vista que essas classes podem apresentar atributos muito similares nas imagens em banda L (predomínio de espalhamento superficial). No caso referente às classes “Cobertura vegetal arbórea-arbustiva” e “Cobertura vegetal herbácea”, as confusões residem no fato de que áreas de Cerrado Típico denso foram agrupadas na classe de vegetação arbórea e, desse modo, podem apresentar similitude com a classe de vegetação herbácea devido a características intrínsecas do Cerrado (dossel espaçado com presença também de vegetação graminosa) (RIBEIRO e WALTER, 2008). Por fim, as confusões verificadas entre “Áreas com edificações” e

“Cobertura vegetal arbórea-arbustiva” se explica pelo fato do espalhamento volumétrico ocasionado por vegetação arbórea em áreas urbanas.

Tabela 4.3. Acurácias do usuário e do produtor para cada classificador utilizado (legenda com nove classes).

|   | Polarimétrico Wishart |        | Naive Bayes |        | J48     |        | Random Forest |         | Multilayer Perceptron |        | Support Vector Machine |        |
|---|-----------------------|--------|-------------|--------|---------|--------|---------------|---------|-----------------------|--------|------------------------|--------|
|   | AU                    | AP     | AU          | AP     | AU      | AP     | AU            | AP      | AU                    | AP     | AU                     | AP     |
| <b>Pastagens</b>  | 50,55%                | 57,11% | 46,93%      | 74,31% | 53,51%  | 67,03% | 59,65%        | 72,73%  | 58,77%                | 78,36% | 59,65%                 | 74,32% |
| <b>Formações florestais</b>                                       | 87,76%                | 53,55% | 72,89%      | 97,64% | 76,41%  | 94,76% | 81,69%        | 93,55%  | 78,52%                | 96,54% | 83,45%                 | 95,18% |
| <b>Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo</b> | 34,53%                | 67,48% | 68,83%      | 68,27% | 77,33%  | 65,41% | 85,02%        | 68,40%  | 87,04%                | 69,81% | 86,64%                 | 70,16% |
| <b>Culturas temporárias</b>                                       | 71,64%                | 11,81% | 73,68%      | 49,41% | 78,95%  | 52,33% | 80,70%        | 54,76%  | 77,19%                | 77,19% | 77,19%                 | 74,58% |
| <b>Solo exposto/palhada</b>                                       | 77,53%                | 23,51% | 49,52%      | 71,23% | 50,48%  | 51,46% | 55,24%        | 62,37%  | 55,24%                | 59,18% | 62,86%                 | 55,93% |
| <b>Silvicultura</b>   | 9,29%                 | 10,41% | 66,67%      | 4,08%  | 33,33%  | 11,11% | 33,33%        | 100,00% | 33,33%                | 7,69%  | 33,33%                 | 25,00% |
| <b>Corpo d'água continental</b>                                   | 0,00%                 | -      | -           | 0,00%  | -       | 0,00%  | -             | -       | -                     | -      | -                      | -      |
| <b>Áreas com edificações</b>                                      | 24,27%                | 51,96% | 100,00%     | 25,00% | 100,00% | 92,86% | 100,00%       | 72,22%  | 100,00%               | 48,15% | 100,00%                | 72,22% |
| <b>Campo Limpo</b>  | 2,01%                 | 16,26% | 72,73%      | 17,20% | 27,27%  | 14,29% | 27,27%        | 28,57%  | 54,55%                | 22,22% | 45,45%                 | 43,48% |

Tabela 4.4. Valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com nove classes).

|   | Polarimétrico Wishart |      | Naive Bayes |      | J48  |      | Random Forest |      | Multilayer Perceptron |      | Support Vector Machine |      |
|---|-----------------------|------|-------------|------|------|------|---------------|------|-----------------------|------|------------------------|------|
|   | KCU                   | KCP  | KCU         | KCP  | KCU  | KCP  | KCU           | KCP  | KCU                   | KCP  | KCU                    | KCP  |
| <b>Pastagens</b>  | 0,37                  | 0,44 | 0,38        | 0,66 | 0,43 | 0,57 | 0,50          | 0,64 | 0,50                  | 0,72 | 0,50                   | 0,66 |
| <b>Formações florestais</b>                                       | 0,82                  | 0,42 | 0,65        | 0,97 | 0,69 | 0,93 | 0,75          | 0,91 | 0,72                  | 0,95 | 0,78                   | 0,93 |
| <b>Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo</b> | 0,19                  | 0,48 | 0,58        | 0,57 | 0,67 | 0,53 | 0,78          | 0,57 | 0,81                  | 0,59 | 0,80                   | 0,60 |
| <b>Culturas temporárias</b>                                       | 0,68                  | 0,10 | 0,71        | 0,46 | 0,77 | 0,49 | 0,79          | 0,52 | 0,76                  | 0,76 | 0,76                   | 0,73 |
| <b>Solo exposto/palhada</b>                                       | 0,74                  | 0,20 | 0,45        | 0,68 | 0,45 | 0,45 | 0,50          | 0,58 | 0,50                  | 0,54 | 0,58                   | 0,51 |
| <b>Silvicultura</b>   | 0,09                  | 0,10 | 0,65        | 0,04 | 0,33 | 0,11 | 0,33          | 1,00 | 0,32                  | 0,07 | 0,33                   | 0,25 |
| <b>Corpo d'água continental</b>                                   | 0,00                  | -    | -           | 0,00 | -    | 0,00 | -             | -    | -                     | -    | -                      | -    |
| <b>Áreas com edificações</b>                                      | 0,23                  | 0,51 | 1,00        | 0,24 | 1,00 | 0,93 | 1,00          | 0,72 | 1,00                  | 0,47 | 1,00                   | 0,72 |
| <b>Campo Limpo</b>  | 0,01                  | 0,08 | 0,70        | 0,15 | 0,24 | 0,12 | 0,26          | 0,27 | 0,52                  | 0,20 | 0,44                   | 0,42 |

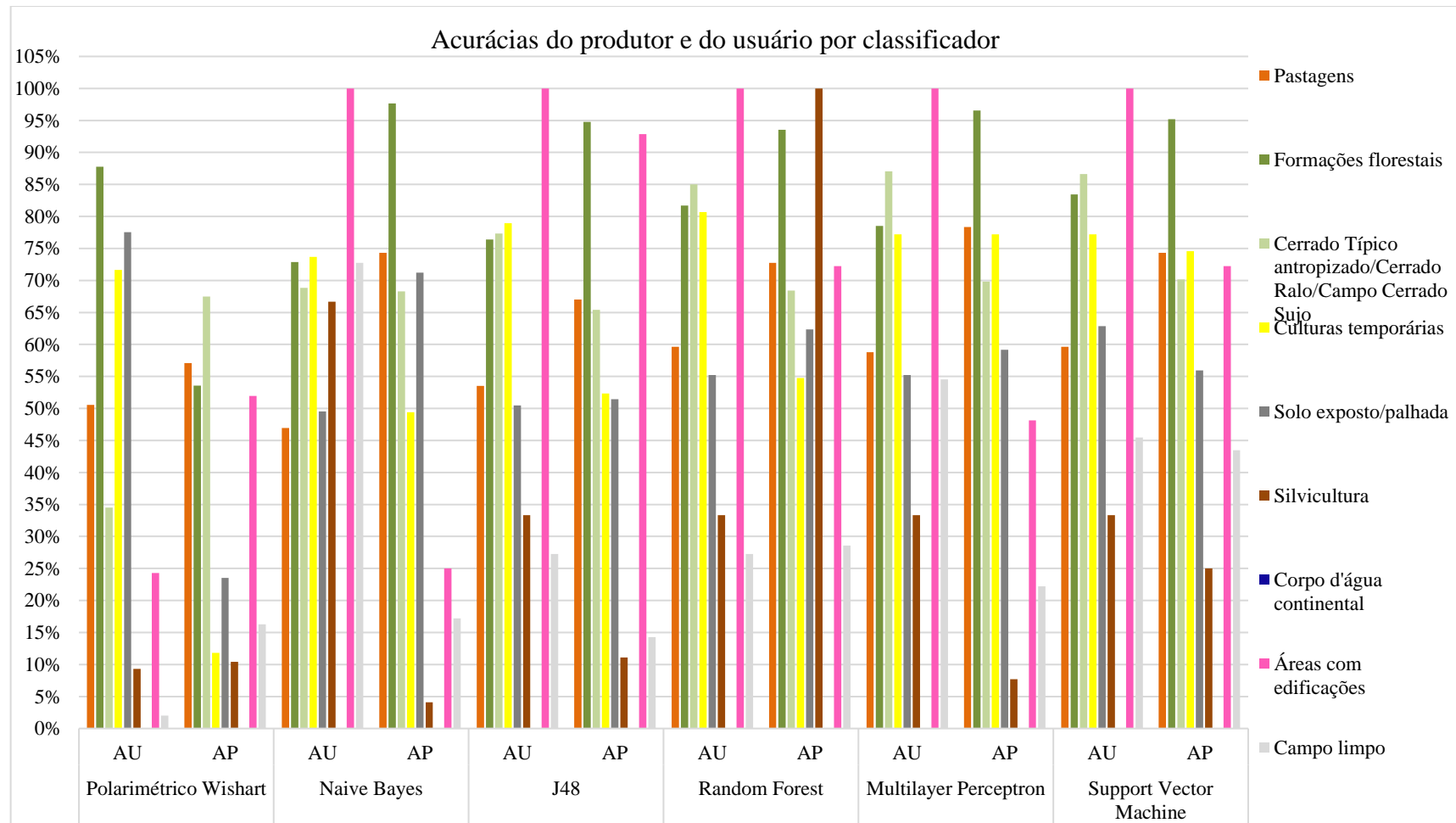


Figura 4.15. Representação gráfica das acurácias do usuário e do produtor para cada classificador utilizado (legenda com nove classes).

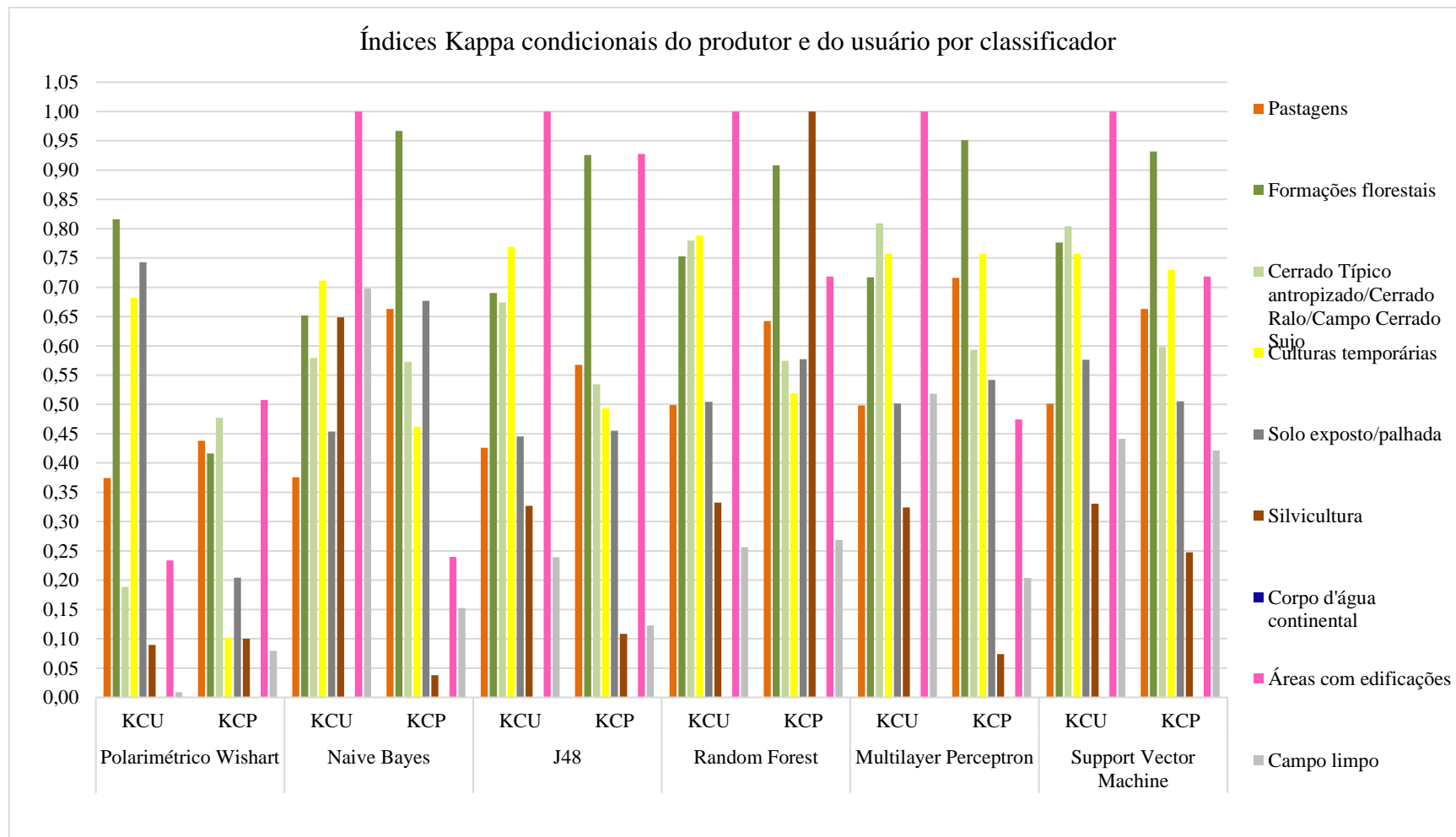


Figura 4.16. Representação gráfica dos valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com nove classes).

Tabela 4.5. Acurácias do usuário e do produtor para cada classificador utilizado (legenda com cinco classes).

|  | Polarimétrico Wishart |        | Naive Bayes |        | J48     |        | Random Forest |        | Multilayer Perceptron |        | Support Vector Machine |        |
|--|-----------------------|--------|-------------|--------|---------|--------|---------------|--------|-----------------------|--------|------------------------|--------|
|  | AU                    | AP     | AU          | AP     | AU      | AP     | AU            | AP     | AU                    | AP     | AU                     | AP     |
| <b>Cobertura vegetal herbácea</b>          | 60,28%                | 48,24% | 58,95%      | 73,36% | 64,21%  | 68,28% | 71,93%        | 75,65% | 64,91%                | 81,14% | 66,67%                 | 78,51% |
| <b>Cobertura vegetal arbórea-arbustiva</b> | 78,52%                | 86,54% | 86,52%      | 90,59% | 90,82%  | 91,51% | 95,13%        | 91,37% | 94,94%                | 91,85% | 95,69%                 | 91,58% |
| <b>Solo exposto</b>                        | 56,42%                | 52,59% | 66,14%      | 50,60% | 57,48%  | 50,34% | 55,12%        | 61,40% | 63,78%                | 53,29% | 65,35%                 | 58,87% |
| <b>Corpo d'água continental</b>            | 0,00%                 | -      | -           | 0,00%  | -       | 0,00%  | -             | -      | -                     | -      | -                      | -      |
| <b>Áreas com edificações</b>               | 24,27%                | 51,96% | 100,00%     | 25,00% | 100,00% | 92,86% | 100,00%       | 72,22% | 100,00%               | 48,15% | 100,00%                | 72,22% |

Tabela 4.6. Valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com cinco classes).

|  | Polarimétrico Wishart |      | Naive Bayes |      | J48  |      | Random Forest |      | Multilayer Perceptron |      | Support Vector Machine |      |
|--|-----------------------|------|-------------|------|------|------|---------------|------|-----------------------|------|------------------------|------|
|  | KCU                   | KCP  | KCU         | KCP  | KCU  | KCP  | KCU           | KCP  | KCU                   | KCP  | KCU                    | KCP  |
| <b>Cobertura vegetal herbácea</b>          | 0,42                  | 0,31 | 0,46        | 0,62 | 0,50 | 0,55 | 0,61          | 0,65 | 0,54                  | 0,73 | 0,55                   | 0,69 |
| <b>Cobertura vegetal arbórea-arbustiva</b> | 0,54                  | 0,67 | 0,71        | 0,79 | 0,79 | 0,81 | 0,88          | 0,81 | 0,88                  | 0,82 | 0,90                   | 0,81 |
| <b>Solo exposto</b>                        | 0,49                  | 0,46 | 0,59        | 0,43 | 0,50 | 0,43 | 0,49          | 0,56 | 0,57                  | 0,46 | 0,59                   | 0,53 |
| <b>Corpo d'água continental</b>            | 0,00                  | -    | -           | 0,00 | -    | 0,00 | 0,00          | -    | -                     | -    | -                      | -    |
| <b>Áreas com edificações</b>               | 0,23                  | 0,51 | 1,00        | 0,24 | 1,00 | 0,93 | 1,00          | 0,72 | 1,00                  | 0,47 | 1,00                   | 0,72 |

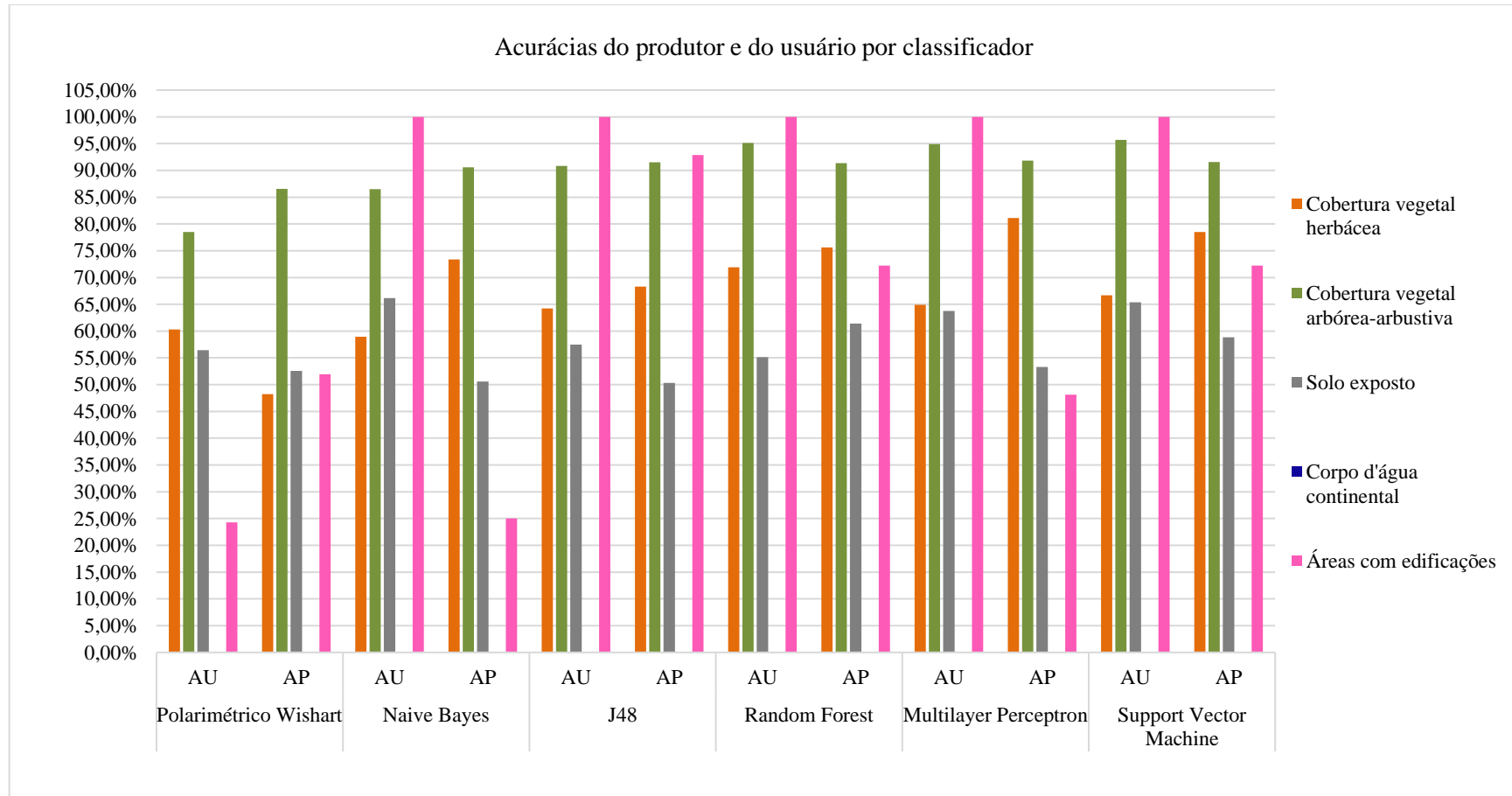


Figura 4.17. Representação gráfica das acurácias do usuário e do produtor para cada classificador utilizado (legenda com cinco classes).



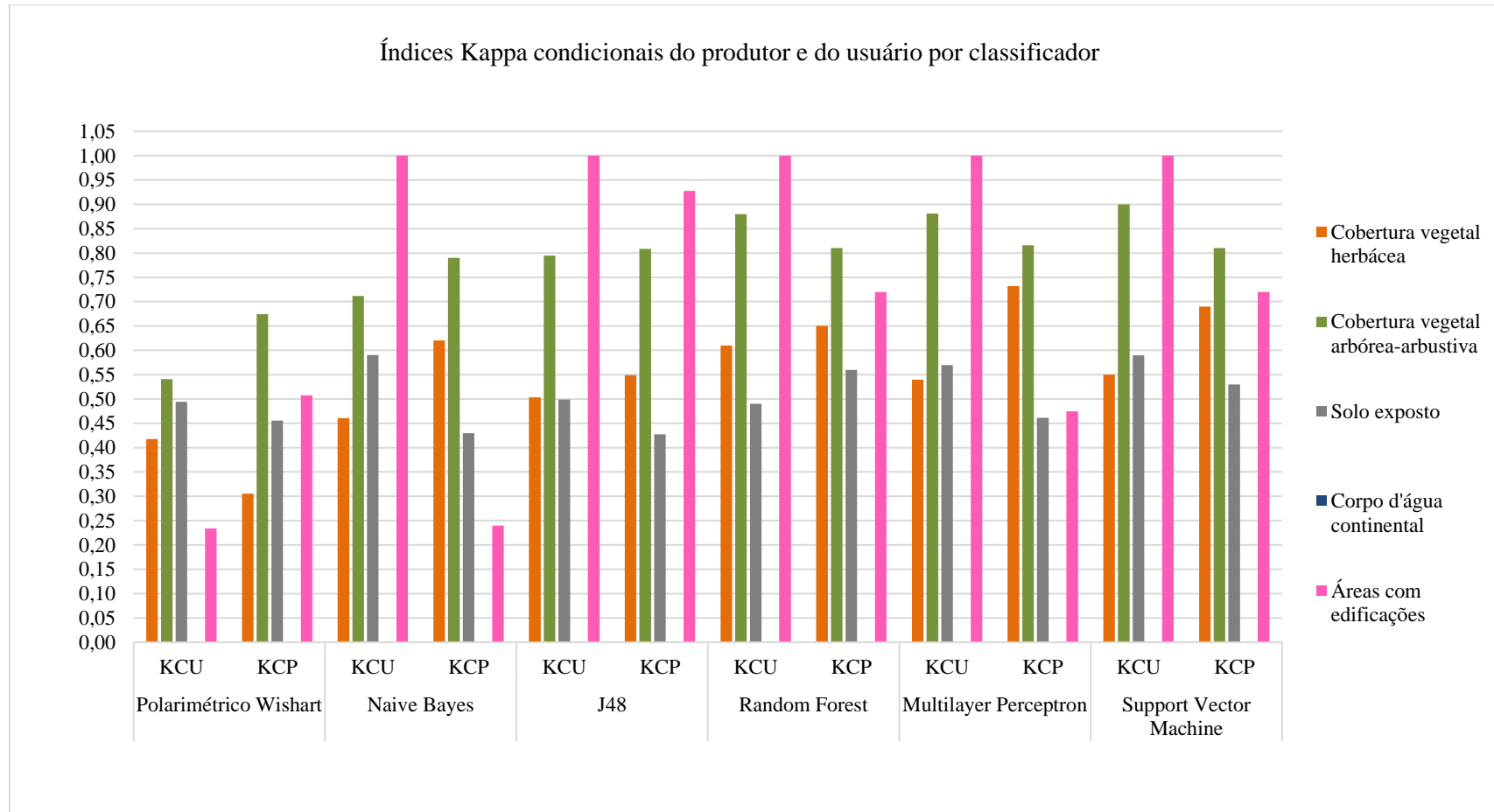


Figura 4.18. Representação gráfica dos valores de índice Kappa condicional do usuário e do produtor para cada classificador utilizado (legenda com cinco classes).

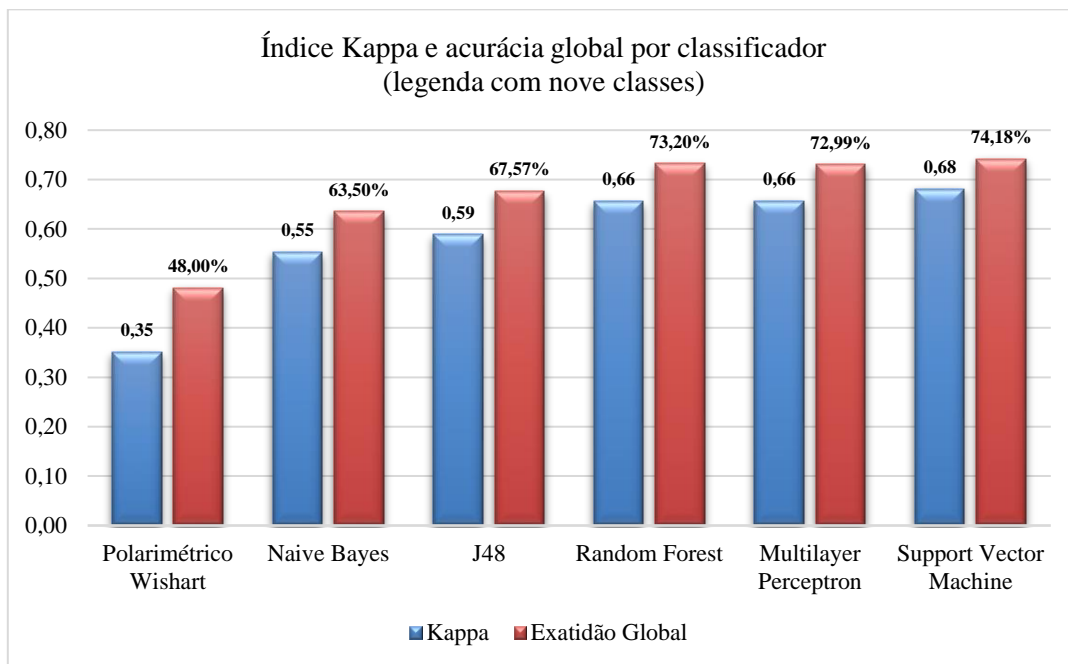


Figura 4.19. Índice Kappa e acurácia global por classificador. Legenda com nove classes e conjunto amostral com 200 amostras por classe.

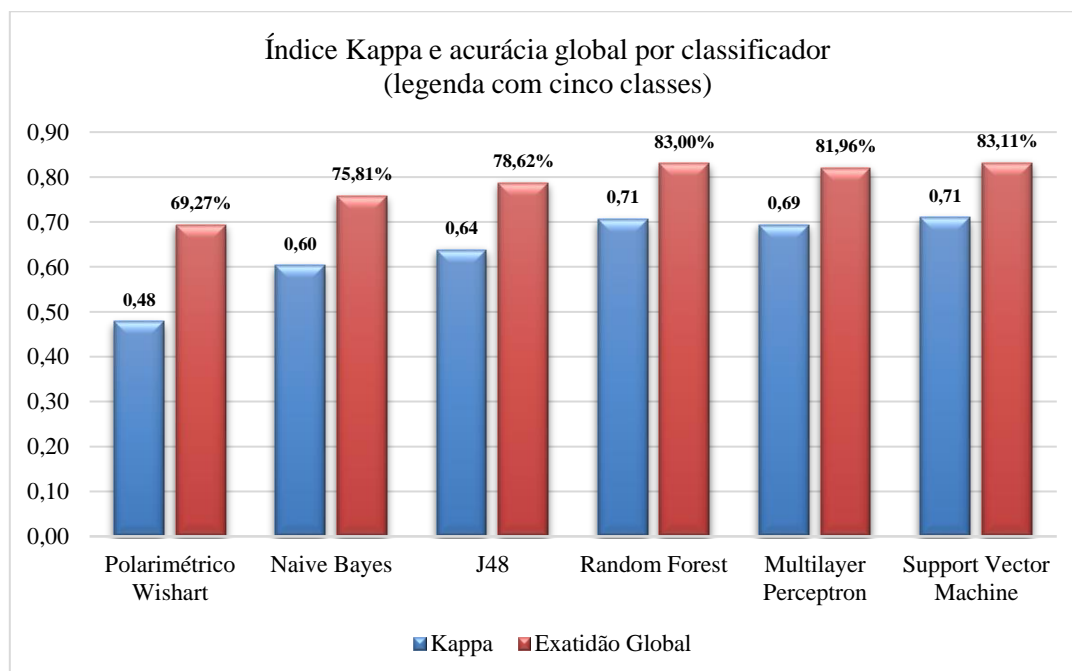
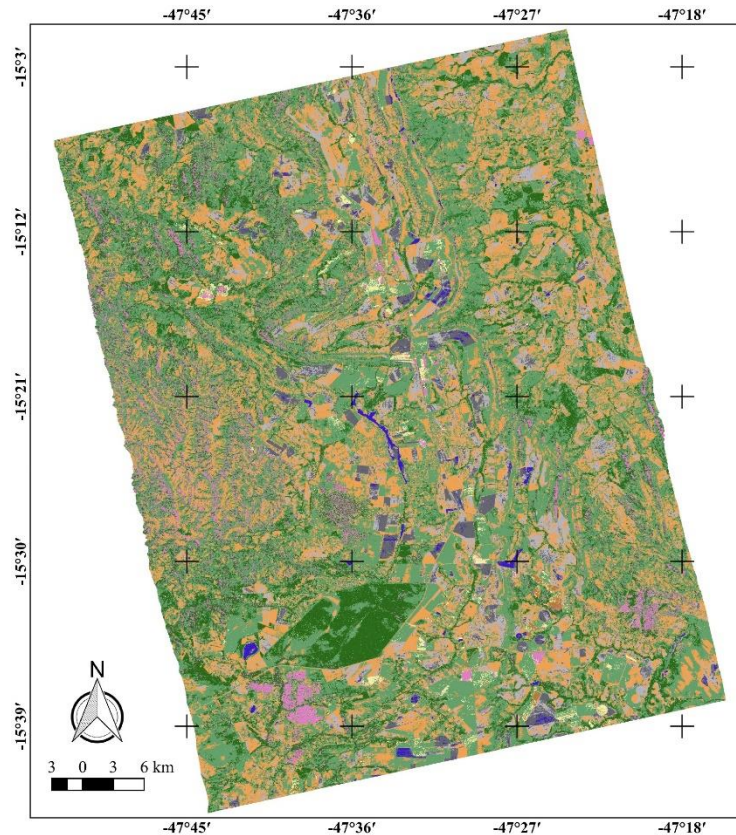
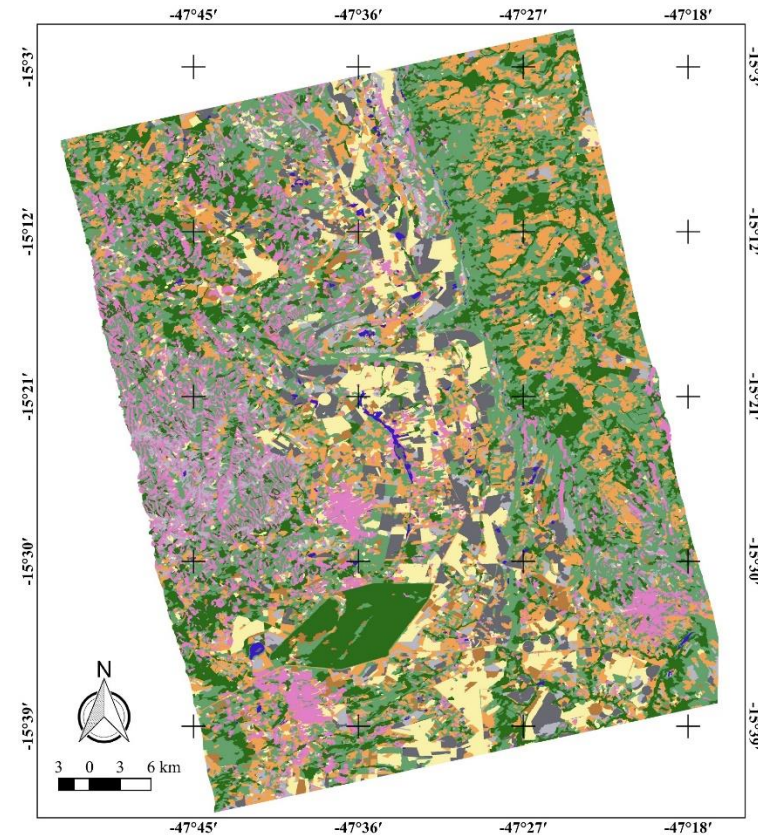


Figura 4.20. Índice Kappa e acurácia global por classificador. Legenda com cinco classes e conjunto amostral com 200 amostras por classe.



a)

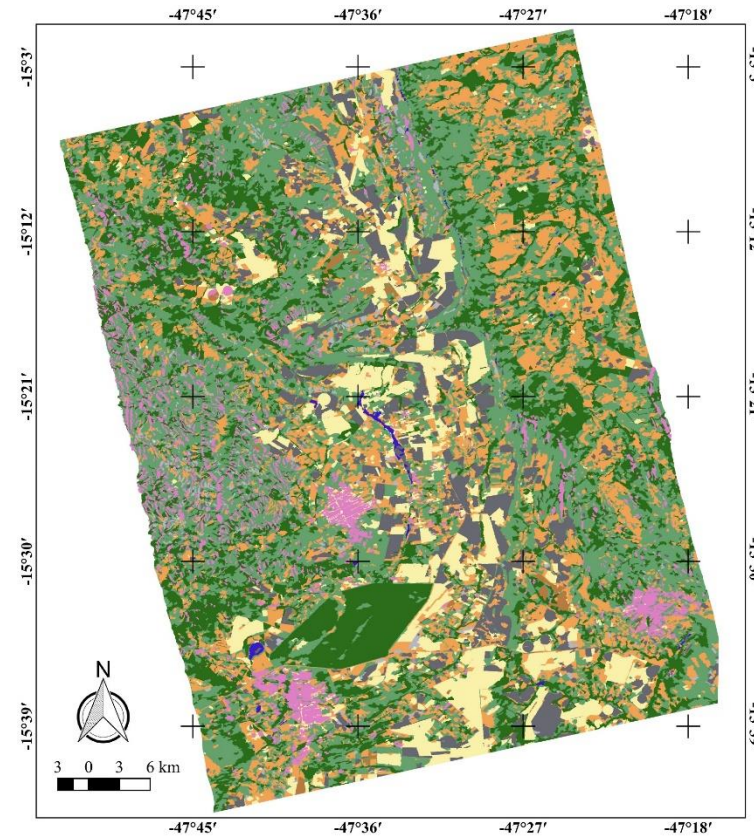
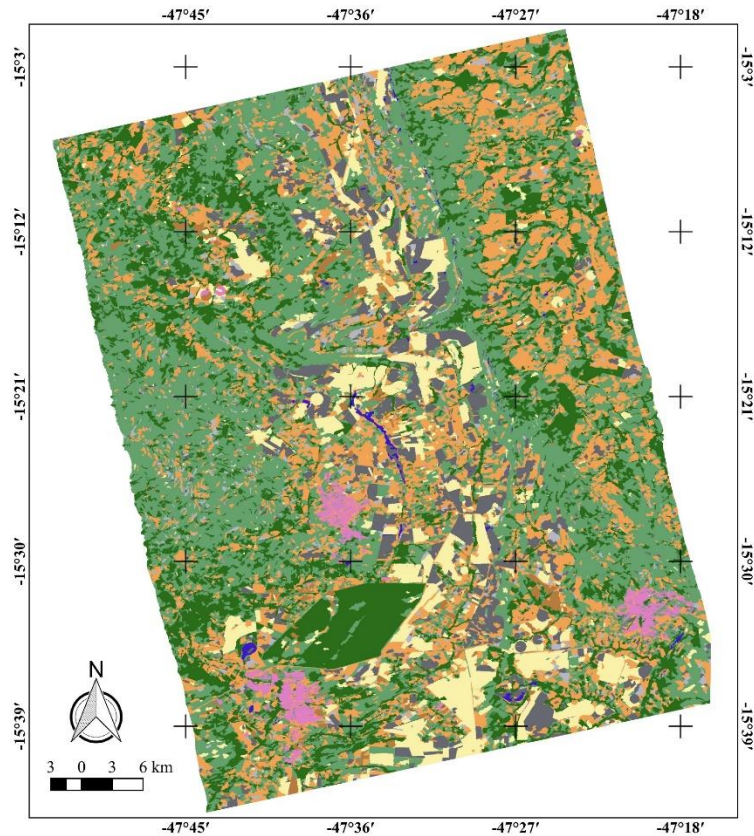


b)



Figura 4.21. Classificações obtidas com os algoritmos PW (a) e NB (b). Legenda com nove classes.

*Datum* horizontal: WGS84; Projeção: UTM 23S.



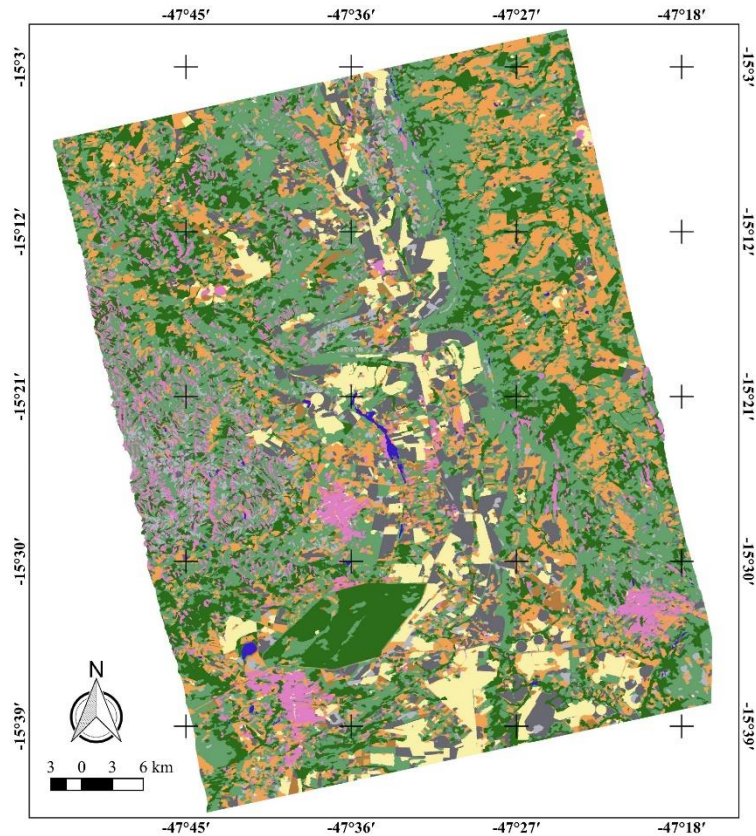
a)

b)

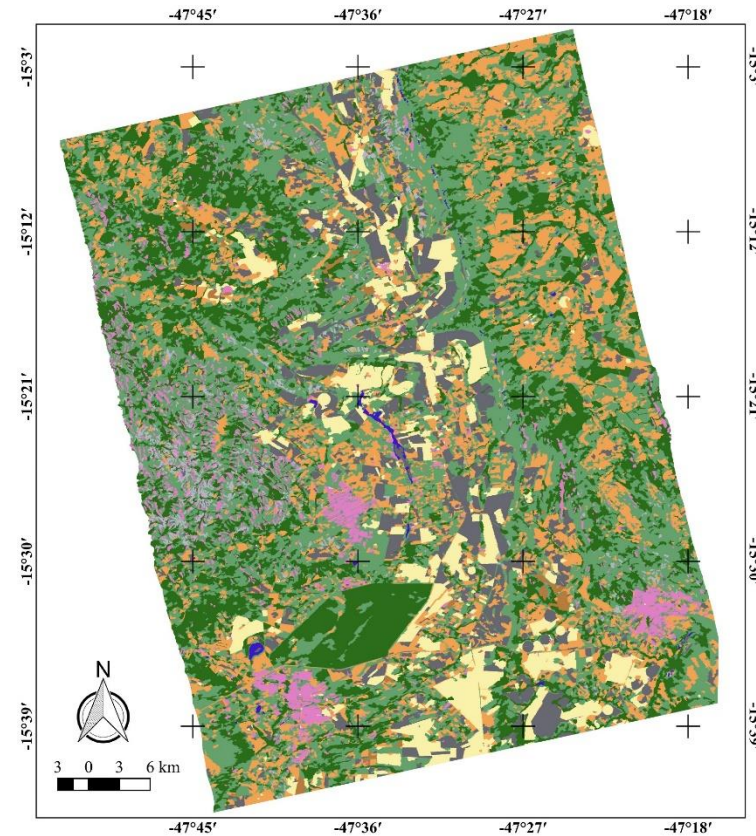


Figura 4.22. Classificações obtidas com os algoritmos AD J48 (a) e RF (b). Legenda com nove classes.

*Datum* horizontal: WGS84; Projeção: UTM 23S.



a)



b)



Figura 4.23. Classificações obtidas com os algoritmos MLP (a) e SVM (b). Legenda com nove classes.  
*Datum* horizontal: WGS84; Projeção: UTM 23S.

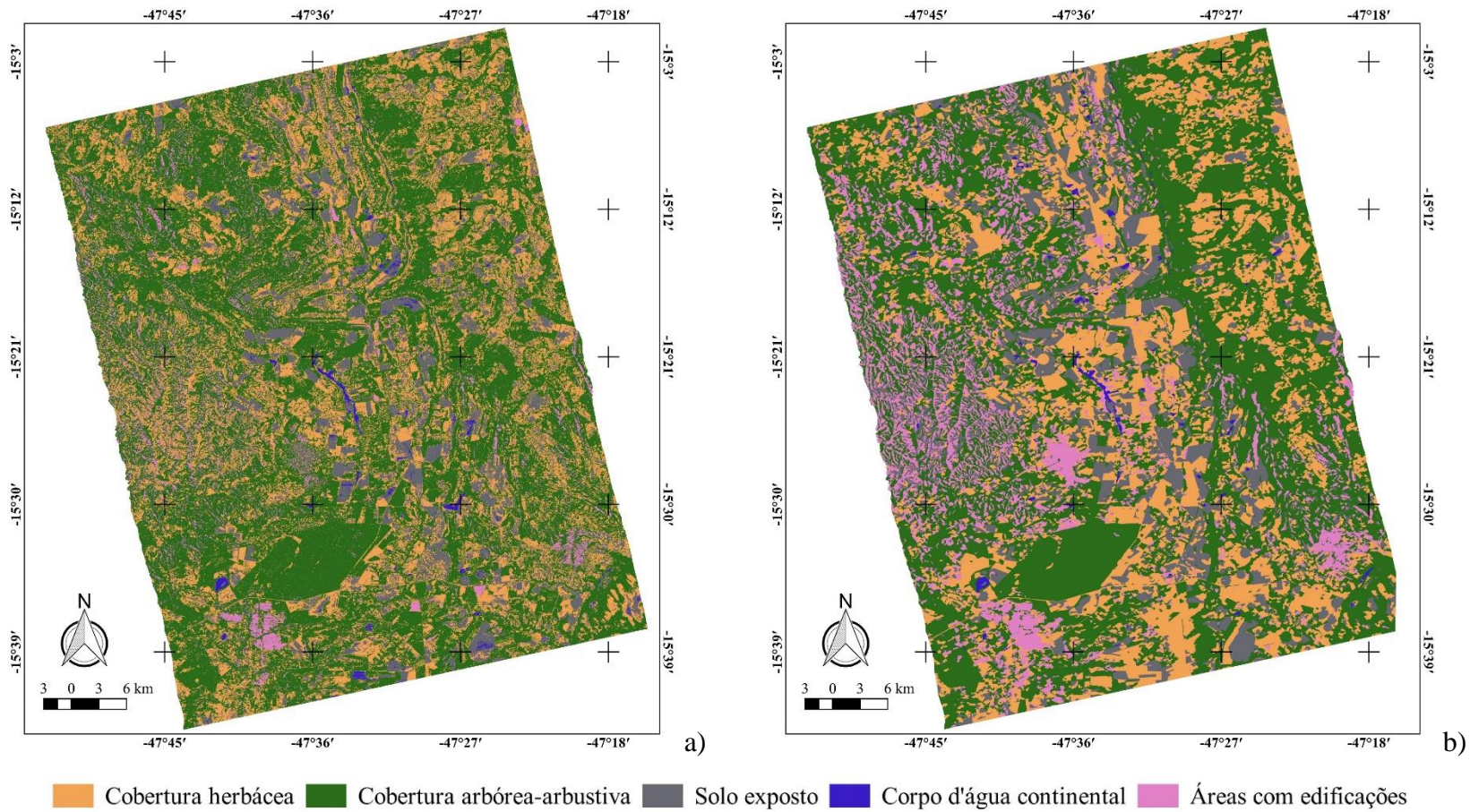


Figura 4.24. Classificações obtidas com os algoritmos PW (a) e NB (b). Legenda com cinco classes.  
*Datum* horizontal: WGS84; Projeção: UTM 23S.

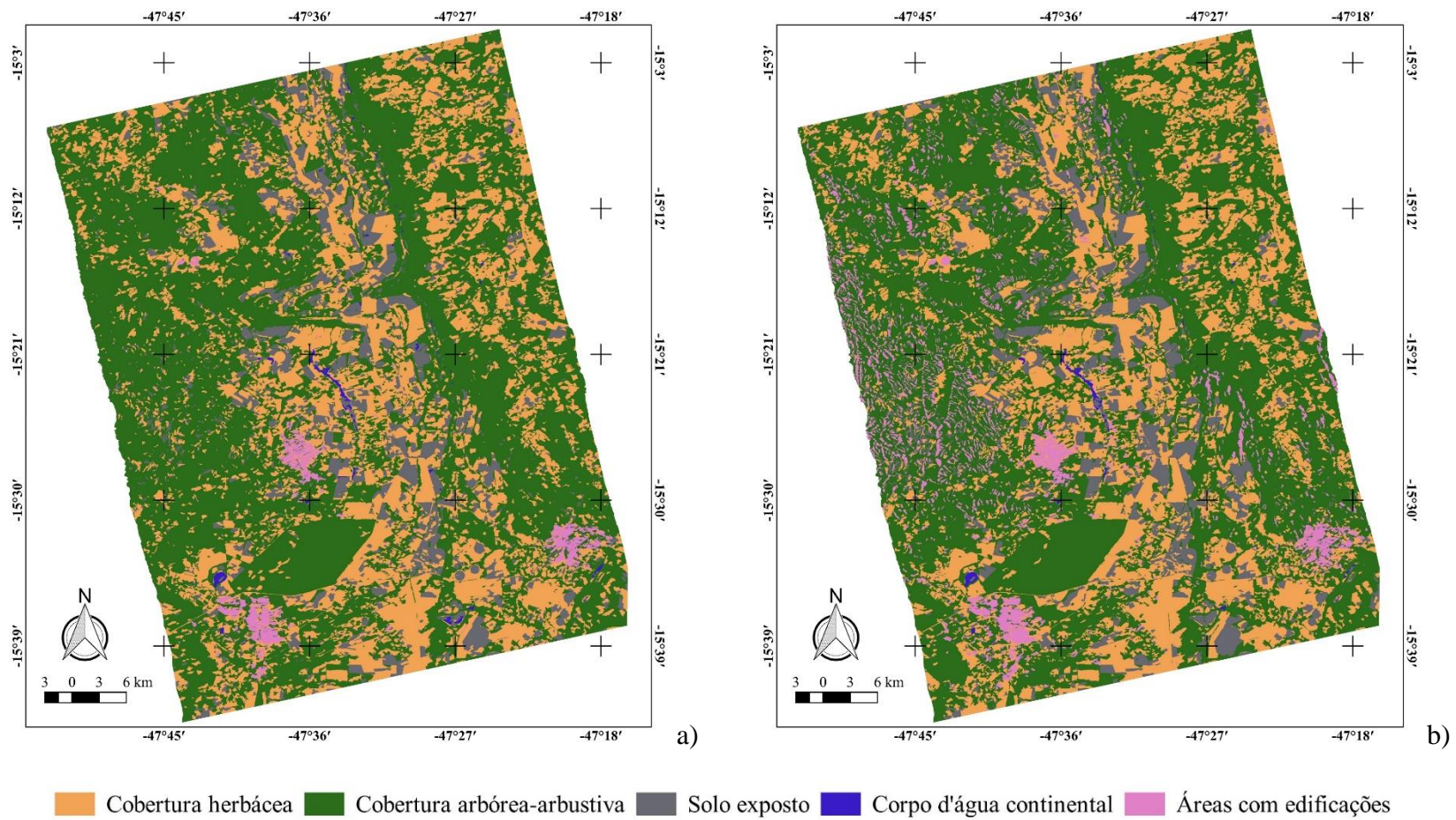


Figura 4.25. Classificações obtidas com os algoritmos AD J48 (a) e RF (b). Legenda com cinco classes.  
*Datum* horizontal: WGS84; Projeção: UTM 23S.

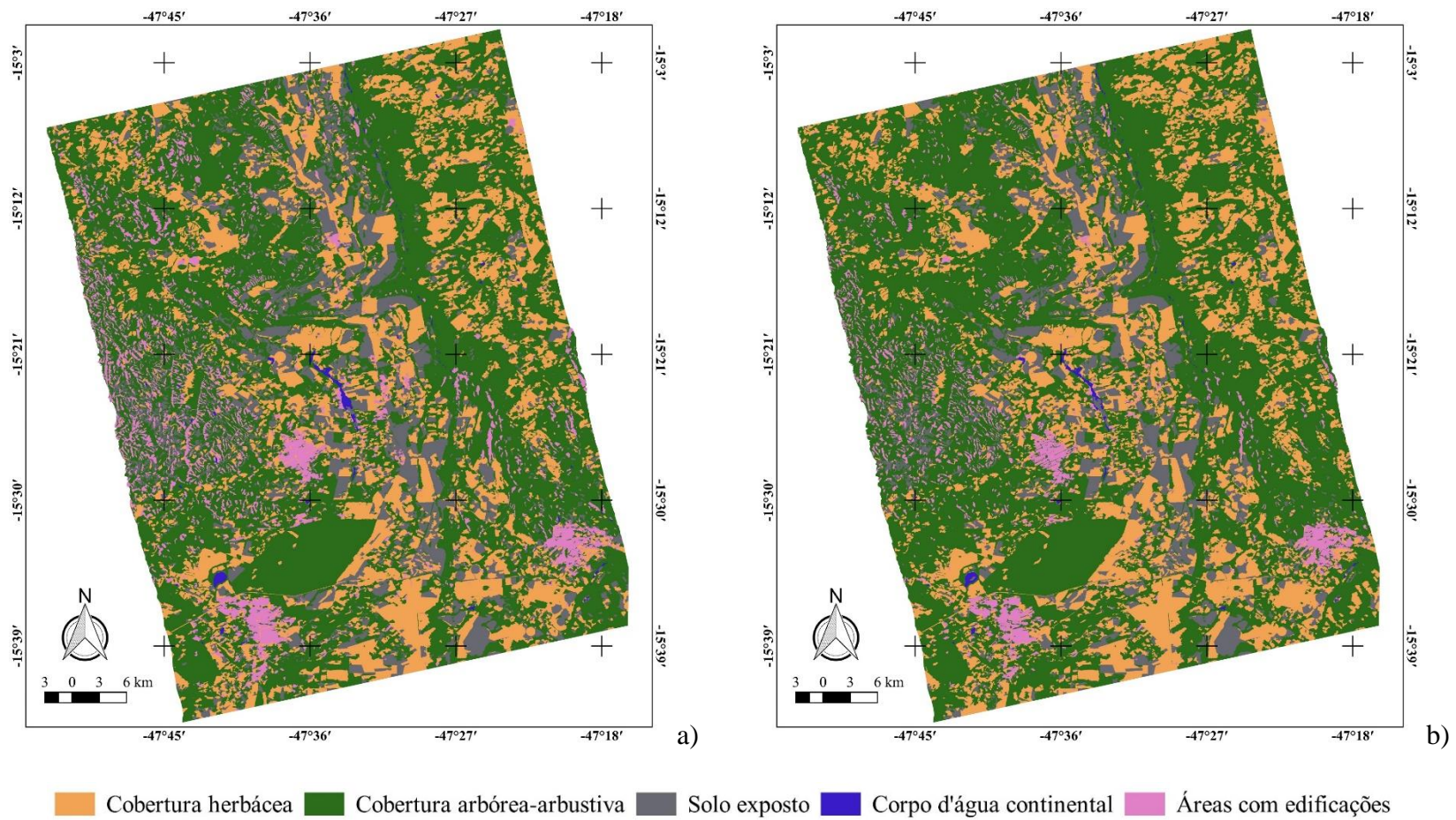


Figura 4.26. Classificações obtidas com os algoritmos MLP (a) e SVM (b). Legenda com cinco classes. Datum horizontal: WGS84; Projeção: UTM 23S.



Nas Figuras 4.21 a 4.26, verifica-se de modo mais fácil o desempenho dos classificadores. O classificador NB apresentou, de modo exagerado, erros de inclusão para a classe “Áreas com edificações”. Houve muitos polígonos classificados nessa classe, sendo que as áreas urbanas são restritas na cena utilizada. Esses erros de classificação se devem à falta de habilidade do NB para utilizar os atributos disponíveis e, também, devido a muitas áreas declivosas apresentando *layover* (inversão de relevo) e *foreshortening* (encurtamento de rampa) que se confundiram com as áreas urbanas.

Com relação ao classificador PW, verifica-se na Figura 4.21a que seu desempenho foi adequado somente na separação entre cobertura vegetal arbórea e não arbórea. Secundariamente, o PW apresentou bom desempenho na identificação das áreas de pastagens a nordeste da cena.

Os demais classificadores apresentaram padrões equilibrados de distribuição das classes. O classificador AD J48 despertou atenção pela identificação mais precisa das áreas com edificações, como se observa na Figura 4.22a sem demasiada confusão com outras classes ou com áreas apresentando *layover* (inversão de relevo) e *foreshortening* (encurtamento de rampa). Esse comportamento do classificador AD J48 também foi identificado na Figura 4.16 e discutido anteriormente.

Nas Figuras 4.27 e 4.28, análises visuais qualitativas foram realizadas comparando as classificações do SVM com o mapa de uso e cobertura do bioma Cerrado produzido pelo projeto TerraClass (INPE, 2017). Apesar das diferenças entre legendas e procedimentos empregados (TerraClass usa dados ópticos e interpretação visual), observaram-se muitas similitudes entre os padrões identificados. Cabe considerar também a defasagem temporal entre os dois levantamentos (o TerraClass é de 2013).

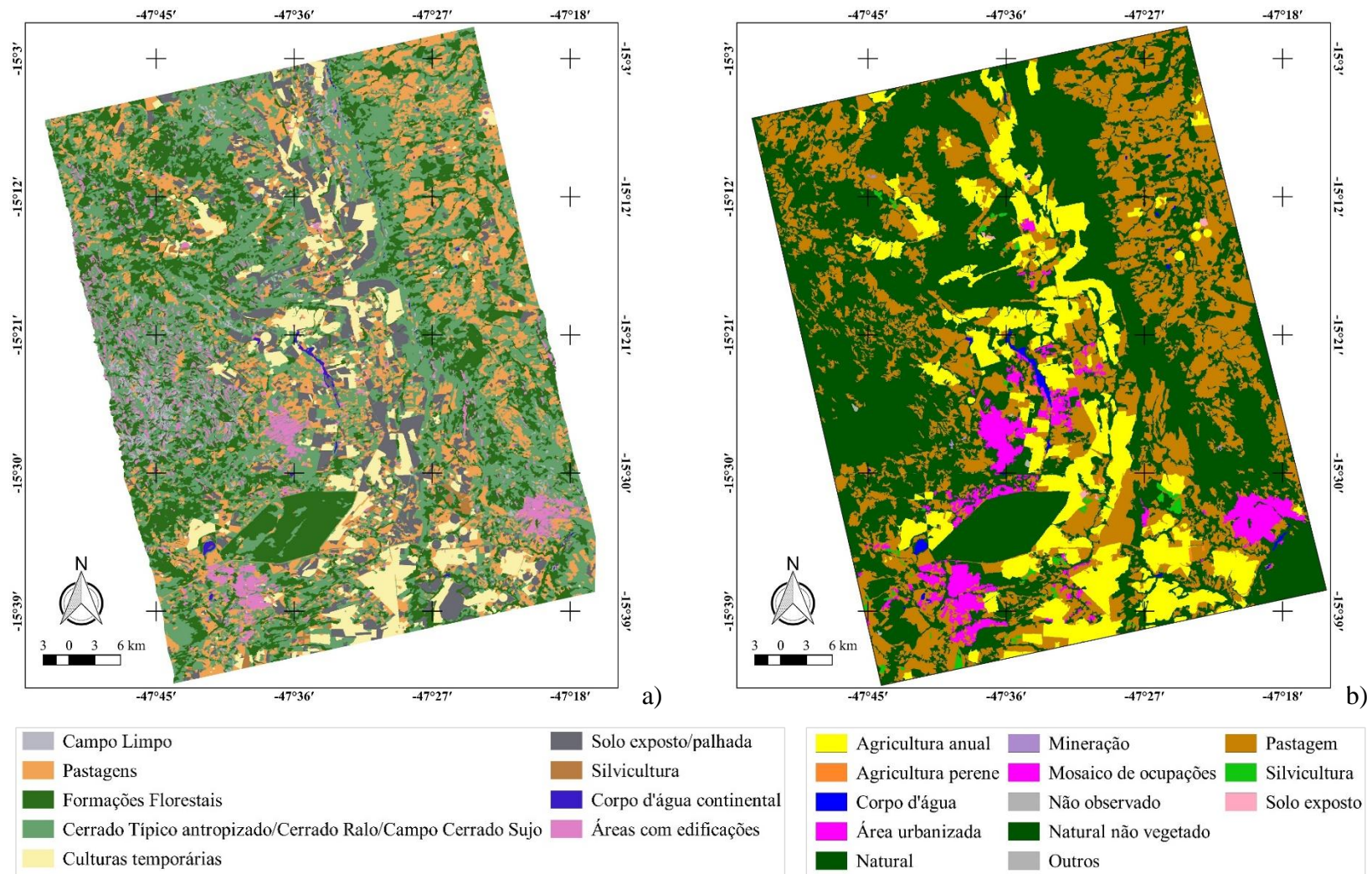


Figura 4.27. a) Classificação obtida com o algoritmo SVM e conjunto de treinamento de 200 amostras por classe (legenda com nove classes). b) Classificação gerada pelo programa TerraClass (INPE, 2017). Datum horizontal: WGS84; Projeção: UTM 23S.

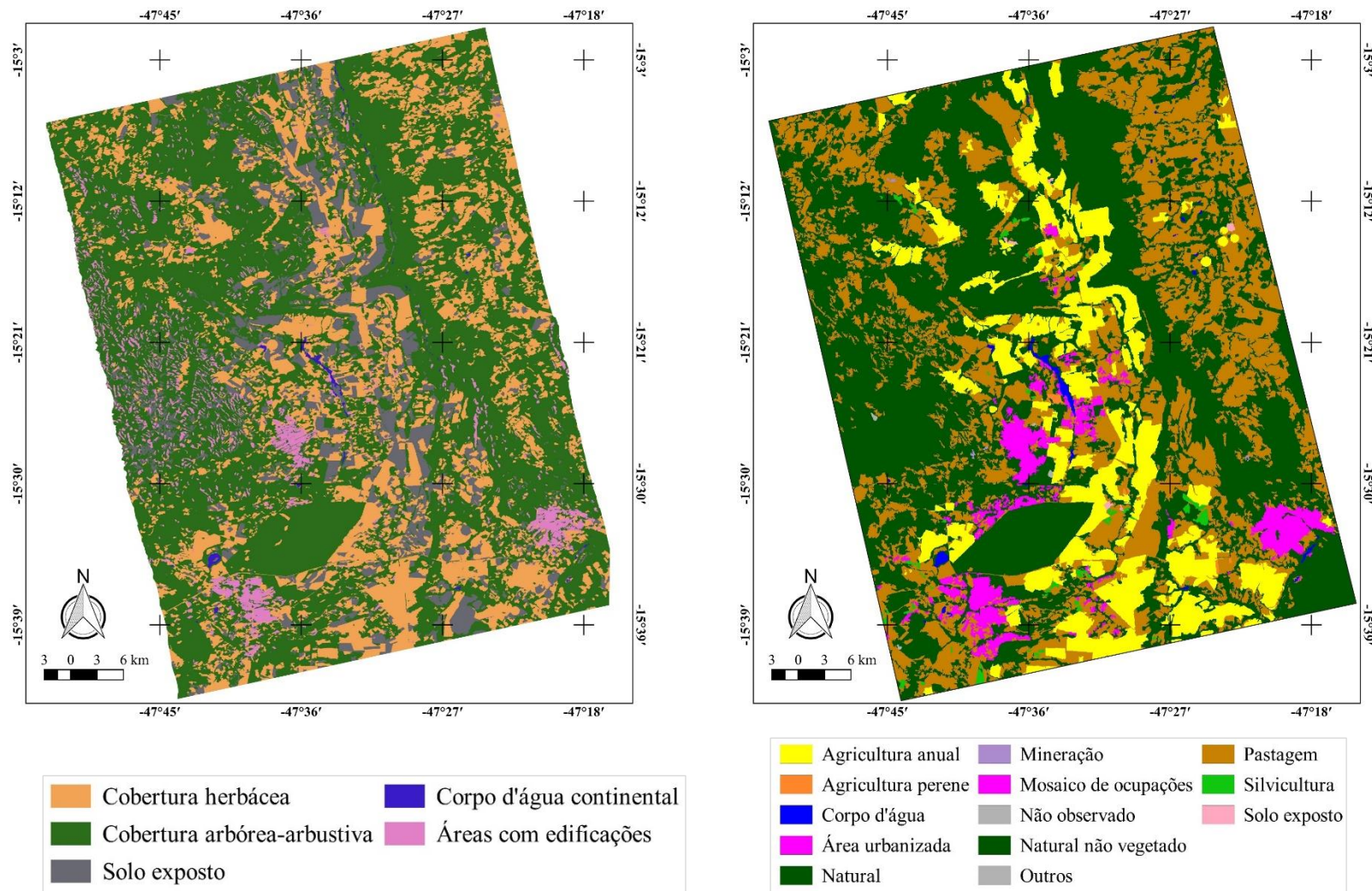


Figura 4.28. a) Classificação obtida com o algoritmo SVM e conjunto de treinamento de 200 amostras por classe (legenda com cinco classes). b) Classificação gerada pelo programa TerraClass (INPE, 2017). Datum horizontal: WGS84; Projeção: UTM 23S.

Para uma análise robusta dos classificadores, visando à geração de um *ranking* objetivo, empregaram-se nesta pesquisa testes de hipóteses (Apêndices G a J) com base na estatística Z (distribuição normal padrão), de modo a verificar se os índices Kappa são iguais ou diferentes em dado nível de significância (CONGALTON e GREEN, 2009). Nas Tabelas 4.7 e 4.8, são apresentadas as matrizes com os p-valores dos testes para cada par de classificadores. Os p-valores destacados em amarelo são maiores que o nível de significância adotado no teste e, desse modo, unem nas matrizes os classificadores iguais em desempenho. As Figuras 4.29 e 4.30 apresentam graficamente as informações das Tabelas 4.7 e 4.8.

A análise da Tabela 4.7 e da Figura 4.29, considerando a legenda de nove classes, permite constatar que os classificadores RF, MLP e SVM são iguais em termos de índices de concordância Kappa e, por conseguinte, em desempenho geral. Verifica-se também que os classificadores NB e AD J48 também são iguais em desempenho, porém, inferiores (e diferentes) aos anteriores mencionados. O classificador PW, por sua vez, foi inferior a todos os demais.

Para os testes referentes à legenda de cinco classes (Apêndice J), observa-se, na Tabela 4.8 e na Figura 4.30, a manutenção das igualdades já verificadas nos testes aplicados nas classificações com nove classes. Desse modo, definiu-se um *ranking* dos classificadores, segundo seus desempenhos relativos ao índice Kappa (Quadro 4.12). Este *ranking* é um resultado objetivo desta pesquisa e pode ser adotado como um indicativo para outros trabalhos com dados, legenda e área semelhantes.

Tabela 4.7. Matriz de p-valores obtidos nas comparações entre os classificadores utilizados. Legenda com nove classes.

|                               | P-valor do teste estatístico Z para comparação de índices Kappa |                    |            |                      |                              |                               |
|-------------------------------|---|--------------------|------------|----------------------|------------------------------|-------------------------------|
|                               | <b>Polarimétrico Wishart</b>                                    | <i>Naive Bayes</i> | <b>J48</b> | <i>Random Forest</i> | <i>Multilayer Perceptron</i> | <i>Support Vector Machine</i> |
| <b>Polarimétrico Wishart</b>  | -   | 0,0000             | 0,0000     | 0,0000               | 0,0000                       | 0,0000                        |
| <i>Naive Bayes</i>            | 0,0000  | -                  | 0,0837     | 0,0000               | 0,0000                       | 0,0000                        |
| <b>J48</b>                    | 0,0000  | 0,0837             | -          | 0,0046               | 0,0038                       | 0,0001                        |
| <i>Random Forest</i>          | 0,0000  | 0,0000             | 0,0046     | -                    | 0,4790                       | 0,1565                        |
| <i>Multilayer Perceptron</i>  | 0,0000  | 0,0000             | 0,0038     | 0,4790               | -                            | 0,1673                        |
| <i>Support Vector Machine</i> | 0,0000  | 0,0000             | 0,0001     | 0,1565               | 0,1673                       | -                             |

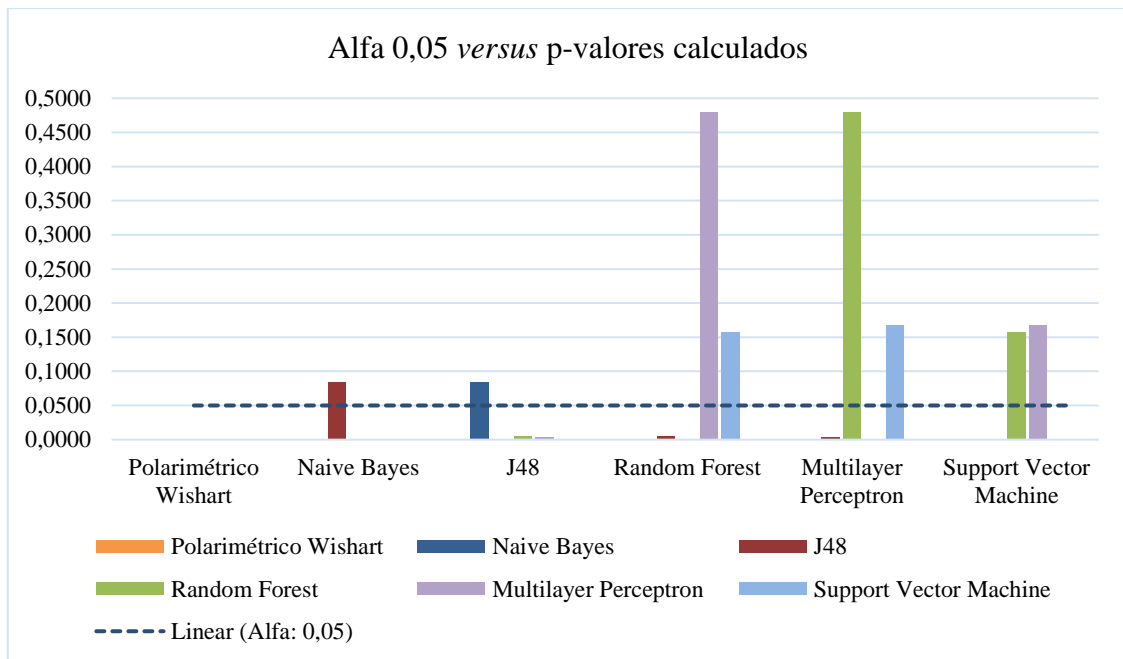


Figura 4.29. Representação gráfica dos resultados dos testes Z (legenda com nove classes). Na abscissa estão os diferentes classificadores utilizados, e na ordenada, os p-valores calculados. Classificadores que apresentarem p-valores maiores que  $\alpha$  (0,05) são considerados iguais aos seus correspondentes na abscissa.

Tabela 4.8. Matriz de p-valores obtidos nas comparações entre os classificadores utilizados. Legenda com cinco classes.

|                               | P-valor do teste estatístico Z para comparação de índices Kappa |             |        |               |                       |                        |
|-------------------------------|---|-------------|--------|---------------|-----------------------|------------------------|
|                               | Polarimétrico Wishart   | Naive Bayes | J48    | Random Forest | Multilayer Perceptron | Support Vector Machine |
| <b>Polarimétrico Wishart</b>  | -   | 0,0000      | 0,0000 | 0,0000        | 0,0000                | 0,0000                 |
| <b>Naive Bayes</b>            | 0,0000  | -           | 0,1242 | 0,0002        | 0,0008                | 0,0001                 |
| <b>J48</b>                    | 0,0000  | 0,1242      | -      | 0,0087        | 0,0260                | 0,0058                 |
| <b>Random Forest</b>          | 0,0000  | 0,0002      | 0,0087 | -             | 0,3265                | 0,4449                 |
| <b>Multilayer Perceptron</b>  | 0,0000  | 0,0008      | 0,0260 | 0,3265        | -                     | 0,2766                 |
| <b>Support Vector Machine</b> | 0,0000  | 0,0001      | 0,0058 | 0,4449        | 0,2766                | -                      |

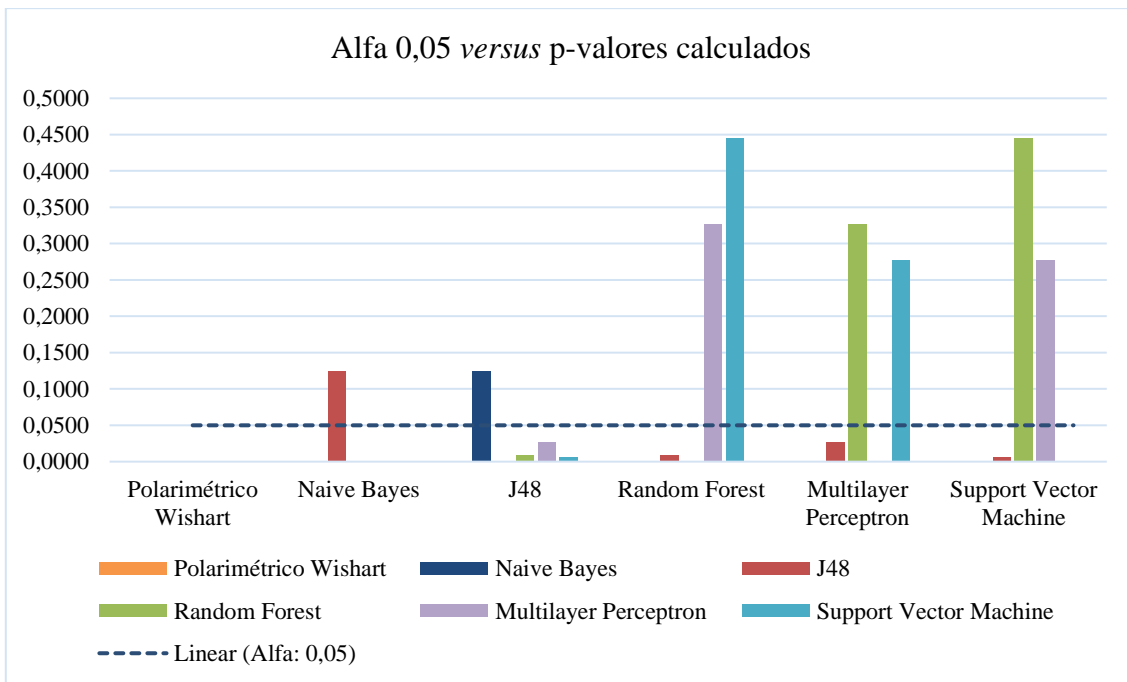


Figura 4.30. Representação gráfica dos resultados do teste Z (legenda com cinco classes). Na abscissa estão os diferentes classificadores utilizados e na ordenada os p-valores calculados. Classificadores que apresentarem p-valores maiores que  $\alpha$  (0,05) são considerados iguais aos seus correspondentes na abscissa.

Quadro 4.12. *Ranking* final dos classificadores para as legendas de nove e cinco classes.

| <b>Ranking Final – Legendas com nove e cinco classes</b> |                               |                             |                                    |                              |                                     |
|--|-------------------------------|-----------------------------|------------------------------------|------------------------------|-------------------------------------|
| Posição  | Classificador                 | Índice Kappa (nove classes) | Acurácia Global (%) (nove classes) | Índice Kappa (cinco classes) | Acurácia Global (%) (cinco classes) |
| 1°   | <i>Support Vector Machine</i> | 0,68                        | 74,18                              | 0,71                         | 83,11                               |
|  | <i>Random Forest</i>          | 0,66                        | 73,20                              | 0,71                         | 83,00                               |
|  | <i>Multilayer Perceptron</i>  | 0,66                        | 72,99                              | 0,69                         | 81,96                               |
| 2°   | J48                           | 0,59                        | 65,57                              | 0,64                         | 78,62                               |
|  | <i>Naive Bayes</i>            | 0,55                        | 63,50                              | 0,60                         | 75,81                               |
| 3°   | Polarimétrico Wishart         | 0,35                        | 48,00                              | 0,48                         | 69,27                               |

Dentre todos os classificadores utilizados, apenas o AD J48 apresentou de forma explícita os atributos selecionados para a classificação (Apêndices B e C). Essa informação dos atributos efetivamente utilizados é relevante para uma análise da importância relativa de cada método de decomposição ou de cada parâmetro polarimétrico, por exemplo. Outra forma de se analisar a importância dos conjuntos de atributos seria o emprego preliminar de métodos de seleção de atributo, tais como o algoritmo *Infogain* empregado nesta pesquisa (Apêndice D).

Com base nas informações do algoritmo *Infogain* e no AD J48, verificou-se, de modo geral, que todas as camadas de entrada (oito parâmetros polarimétricos, 13 componentes de decomposições e quatro polarizações) foram utilizadas para classificação. O que variou foi a métrica aplicada ao segmento. Ou seja, ora se utiliza a média, ora o desvio-padrão, valor mínimo, etc. No caso das classes do AD J48 com melhor desempenho (“Áreas com edificações” e “Formações florestais”), os atributos finais responsáveis pela classificação foram respectivamente a razão de polarização cruzada (HH/HV) e o índice VSI. O emprego desse índice para separar formação florestal é coerente, haja vista que representa ou estima o espalhamento volumétrico de dosséis florestais (POPE et al., 1994).

Observou-se também que os componentes referentes ao espalhamento volumétrico (van Zyl, Freeman-Durden e Yamaguchi) foram elencadas tanto na AD J48 quanto na relação obtida com o *Infogain*. Esses componentes mencionados têm relação com a estrutura do dossel da vegetação (SANO et al., 2001; 2005; BITENCOURT et al., 2007; URBAZAEV et al., 2015; ODIPO et al., 2016; BOUVET et al., 2018). Os componentes H, A e  $\alpha$  também foram elencados. Destaque deve ser dado aos índices CSI, VSI, BMI e RFDI, tanto na AD J48 quanto na relação do *Infogain*. As razões de polarização cruzada também participaram em menor grau e as polarizações cruzadas em amplitude (HV e VH) aparecem também nas duas listagens (AD J48 e *Infogain*).

#### **4.6 - Conclusões**

O fluxo de procedimentos apresentado neste trabalho mostrou-se viável para a tarefa de mapeamento de uso e cobertura de terras em área de Cerrado. De acordo com a classificação proposta por Landis e Koch (1977) e considerando a legenda com nove classes, os classificadores baseados em aprendizado de máquina atingiram as classes “boa” (casos do NB e AD J48) ou “muito boa” (RF, MLP e SVM). Para a legenda agregada com cinco classes, todos os classificadores de mineração de dados atingiram a categoria “muito boa”.

O classificador PW se mostrou mais adequado para a identificação de padrões mais genéricos de uso e cobertura de terras, tais como vegetação e não-vegetação. Alguns trabalhos que testaram o PW também aplicaram legendas com poucas classes e, assim, corroboram esta afirmação (LEE e POTTIER, 2009; SINGH et al., 2014). De qualquer modo, as classificações do PW foram categorizadas como “Razoável”, na legenda com nove classes, e como “Boa” na legenda com cinco classes.

Com base nas classificações e resultados, foi possível identificar três grupos de classificadores. Um primeiro grupo de classificadores, que obteve os melhores resultados, constitui-se dos algoritmos RF, MLP e SVM, os quais obtiveram desempenhos semelhantes e apresentaram índices Kappa estatisticamente iguais. Um segundo grupo, que apresentou resultados intermediários, compõem-se dos classificadores NB e AD J48, os quais apresentaram índices Kappa estatisticamente iguais. Um terceiro e último grupo é constituído unicamente pelo classificador *pixel-a-pixel* PW, inferior a todos os demais.

Verifica-se que o classificador RF supera em desempenho a AD J48, em consonância com a literatura científica pertinente (BELGIU e DRAGUT, 2016). A AD J48, apesar de mais complexa, não apresentou resultado superior ao classificador NB. Todavia, o alto índice de acerto da AD J48, em relação à classe “Áreas com edificações”, propiciou mapas temáticos com menos erros de inclusão nesta classe e com uma aparência melhor nas áreas de vegetação.

Com relação aos atributos polarimétricos, verificou-se que as decomposições tiveram participação importante na identificação e classificação dos padrões de uso e cobertura de terras. As componentes referentes ao espalhamento volumétrico (Van Zyl, Freeman-Durden e Yamaguchi) foram elencadas tanto na AD J48 quanto na relação obtida com o *Infogain*. Estas componentes relativas ao espalhamento volumétrico têm relação com a estrutura do dossel da vegetação (SANO et al., 2001; SANO et al., 2005; BITENCOURT et al., 2007; URBAZAEV et al., 2015; ODIPO et al., 2016; BOUVET et al., 2018). As componentes H, A e  $\alpha$  também foram elencadas. Todavia, para estas componentes, deve ser realizado um estudo mais detalhado sobre as faixas de valores utilizados para entendimento de quais mecanismos e tipos de espalhamento prevaleceram na cena e quais as componentes mais importantes nas classificações.

Ainda com relação aos atributos polarimétricos, os parâmetros ditos incoerentes também tiveram participação importante. Destaque aos índices CSI, VSI, BMI e RFDI, tanto na AD J48 quanto na relação do *Infogain*. As razões de polarização cruzada também participaram em menor grau e destaque deve ser dado às polarizações cruzadas em amplitude (HV e VH), que aparecem nas duas listagens (AD J48 e *Infogain*).

Visando a trabalhos futuros, seria importante empregar imagens multitemporais e também, se possível, técnicas PolInSAR (Polarimetria e Interferometria SAR) nas classificações. Dados SENTINEL-1 (banda C) também poderiam ser testados isoladamente ou em conjunto com os dados ALOS-2/PALSAR-2 (banda L). Também poderia ser utilizada uma classificação não-supervisionada prévia empregando o espaço de atributos H-  $\alpha$  visando



ao entendimento dos tipos de espalhamento e assim um melhor treinamento dos classificadores, bem como para se ter uma base classificada prévia, objetivando realizar uma amostragem estratificada para fins de validação temática.

Também visando a trabalhos futuros, devem-se realizar testes com outros classificadores polarimétricos *pixel-a-pixel*, tais como, o algoritmo MaxVer-ICM polarimétrico desenvolvido por Correia (1998).

Por fim, considerando a complexidade da paisagem da área de estudo desta pesquisa, bem como os índices de acerto obtidos, os procedimentos e atributos utilizados nesta pesquisa podem apresentar bom desempenho também em outros biomas ou domínios de vegetação.

## Agradecimentos

Os autores agradecem à Agência Espacial Japonesa (JAXA), mais especificamente, ao *Kyoto & Carbon Protocol*, pela cessão das imagens do satélite ALOS-2/PALSAR-2.

## Referências

ANJOS, C. S.; ALMEIDA, C. M.; GALVÃO, L. S.; FILHO, C. R. S.; LACERDA, M. G.; PRATI, R. C. Análise do nível de legenda de classificação de áreas urbanas empregando imagens multiespectrais e hiperespectrais com os métodos árvore de decisão C4.5 e Floresta randômica. **Boletim de Ciências Geodésicas**, v. 23, n. 2, p. 371- 388, 2017.

BAATZ, M.; SCHÄPE, A. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: ANGEWANDTE GEOGRAPHISCHE INFORMATIONSVARBEITUNG XII. BEITRÄGE ZUM AGIT-SYMPOSIUM, Salzburg, 2000. **Anais...** Karlsruhe: Herbert Wichmann Verlag, p. 12–23.

BARGIEL, D.; HERRMANN, S. Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight TerraSAR-X data. **Remote Sensing**, v. 3, n. 5, p. 859–877, 2011.

BENZ, U. C.; HOFMANN, P.; WILLHAUCK, G.; LINGENFELDER, I.; HEYNEN, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS ready information. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 58, n. 3–4, p. 239–258, 2004.

BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BERGEN, K. M.; GOETZ, S. J.; DUBAYAH, R. O.; HENEERY, G. M.; HUNSAKER, C. T.; IMHOFF, M. L.; NELSON, R. F.; PARKER, G. G.; RADELOFF, V. C. Remote sensing

of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. **Journal of Geophysical Research**, v. 114, G00E06, p. 1–13, 2009.

BIRO, K.; PRADHAN, B.; SULIEMAN, H.; BUCHROITHNER, M. Exploitation of TerraSAR-X data for land use/land cover analysis using object-oriented classification approach in the African Sahel area, Sudan. **Journal of the Indian Society of Remote Sensing**, v. 41, n. 3, p. 539–553, 2013.

BISWAL, S.; GHOSH, A.; SHARMA, R.; JOSHI, P. K. Satellite data classification using open source support. **Journal of the Indian Society of Remote Sensing**, v. 41, n. 3, p. 523–530, 2013.

BITENCOURT, M. D.; MESQUITA Jr., H. N.; KUNTSCHIK, G.; ROCHA, H. R.; FURLEY, P. A. Cerrado vegetation study using optical and radar remote sensing: two Brazilian case studies. **Canadian Journal of Remote Sensing**, v. 33, n. 6, p. 468–480, 2007.

BOERNER, W.; MOTT, H.; LÜNENBERG, E.; LIVINGSTONE, C.; BRISCO, B.; BROWN, R. J.; PATERSON, J. S. Polarimetry in radar remote sensing: basic and applied concepts. In: HENDERSON, F. M.; LEWIS, A. J. **Manual of Remote Sensing: Principles and Applications of Imaging Radars** (3<sup>a</sup> ed.). Nova York: John Wiley & Sons, p. 271–356, 1998.

BOUCKAERT, R. R.; FRANK, E.; HALL, M. A.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. WEKA - experiences with a Java open-source project. **Journal of Machine Learning Research**, v. 11, p. 2533–2541, 2010.

BOUVET, A.; MERMÓZ, S.; TOAN, T. L.; VILLARD, L.; MATHIEU, R.; NAIDOO, L.; ASNER, G. P. An above-ground biomass map of African savannas and woodlands at 25m resolution derived from ALOS PALSAR. **Remote Sensing of Environment**, v. 206, p. 156–173, 2018.

BRAUN, A.; HOCHSCHILD, V. A SAR-based index for landscape changes in African savannas. **Remote Sensing**, v. 9, n. 4, Paper n. 359, 2017, 23 p.

CLOUDE, S. R.; POTTIER, E. A review of target decomposition theorems in radar polarimetry. **IEEE Transactions on Geoscience and Remote Sensing**, v. 34, n. 2, p. 498–518, 1996.

CLOUDE, S. R.; POTTIER, E. An entropy based classification scheme for land applications of polarimetric SAR. **IEEE Transactions on Geoscience Remote Sensing**, v. 35, n. 1, p. 68–78, 1997.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Boca Raton: CRC Press, 2009, 200 p.

CONGALTON, R. G.; GU, J.; YADAV, K.; THENKABAIL, P.; OZDOGAN, M. Global land cover mapping: a review and uncertainty analysis. **Remote Sensing**, v. 6, n. 12, p. 12070–12093, 2014.

CORREIA, A.H. **Projeto, desenvolvimento e avaliação de classificadores estatísticos pontuais e contextuais para imagens SAR polarimétricas**. 1998. 273 p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos/SP, 1998. (INPE-7178-TDI/679).

Disponível em:<<http://urlib.net/6qtX3pFwXQZ4PKzA/isJ7N>>.

DUARTE, R. M.; WOZNIAK, E.; RECONDO, C.; CABO, C.; MARQUÍNEZ, J.; FERNÁNDEZ, S. Estimation of surface roughness and stone cover in burnt soils using SAR images. **Catena**, v. 74, p. 264–272, 2008.

DU, P.; SAMAT, A.; WASKE, B.; LIU, S.; LI, Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 105, p. 38–53, 2015.

EILANDER, D.; ANNOR, F. O.; IANNINI, L.; VAN DE GIESEN, N. Remotely sensed monitoring of small reservoir dynamics: a bayesian approach. **Remote Sensing**, v. 6, p. 1191–1210, 2014.

ESA. European Space Agency. **Sentinel Application Platform (SNAP). SNAP 6.0 Help Guide**. ESA, 2018.

EVANS, T. L.; COSTA, M. Landcover classification of the lower Nhecolândia subregion of the Brazilian Pantanal wetlands using ALOS/PALSAR, RADARSAT-2 and ENVISAT/ASAR imagery. **Remote Sensing of Environment**, v. 128, p. 118–137, 2013.

FEDDEMA, J. J.; OLESON, K. W.; BONAN, G. B.; MEARNNS, L. O.; BUJA, L. E.; MEEHL, G. A.; WASHINGTON, W. M. The importance of land-cover change in simulating future climates, **Science**, v. 310, n. 5754, p. 1674–1678, 2005.

FOODY, G. M. Impacts of sample design for validation data on the accuracy of feedforward neural network classification. **Applied Sciences**, v. 7, Paper # 888, 2017, 15 p.

FREITAS, C. C.; SOLER, L. S.; SANT’ANNA, S. J. S.; DUTRA, L. V.; SANTOS, J. R.; MURA, J. C.; CORREIA, A. H. Land use and land cover mapping in the Brazilian Amazon using polarimetric airborne P-band SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, n. 10, p. 2956–2970, 2008.

FREEMAN, A.; DURDEN, S. L. A three-component scattering model for polarimetric SAR Data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 3, p. 963–973, 1998.

GAMBA, P.; ALDRIGHI, M. SAR data classification of urban areas by means of segmentation techniques and ancillary optical data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 5, n. 4, p. 1140–1148, 2012.

GOIÁS. **Geomorfologia do Estado de Goiás e Distrito Federal**. Goiânia: Secretaria de Indústria e Comércio/Superintendência de Geologia e Mineração, 2006, 128 p.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. 3<sup>a</sup> ed. Whaltan: Elsevier, 2012. 744 p.

HAARPAINNER, J.; BLANCO, D. F.; ENBLE, F.; DATTA, P.; MAZINGA, A.; SINGA, C.; MANE, L. Tropical forest remote sensing services for the Democratic Republic of Congo inside the EU FP7 'Recover' Project (Final Results 2000-2012). In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 36. Berlim. **Anais...** p. 397–402, 2015.

HELLMANN, M. P. SAR polarimetry tutorial. Rennes: University of Rennes, 2001. 109 p. Disponível em: <<http://epsilon.nought.de/>>. Acesso em: 01 fev. 2017.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Produção Agrícola Municipal – PAM**. 2017. Disponível em: <<https://ww2.ibge.gov.br/home/estatistica/economia/pam/2016/default.shtm>>. Acesso em: 10 ago. 2017.

INMET. Instituto Nacional de Meteorologia. **Dados meteorológicos da estação automática (A045) instalada na Estação Ecológica Águas Emendadas (período: abril a maio de 2016)**. Brasília: SADMET, 2018.

INPE. Instituto Nacional de Pesquisas Espaciais. **Projeto TerraClass Cerrado: mapeamento do uso e cobertura vegetal do Cerrado**. 2017. Disponível em: <<http://www.dpi.inpe.br/tccerrado/download.php>>. Acesso em: 01 jul. 2017.

JAXA. Japan Aerospace Exploration Agency. **ALOS-2 basic observation scenario** (1ª ed.). 2014. Disponível em: <[http://www.eorc.jaxa.jp/ALOS-2/en/obs/scenario/ALOS-2\\_Basic\\_Observation\\_Scenario\\_First-Ed\\_E\\_v00.pdf](http://www.eorc.jaxa.jp/ALOS-2/en/obs/scenario/ALOS-2_Basic_Observation_Scenario_First-Ed_E_v00.pdf)>. Acesso em: 10 ago. 2015.

JAXA. Japan Aerospace Exploration Agency. **Calibration results of ALOS-2/PALSAR-2 JAXA Standard Products**. 2018. Disponível em: <[https://www.eorc.jaxa.jp/ALOS-2/en/calval/calval\\_index.htm](https://www.eorc.jaxa.jp/ALOS-2/en/calval/calval_index.htm)>. Acesso em: 15 jan. 2018.

KASPERCZUK, A.; DARDZIŃSKA, A. Comparative evaluation of the different data mining techniques used for the medical database. **Acta Mechanica et Automatica**, v. 10, n. 3, p. 233–238, 2016.

KIM, Y.; VAN ZYL, J. J. A time-series approach to estimate soil moisture using polarimetric radar data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 47, n. 8, p. 2519–2527, 2009.

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA – geographic data mining analyst. **Computers & Geosciences**, v. 57, p. 133–145, 2013.

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

LEE, J.; POTTIER, E. **Polarimetric radar imaging from basics to applications**. Nova York: CRC Press, 2009. 398 p.

MAPBIOMAS. **Mapeamento anual da cobertura e uso do solo no Brasil**. 2017. Disponível em: <<http://mapbiomas.org>>. Acesso em: 15 jun. 2017.

MEINEL, G.; NEUBERT, M. A. A comparison of segmentation programs for high resolution remote sensing data. **International Archives of Photogrammetry and Remote Sensing**, v. 23, Part B, p. 1097–1105, 2004.

MILES, L.; KAPOS, V. Reducing greenhouse gas emissions from deforestation and forest degradation: global land-use implications. **Science**, v. 320, n. 5882, p. 1454–1455, 2008.

MITCHARD, E. T. A.; SAATCHI, S. S.; WHITE, L. J. T.; ABERNETHY, K. A.; JEFFERY, K. J.; LEWIS, S. L.; COLLINS, M.; LEFSKY, M. A.; LEAL, M. E.; WOODHOUSE, I. H.; MEIR, P. Mapping tropical forest biomass with radar and spaceborne LIDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. **Biogeosciences**, v. 9, p. 179–191, 2012.

NEXTGIS. **QuickMapServices plugin v. 0.19.10.1**.

Disponível em: <[http://plugins.qgis.org/plugins/quick\\_map\\_services/](http://plugins.qgis.org/plugins/quick_map_services/)>. Acesso em: 16 dez. 2018.

ODIPO, V. O.; NICKLESS, A.; BERGER, C.; BAADE, J.; URBAZAEV, M.; WALTHER, C.; SCHMULLIUS, C. Assessment of aboveground woody biomass dynamics using terrestrial laser scanner and L-band ALOS PALSAR data in South African savanna. **Forests**, v. 7, Paper # 294, 2016, 24 p.

OK, A. O.; AKAR, O.; GUNGOR, O. Evaluation of random forest method for agricultural crop classification. **European Journal of Remote Sensing**, v. 45, p. 421–432, 2012.

OUCHI, K. Recent trend and advance of synthetic aperture radar with selected topics. **Remote Sensing**, v. 5, n. 2, p. 716–807, 2013.

PASSO, D.; BIAS, E.; BRITES, R.; COSTA, G. Comparison of the accuracy of classifications generated by interimage and by INTERIMAGE integrated with data mining. **South-Eastern European Journal of Earth Observation and Geomatics**, v. 3, p. 93–98, 2014.

PIATETSKY-SHAPIRO, G.; FAYYAD, U. An introduction to SIGKDD and a reflection on the term ‘data mining’. **SIGKDD Explorations**, v. 13, n. 2, p. 103–104, 2012.

QI, Z.; YEH, A. G. O.; LI, X.; LIN, Z. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. **Remote Sensing of Environment**, v. 118, p. 21–39, 2012.

PCI GEOMATICS. **PCI Geomatica 2017 Help Guide**. PCI Geomatics Enterprises, 2018.

POPE, K. O.; REY-BENAYAS, J. M.; PARIS, J. F. Radar remote sensing of forest and wetland ecosystems in the central American tropics. **Remote Sensing of Environment**, v. 48, p. 205–219, 1994.

RAHMAN, M. M.; MORAN, M. S.; THOMA, D. P.; BRYANT, R.; COLLINS, C. D. H.; JACKSON, T.; ORR, B. J.; TISCHLER, M. Mapping surface roughness and soil moisture using multi-angle radar imagery without ancillary data. **Remote Sensing of Environment**, v. 112, p. 391–402, 2008.

REYNOLDS, J.; WESSON, K.; DESBIEZ, A. L. J.; OCHOA-QUINTERO, J. M.; LEIMGRUBER, P. Using remote sensing and random forest to assess the conservation status of critical Cerrado habitats in Mato Grosso do Sul, Brazil. **Land**, v. 5, Paper 5020012, 12 p., 2016.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. (Eds.). **Cerrado: Ecologia e Flora**, Planaltina: Embrapa Cerrados, 2008, p. 151–199.

RICHARDS, J. A. **Remote Sensing with Imaging Radar**. Berlim: Springer, 2009, 361 p.

ROSENQVIST, A.; MILNE, A.; LUCAS, R.; IMHOFF, M.; DOBSON, C. A review of remote sensing technology in support of the Kyoto Protocol. **Environmental Science & Policy**, v. 6, p. 441–455, 2003.

ROSENQVIST, A.; SHIMADA, M.; SUZUKI, S.; OHGUSHI, F.; TADONO, T.; WATANABE, M.; TSUZUKU, K.; WATANABE, T.; KAMIJO, S.; AOKI, E. Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. **Remote Sensing of Environment**, v. 155, p. 3–12, 2014.

SANO, E. E.; PINHEIRO, G. C. C.; MENESES, P. R. Assessing JERS-1 synthetic aperture radar data for vegetation mapping in the Brazilian savanna. **Journal of Remote Sensing Society of Japan**, v. 21, p. 158–167, 2001.

SANO, E. E.; FERREIRA, L. G.; HUETE, A. R. Synthetic aperture radar (L-band) and optical vegetation indices for discriminating the Brazilian savanna physiognomies: A comparative analysis. **Earth Interactions**, v. 9, Paper no. 15, p. 1–15, 2005.

SANO, E. E.; SANTOS, E. M.; MENESES, P. R. Análise de imagens do satélite ALOS PALSAR para o mapeamento de uso e cobertura da terra do Distrito Federal. **Geociências**, v. 28, n. 4, p. 441–451, 2009.

SHARMA, R.; GHOSH, A.; JOSHI, P. K. Decision tree approach for classification of remotely sensed satellite data using open source support. **Journal of Earth System Science**, v. 122, n. 5, p. 1237–1247, 2013.

SINGH, G.; VENKATARAMAN, G.; YAMAGUCHI, Y.; PARK, S. Capability assessment of fully polarimetric ALOS-PALSAR data for discriminating wet snow from other scattering types in mountainous regions. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 2, p. 1177–1196, 2014.

SHIMADA, M.; WATANABE, M.; MOTOOKA, T.; KANKAKU, Y. PALSAR-2 polarimetric performance and the simulation study using the PI-SAR-L2. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS 2013), Melbourne, Austrália, **Anais...** p. 2309–2312, 2013.

SHIRAISHI, T.; MOTOHKA, T.; THAPA, R. B.; WATANABE, M.; SHIMADA, M. Comparative assessment of supervised classifiers for land use–land cover classification in a

tropical region using time-series PALSAR mosaic data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 4, p. 1186–1199, 2014.

SILVA, M. P. S.; CÂMARA, G.; ESCADA, M. I. S.; SOUZA, R. C. M. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. **International Journal of Remote Sensing**, v. 29, n. 16, p. 4803–4822, 2008.

SUZUKI, S.; KANKAKU, Y.; SHIMADA, M. ALOS-2 acquisition strategy. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS 2013), Melbourne, Austrália. **Anais...** p. 2412-2415, 2013.

SYMEONAKIS, E.; HIGGINBOTTOM, T. P.; PETROULAKI, K.; RABE, A. Optimisation of savannah land cover characterisation with optical and SAR data. **Remote Sensing**, v. 10, Paper 499, 2018, 18 p.

TOLLERUD, H. J.; FANTLE, M. S. The temporal variability of centimeter-scale surface roughness in a playa dust source: synthetic aperture radar investigation of playa surface dynamics. **Remote Sensing of Environment**, v. 154, p. 285–297, 2014.

TOPOUZELIS, K.; PSYLLOS, A. Oil spill feature selection and classification using decision tree forest on SAR image data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 68, p. 135–143, 2012.

TRIMBLE. **eCognition Developer 8.7: Reference Book**. Munique: Trimble, 2011.

TSAI, H. Global data mining: an empirical study of current trends, future forecasts and technology diffusions. **Expert Systems with Applications**, v. 39, p. 8172–8181, 2012.

TSAI, H. Knowledge management vs. data mining: research trend, forecast and citation approach. **Expert Systems with Applications**, v. 40, p. 3160–3173, 2013.

UHLMANN, S.; KIRANYAZ, S. Integrating color features in polarimetric SAR image classification. **IEEE Transactions on Geosciences and Remote Sensing**, v. 52, p. 2197–2216, 2014.

USGS. United States Geological Survey. **Global Visualization (GloVis) Viewer**. Reston: USGS, 2017. Disponível em: < <https://glovis.usgs.gov/>>. Acesso em: 05 fev. 2017.

URBAZAEV, M.; THIEL, C.; MATHIEU, R.; NAIDOO, L.; LEVICK, S. R.; SMIT, I. P. J.; ASNER, G. P.; SCHMULLIUS, C. Assessment of the mapping of fractional woody cover in southern African savannas using multi-temporal and polarimetric ALOS PALSAR L-band images. **Remote Sensing of Environment**, v. 166, p. 138–153, 2015.

VAN ZYL, J. J. Unsupervised classification of scattering behavior using radar polarimetry data. **IEEE Transactions on Geoscience Remote Sensing**, v. 27, n. 1, p. 36–45, 1989.

WASKE, B.; BRAUN, M. Classifier ensembles for land cover mapping using multitemporal SAR imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, p. 450–457, 2009.

WENTZ, E. A.; NELSON, D.; RAHMAN, A.; STEFANOV, W. L.; ROY, S. S. Expert system classification of urban land use/cover for Delhi, India. **International Journal of Remote Sensing**, v. 29, n. 15–16, p. 4405–4427, 2008.

WESTRA, T.; WULF, R.; COILLIE, F.; CRABBE, S. Optimal ENVISAT advanced synthetic aperture radar image parameters for mapping and monitoring Sahelian floodplains. **Journal of Applied Remote Sensing**, v. 4, n. 1, Paper 043511, 17 p., 2010.

WITTEN, I. H.; FRANK, E. **Data Mining: Practical Machine Learning. Tools and Techniques**. (2<sup>a</sup> ed.). San Francisco: Morgan Kaufmann, 2005, 560 p.

YAMAGUCHI, Y.; MORIYAMA, T.; ISHIDO, M.; YAMADA, H. Four-component scattering model for polarimetric SAR image decomposition. **IEEE Transactions on Geoscience Remote Sensing**, v. 43, n. 8, p. 1699–1706, 2005.

YOMMY, A. S.; LIU, R.; WU, S. SAR image despeckling using refined Lee filter. In: INTERNATIONAL CONFERENCE ON INTELLIGENT HUMAN-MACHINE SYSTEMS AND CYBERNETICS, 7. Hangzhou, China. **Anais...** p. 260–265, 2015.

ZHANG, H.; LIN, H.; LI, Y. Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification. **IEEE Geoscience and Remote Sensing Letters**, v. 12, n. 5, p. 1061–1065, 2015.



## CONSIDERAÇÕES FINAIS

Nesta pesquisa, foi verificada a viabilidade de um conjunto de técnicas de mineração de dados, métodos SAR polarimétricos e de imagens do sensor ALOS-2/PALSAR-2, para fins de identificação automática de padrões de uso e cobertura de terras no Cerrado. Os experimentos realizados foram divididos em dois artigos apresentados nos Capítulos 3 e 4. Ambos os artigos apresentaram contribuições inovadoras e relevantes no âmbito da aplicação de imagens SAR para o Cerrado. O artigo do Capítulo 3 focou principalmente nas técnicas de mineração de dados e destacou as limitações e potencialidades dos algoritmos de aprendizado de máquina AD J48, RF e MLP. Por outro lado, o artigo do Capítulo 4 apresentou uma análise mais completa das técnicas de classificação e dos dados SAR multipolarimétricos, explorando a teoria de decomposição de alvos, parâmetros polarimétricos, segmentação de imagens, cinco algoritmos de aprendizado de máquina (NB, AD J48, RF, MLP e SVM), além do classificador supervisionado PW.

Ampla discussão foi realizada sobre os resultados dos dois artigos mencionados, podendo-se concluir que a hipótese apresentada no Capítulo 1 foi verdadeira e pôde ser aceita. Os resultados dos Artigos 1 e 2 demonstraram a viabilidade do método apresentado nesta tese, reforçando também a necessidade de maior exploração de dados SAR obtidos na banda L para estudo da vegetação de Cerrado, haja vista a vocação/habilidade desse tipo de sensor na extração de parâmetros indicativos da estrutura vegetal e, por conseguinte, do volume de biomassa acima do solo.

Como propostas de trabalhos futuros no Cerrado, destacam-se:

- A utilização do fluxo metodológico proposto para dados SAR orbitais multisensores e multifrequência com banda L e banda C (ALOS-2/PALSAR-2 e SENTINEL-1, por exemplo);
- O uso de séries temporais de imagens SAR polarimétricas;
- O emprego, se possível, de técnicas PolInSAR para classificação, unindo assim métodos e atributos avançados tanto de polarimetria quanto de interferometria;
- Uso combinado de levantamentos LIDAR para construção de modelos de estimativa de biomassa, relacionando assim parâmetros SAR ( $\sigma^0$ , por exemplo) com aspectos físicos da

vegetação. Os modelos calibrados podem ser empregados na estimativa da biomassa e também na classificação das fisionomias de vegetação.

**APÊNDICE A: RELAÇÃO DE ATRIBUTOS UTILIZADOS NAS CLASSIFICAÇÕES REALIZADAS NO WEKA 3.8**

| <b>Identificador</b> | <b>NOME DO ATRIBUTO<br/>(POR SEGMENTO/OBJETO)</b> |
|----------------------|---|
| 1                    | Mínimo <i>pixel</i> VV_SIGMA0                     |
| 2                    | Mínimo <i>pixel</i> VVVH_RAZAO_dB                 |
| 3                    | Mínimo <i>pixel</i> VSI_dB                        |
| 4                    | Mínimo <i>pixel</i> VH_SIGMA0                     |
| 5                    | Mínimo <i>pixel</i> DBL_YAMAGUCHI                 |
| 6                    | Mínimo <i>pixel</i> DBL_VANZYL                    |
| 7                    | Mínimo <i>pixel</i> DBL_FREEMAN-DURDEN            |
| 8                    | Mínimo <i>pixel</i> RVI_dB                        |
| 9                    | Mínimo <i>pixel</i> RFDI_dB                       |
| 10                   | Mínimo <i>pixel</i> HV_SIGMA0                     |
| 11                   | Mínimo <i>pixel</i> HLX_YAMAGUCHI                 |
| 12                   | Mínimo <i>pixel</i> HH_SIGMA0                     |
| 13                   | Mínimo <i>pixel</i> HHVV_RAZAO_dB                 |
| 14                   | Mínimo <i>pixel</i> HHHV_RAZAO_dB                 |
| 15                   | Mínimo <i>pixel</i> VOL_YAMAGUCHI                 |
| 16                   | Mínimo <i>pixel</i> VOL_VANZYL                    |
| 17                   | Mínimo <i>pixel</i> VOL_FREEMAN-DURDEN            |
| 18                   | Mínimo <i>pixel</i> ENTROPIA                      |
| 19                   | Mínimo <i>pixel</i> CSI_dB                        |
| 20                   | Mínimo <i>pixel</i> SURF_YAMAGUCHI                |
| 21                   | Mínimo <i>pixel</i> SURF_VANZYL                   |
| 22                   | Mínimo <i>pixel</i> SURF_FREEMAN-DURDEN           |
| 23                   | Mínimo <i>pixel</i> BMI_dB                        |
| 24                   | Mínimo <i>pixel</i> ANISOTROPIA                   |
| 25                   | Mínimo <i>pixel</i> ALFA                          |
| 26                   | Desvio padrão VV_SIGMA0                           |
| 27                   | Desvio padrão VVVH_RAZAO_dB                       |
| 28                   | Desvio padrão VSI_dB                              |
| 29                   | Desvio padrão VH_SIGMA0                           |
| 30                   | Desvio padrão DBL_YAMAGUCHI                       |
| 31                   | Desvio padrão DBL_VANZYL                          |
| 32                   | Desvio padrão DBL_FREEMAN-DURDEN                  |
| 33                   | Desvio padrão RVI_dB                              |
| 34                   | Desvio padrão RFDI_dB                             |

| <b>Identificador</b> | <b>NOME DO ATRIBUTO<br/>(POR SEGMENTO/OBJETO)</b> |
|----------------------|---|
| 35                   | Desvio padrão HV_SIGMA0                           |
| 36                   | Desvio padrão HLX_YAMAGUCHI                       |
| 37                   | Desvio padrão HH_SIGMA0                           |
| 38                   | Desvio padrão HHVV_RAZAO_dB                       |
| 39                   | Desvio padrão HHHV_RAZAO_dB                       |
| 40                   | Desvio padrão VOL_YAMAGUCHI                       |
| 41                   | Desvio padrão VOL_VANZYL                          |
| 42                   | Desvio padrão VOL_FREEMAN-DURDEN                  |
| 43                   | Desvio padrão ENTROPIA                            |
| 44                   | Desvio padrão CSI_dB                              |
| 45                   | Desvio padrão SURF_YAMAGUCHI                      |
| 46                   | Desvio padrão SURF_VANZYL                         |
| 47                   | Desvio padrão SURF_FREEMAN-DURDEN                 |
| 48                   | Desvio padrão BMI_dB                              |
| 49                   | Desvio padrão ANISOTROPIA                         |
| 50                   | Desvio padrão ALFA                                |
| 51                   | Média VV_SIGMA0                                   |
| 52                   | Média VVVH_RAZAO_dB                               |
| 53                   | Média VSI_dB                                      |
| 54                   | Média VH_SIGMA0                                   |
| 55                   | Média DBL_YAMAGUCHI                               |
| 56                   | Média DBL_VANZYL                                  |
| 57                   | Média DBL_FREEMAN-DURDEN                          |
| 58                   | Média RVI_dB                                      |
| 59                   | Média RFDI_dB                                     |
| 60                   | Média HV_SIGMA0                                   |
| 61                   | Média HLX_YAMAGUCHI                               |
| 62                   | Média HH_SIGMA0                                   |
| 63                   | Média HHVV_RAZAO_dB                               |
| 64                   | Média HHHV_RAZAO_dB                               |
| 65                   | Média VOL_YAMAGUCHI                               |
| 66                   | Média VOL_VANZYL                                  |
| 67                   | Média VOL_FREEMAN-DURDEN                          |
| 68                   | Média ENTROPIA                                    |
| 69                   | Média CSI_dB                                      |
| 70                   | Média SURF_YAMAGUCHI                              |
| 71                   | Média SURF_VANZYL                                 |
| 72                   | Média SURF_FREEMAN-DURDEN                         |
| 73                   | Média BMI_dB                                      |
| 74                   | Média ANISOTROPIA                                 |
| 75                   | Média ALFA  |

| <b>Identificador</b> | <b>NOME DO ATRIBUTO<br/>(POR SEGMENTO/OBJETO)</b> |
|----------------------|---|
| 76                   | Máximo <i>pixel</i> VV_SIGMA0                     |
| 77                   | Máximo <i>pixel</i> VVVH_RAZAO_dB                 |
| 78                   | Máximo <i>pixel</i> VSI_dB                        |
| 79                   | Máximo <i>pixel</i> VH_SIGMA0                     |
| 80                   | Máximo <i>pixel</i> DBL_YAMAGUCHI                 |
| 81                   | Máximo <i>pixel</i> DBL_VANZYL                    |
| 82                   | Máximo <i>pixel</i> DBL_FREEMAN-DURDEN            |
| 83                   | Máximo <i>pixel</i> RVI_dB                        |
| 84                   | Máximo <i>pixel</i> RFDI_dB                       |
| 85                   | Máximo <i>pixel</i> HV_SIGMA0                     |
| 86                   | Máximo <i>pixel</i> HLX_YAMAGUCHI                 |
| 87                   | Máximo <i>pixel</i> HH_SIGMA0                     |
| 88                   | Máximo <i>pixel</i> HHVV_RAZAO_dB                 |
| 89                   | Máximo <i>pixel</i> HHHV_RAZAO_dB                 |
| 90                   | Máximo <i>pixel</i> VOL_YAMAGUCHI                 |
| 91                   | Máximo <i>pixel</i> VOL_VANZYL                    |
| 92                   | Máximo <i>pixel</i> VOL_FREEMAN-DURDEN            |
| 93                   | Máximo <i>pixel</i> ENTROPIA                      |
| 94                   | Máximo <i>pixel</i> CSI_dB                        |
| 95                   | Máximo <i>pixel</i> SURF_YAMAGUCHI                |
| 96                   | Máximo <i>pixel</i> SURF_VANZYL                   |
| 97                   | Máximo <i>pixel</i> SURF_FREEMAN-DURDEN           |
| 98                   | Máximo <i>pixel</i> BMI_dB                        |
| 99                   | Máximo <i>pixel</i> ANISOTROPIA                   |
| 100                  | Máximo <i>pixel</i> ALFA                          |
| 101                  | <i>Skewness</i> VV_SIGMA0                         |
| 102                  | <i>Skewness</i> VVVH_RAZAO_dB                     |
| 103                  | <i>Skewness</i> VSI_dB                            |
| 104                  | <i>Skewness</i> VH_SIGMA0                         |
| 105                  | <i>Skewness</i> DBL_YAMAGUCHI                     |
| 106                  | <i>Skewness</i> DBL_VANZYL                        |
| 107                  | <i>Skewness</i> DBL_FREEMAN-DURDEN                |
| 108                  | <i>Skewness</i> RVI_dB                            |
| 109                  | <i>Skewness</i> RFDI_dB                           |
| 110                  | <i>Skewness</i> HV_SIGMA0                         |
| 111                  | <i>Skewness</i> HLX_YAMAGUCHI                     |
| 112                  | <i>Skewness</i> HH_SIGMA0                         |
| 113                  | <i>Skewness</i> HHVV_RAZAO_dB                     |
| 114                  | <i>Skewness</i> HHHV_RAZAO_dB                     |
| 115                  | <i>Skewness</i> VOL_YAMAGUCHI                     |
| 116                  | <i>Skewness</i> VOL_VANZYL                        |

| <b>Identificador</b> | <b>NOME DO ATRIBUTO<br/>(POR SEGMENTO/OBJETO)</b> |
|----------------------|---|
| 117                  | <i>Skewness</i> VOL_FREEMAN-DURDEN                |
| 118                  | <i>Skewness</i> ENTROPIA                          |
| 119                  | <i>Skewness</i> CSI_dB                            |
| 120                  | <i>Skewness</i> SURF_YAMAGUCHI                    |
| 121                  | <i>Skewness</i> SURF_VANZYL                       |
| 122                  | <i>Skewness</i> SURF_FREEMAN-DURDEN               |
| 123                  | <i>Skewness</i> BMI_dB                            |
| 124                  | <i>Skewness</i> ANISOTROPIA                       |
| 125                  | <i>Skewness</i> ALFA                              |

**APÊNDICE B: ÁRVORE DE DECISÃO GERADA PELO CLASSIFICADOR J48  
(LEGENDA COM NOVE CLASSES)**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: TREINAMENTO\_200POLIGONOS

Instances: 1275

Attributes: 126

[list of attributes omitted]

Test mode: user supplied test set: size unknown (reading incrementally)

=== Classifier model (full training set) ===

J48 pruned tree

-----

```

Média VOL_FREEMAN-DURDEN <= 107,650056
| Média DBL_FREEMAN-DURDEN <= 93,044635
| | Média VH_SIGMA0 <= -23,174936
| | | Média VOL_VANZYL <= 94,763599
| | | | Média ALFA <= 37,526453
| | | | | Mínimo pixel CSI_dB <= -4,226604
| | | | | | Mínimo pixel HH_SIGMA0 <= -29,692179: Corpodaguacontinental (2,0)
| | | | | | Mínimo pixel HH_SIGMA0 > -29,692179
| | | | | | Média HHVV_Ratio_dB <= -0,371261
| | | | | | | Mínimo pixel RFDI_dB <= -21,984598: Pastagens (5,0)
| | | | | | | Mínimo pixel RFDI_dB > -21,984598: Soloexposto/palhada (6,0/1,0)
| | | | | | | Média HHVV_Ratio_dB > -0,371261
| | | | | | | | Skewness HV_SIGMA0 <= 2,116835
| | | | | | | | | Mínimo pixel VH_SIGMA0 <= -34,140015: Soloexposto/palhada (154,0/1,0)
| | | | | | | | | Mínimo pixel VH_SIGMA0 > -34,140015
| | | | | | | | | Média VOL_YAMAGUCHI <= 92,146462: Campolimpo (2,0)
| | | | | | | | | Média VOL_YAMAGUCHI > 92,146462: Soloexposto/palhada (11,0)

```

| | | | | | | Skewness HV\_SIGMA0 > 2,116835: Pastagens (3,0/1,0)  
 | | | | | Mínimo pixel CSI\_dB > -4,226604: Corpodaguacontinental (4,0)  
 | | | | Média ALFA > 37,526453: Corpodaguacontinental (19,0)  
 | | | Média VOL\_VANZYL > 94,763599  
 | | | | Skewness VOL\_YAMAGUCHI <= -0,512503  
 | | | | | Skewness HV\_SIGMA0 <= 0,232079: Culturastemporarias (8,0)  
 | | | | | Skewness HV\_SIGMA0 > 0,232079: Pastagens (3,0/1,0)  
 | | | | Skewness VOL\_YAMAGUCHI > -0,512503  
 | | | | | Desvio padrão RFDI\_dB <= 0,901556: Soloexposto/palhada (9,0)  
 | | | | | Desvio padrão RFDI\_dB > 0,901556  
 | | | | | | Máximo pixel DBL\_VANZYL <= 100,938652  
 | | | | | | Mínimo pixel HH\_SIGMA0 <= -24,459044  
 | | | | | | | Máximo pixel BMI\_dB <= -11,939823: Soloexposto/palhada (3,0)  
 | | | | | | | Máximo pixel BMI\_dB > -11,939823: Pastagens (4,0)  
 | | | | | | | Mínimo pixel HH\_SIGMA0 > -24,459044  
 | | | | | | | | Desvio padrão DBL\_FREEMAN-DURDEN <= 1,360454: Soloexposto/palhada (4,0/1,0)  
 | | | | | | | | Desvio padrão DBL\_FREEMAN-DURDEN > 1,360454: Campolimpo (14,0)  
 | | | | | | | Máximo pixel DBL\_VANZYL > 100,938652  
 | | | | | | | Máximo pixel RFDI\_dB <= -1,453654: Soloexposto/palhada (7,0/1,0)  
 | | | | | | | Máximo pixel RFDI\_dB > -1,453654  
 | | | | | | | Máximo pixel HLX\_YAMAGUCHI <= 99,40979  
 | | | | | | | | Skewness HV\_SIGMA0 <= 0,604152  
 | | | | | | | | | Mínimo pixel VV\_SIGMA0 <= -22,439281: Campolimpo (5,0)  
 | | | | | | | | | Mínimo pixel VV\_SIGMA0 > -22,439281: Soloexposto/palhada (2,0)  
 | | | | | | | | | Skewness HV\_SIGMA0 > 0,604152: Pastagens (7,0/1,0)  
 | | | | | | | | | Máximo pixel HLX\_YAMAGUCHI > 99,40979  
 | | | | | | | | | Máximo pixel DBL\_YAMAGUCHI <= 99,976173  
 | | | | | | | | | | Mínimo pixel ENTROPIA <= 0,26134: Soloexposto/palhada (3,0/1,0)  
 | | | | | | | | | | Mínimo pixel ENTROPIA > 0,26134: Pastagens (3,0)  
 | | | | | | | | | | Máximo pixel DBL\_YAMAGUCHI > 99,976173: Pastagens (170,0/2,0)  
 | | Média VH\_SIGMA0 > -23,174936  
 | | | Desvio padrão HV\_SIGMA0 <= 1,880668: Culturastemporarias (5,0/1,0)  
 | | | Desvio padrão HV\_SIGMA0 > 1,880668  
 | | | | Skewness VH\_SIGMA0 <= 0,651136  
 | | | | | Mean\_HHHV\_Ratio\_dB <= 3,992983:  
 Cerradoralo/Campocerradosujo/Cerradotipicoantropizado (186,0)



| | | | | Mean\_HHHV\_Ratio\_dB > 3,992983  
 | | | | | Mean\_RFDI\_dB <= -3,869343: Pastagens (3,0/1,0)  
 | | | | | Mean\_RFDI\_dB > -3,869343: Cerradoralo/Campocerradosujo/Cerradotipicoantropizado  
 (5,0)  
 | | | | Skewness VH\_SIGMA0 > 0,651136  
 | | | | | Desvio padrão SURF\_YAMAGUCHI <= 4,485622: Pastagens (5,0)  
 | | | | | Desvio padrão SURF\_YAMAGUCHI > 4,485622:  
 Cerradoralo/Campocerradosujo/Cerradotipicoantropizado (2,0/1,0)  
 | Média DBL\_FREEMAN-DURDEN > 93,044635  
 | | Desvio padrão VOL\_FREEMAN-DURDEN <= 2,511051  
 | | | Média VOL\_YAMAGUCHI <= 103,548319: Culturastemporarias (177,0/2,0)  
 | | | Média VOL\_YAMAGUCHI > 103,548319  
 | | | | Desvio padrão BMI\_dB <= 0,690225: Culturastemporarias (8,0/1,0)  
 | | | | Desvio padrão BMI\_dB > 0,690225: Silvicultura (9,0/1,0)  
 | | Desvio padrão VOL\_FREEMAN-DURDEN > 2,511051  
 | | | Mínimo pixel VV\_SIGMA0 <= -19,586504: Pastagens (2,0)  
 | | | Mínimo pixel VV\_SIGMA0 > -19,586504:  
 Cerradoralo/Campocerradosujo/Cerradotipicoantropizado (2,0/1,0)  
 Média VOL\_FREEMAN-DURDEN > 107,650056  
 | Média ANISOTROPIA <= 0,380778  
 | | Média DBL\_YAMAGUCHI <= 96,068385  
 | | | Média ALFA <= 40,1854: Cerradoralo/Campocerradosujo/Cerradotipicoantropizado (3,0)  
 | | | Média ALFA > 40,1854  
 | | | | Média VH\_SIGMA0 <= -16,358512  
 | | | | | Mínimo pixel VSI\_dB <= -8,251739: Matagaleria/Cerradotipico/Cerradodense (3,0)  
 | | | | | Mínimo pixel VSI\_dB > -8,251739:  
 Cerradoralo/Campocerradosujo/Cerradotipicoantropizado (2,0)  
 | | | | Média VH\_SIGMA0 > -16,358512: Matagaleria/Cerradotipico/Cerradodense (197,0/1,0)  
 | | Média DBL\_YAMAGUCHI > 96,068385: Silvicultura (14,0)  
 | Média ANISOTROPIA > 0,380778  
 | | Desvio padrão HHHV\_Ratio\_dB <= 1,081676: Culturastemporarias (5,0)  
 | | Desvio padrão HHHV\_Ratio\_dB > 1,081676: Areascomedificacoes (199,0)

Number of Leaves : 41

Size of the tree : 81

Time taken to build model: 0,39 seconds

**APÊNDICE C: ATRIBUTOS EMPREGADOS PELA ÁRVORE DE DECISÃO J48  
(APÊNDICE B)**

| <b>Identificador</b> | <b>NOMES DOS ATRIBUTOS<br/>(POR SEGMENTO/OBJETO)</b> |
|----------------------|--|
| 1                    | Mínimo <i>pixel</i> VV_SIGMA0                        |
| 2                    | Mínimo <i>pixel</i> VSI_dB                           |
| 3                    | Mínimo <i>pixel</i> VH_SIGMA0                        |
| 4                    | Mínimo <i>pixel</i> RFDI_dB                          |
| 5                    | Mínimo <i>pixel</i> HH_SIGMA0                        |
| 6                    | Mínimo <i>pixel</i> ENTROPIA                         |
| 7                    | Mínimo <i>pixel</i> CSI_dB                           |
| 8                    | Desvio padrão DBL_FREEMAN-DURDEN                     |
| 9                    | Desvio padrão RFDI_dB                                |
| 10                   | Desvio padrão HV_SIGMA0                              |
| 11                   | Desvio padrão HHHV_RAZAO_dB                          |
| 12                   | Desvio padrão VOL_FREEMAN-DURDEN                     |
| 13                   | Desvio padrão SURF_YAMAGUCHI                         |
| 14                   | Desvio padrão BMI_dB                                 |
| 15                   | Média VH_SIGMA0                                      |
| 16                   | Média DBL_YAMAGUCHI                                  |
| 17                   | Média DBL_FREEMAN-DURDEN                             |
| 18                   | Média RFDI_dB  |
| 19                   | Média HHVV_RAZAO_dB                                  |
| 20                   | Média HHHV_RAZAO_dB                                  |
| 21                   | Média VOL_YAMAGUCHI                                  |
| 22                   | Média VOL_VANZYL                                     |
| 23                   | Média VOL_FREEMAN-DURDEN                             |
| 24                   | Média ANISOTROPIA                                    |
| 25                   | Média ALFA   |
| 26                   | Máximo <i>pixel</i> DBL_YAMAGUCHI                    |
| 27                   | Máximo <i>pixel</i> DBL_VANZYL                       |
| 28                   | Máximo <i>pixel</i> RFDI_dB                          |
| 29                   | Máximo <i>pixel</i> HLX_YAMAGUCHI                    |
| 30                   | Máximo <i>pixel</i> BMI_dB                           |
| 31                   | <i>Skewness</i> VH_SIGMA0                            |
| 32                   | <i>Skewness</i> HV_SIGMA0                            |
| 33                   | <i>Skewness</i> VOL_YAMAGUCHI                        |

**APÊNDICE D: RELAÇÃO DE ATRIBUTOS OBTIDOS UTILIZANDO O PROCEDIMENTO DE SELEÇÃO DO ALGORITMO *INFOGAIN* (LEGENDA COM NOVE CLASSES)**

| <b>Identificador</b> | <b>ATRIBUTOS</b>                          | <b>VALOR DE GANHO</b> |
|----------------------|---|-----------------------|
| 1                    | Média HLX_YAMAGUCHI                       | 1,8359                |
| 2                    | Média VOL_VANZYL                          | 1,8354                |
| 3                    | Média VOL_FREEMAN-DURDEN                  | 1,8301                |
| 4                    | Média VH_SIGMA0                           | 1,8203                |
| 5                    | Média HV_SIGMA0                           | 1,8097                |
| 6                    | Média VOL_YAMAGUCHI                       | 1,8036                |
| 7                    | Desvio padrão<br>HLX_YAMAGUCHI            | 1,7509                |
| 8                    | <i>Skewness</i><br>HLX_YAMAGUCHI          | 1,6276                |
| 9                    | Média BMI_dB                              | 1,6247                |
| 10                   | Média VV_SIGMA0                           | 1,5921                |
| 11                   | Média HH_SIGMA0                           | 1,5896                |
| 12                   | Máximo <i>pixel</i> VH_SIGMA0             | 1,4666                |
| 13                   | Máximo <i>pixel</i> HV_SIGMA0             | 1,4584                |
| 14                   | Média DBL_YAMAGUCHI                       | 1,4353                |
| 15                   | Máximo <i>pixel</i><br>VOL_YAMAGUCHI      | 1,4337                |
| 16                   | Máximo <i>pixel</i><br>VOL_VANZYL         | 1,4161                |
| 17                   | Máximo <i>pixel</i><br>VOL_FREEMAN-DURDEN | 1,4032                |
| 18                   | Desvio padrão<br>DBL_FREEMAN-DURDEN       | 1,3819                |
| 19                   | Mínimo <i>pixel</i><br>VOL_FREEMAN-DURDEN | 1,3742                |

| <b>Identificador</b> | <b>ATRIBUTOS</b>                       | <b>VALOR DE GANHO</b> |
|----------------------|--|-----------------------|
| 20                   | Mínimo <i>pixel</i> VOL_VANZYL         | 1,3484                |
| 21                   | Desvio padrão<br>DBL_VANZYL            | 1,3389                |
| 22                   | Máximo <i>pixel</i><br>HLX_YAMAGUCHI   | 1,3289                |
| 23                   | Desvio padrão<br>SURF_YAMAGUCHI        | 1,319                 |
| 24                   | Mínimo <i>pixel</i> VH_SIGMA0          | 1,2936                |
| 25                   | Média DBL_VANZYL                       | 1,2911                |
| 26                   | Mínimo <i>pixel</i> HV_SIGMA0          | 1,2853                |
| 27                   | Média ALFA                             | 1,2801                |
| 28                   | Desvio padrão<br>SURF_FREEMAN-DURDEN   | 1,2511                |
| 29                   | Desvio padrão<br>DBL_YAMAGUCHI         | 1,2156                |
| 30                   | Média ANISOTROPIA                      | 1,2147                |
| 31                   | Média VVVH_RAZAO_dB                    | 1,2146                |
| 32                   | Média SURF_VANZYL                      | 1,1788                |
| 33                   | Média DBL_FREEMAN-<br>DURDEN           | 1,1718                |
| 34                   | Média RVI_dB                           | 1,1708                |
| 35                   | <i>Skewness</i><br>VOL_YAMAGUCHI       | 1,1699                |
| 36                   | Mínimo <i>pixel</i> BMI_dB             | 1,163                 |
| 37                   | <i>Skewness</i><br>DBL_YAMAGUCHI       | 1,1553                |
| 38                   | Média VSI_dB                           | 1,1484                |
| 39                   | Mínimo <i>pixel</i> HH_SIGMA0          | 1,1468                |
| 40                   | Desvio padrão CSI_dB                   | 1,0974                |
| 41                   | Desvio padrão RFDI_dB                  | 1,0886                |
| 42                   | Média ENTROPIA                         | 1,0705                |
| 43                   | <i>Skewness</i> DBL_FREEMAN-<br>DURDEN | 1,0656                |
| 44                   | Desvio padrão                          | 1,0613                |

| <b>Identificador</b> | <b>ATRIBUTOS</b>                          | <b>VALOR DE GANHO</b> |
|----------------------|---|-----------------------|
|                      | HHVV_Ratio_dB                             |                       |
| 45                   | Média RFDI_dB                             | 1,0567                |
| 46                   | Mínimo <i>pixel</i> VV_SIGMA0             | 1,0533                |
| 47                   | <i>Skewness</i> DBL_VANZYL                | 1,0472                |
| 48                   | Máximo <i>pixel</i><br>DBL_FREEMAN-DURDEN | 1,0325                |
| 49                   | Desvio padrão<br>SURF_VANZYL              | 1,0237                |
| 50                   | Máximo <i>pixel</i> HH_SIGMA0             | 1,0136                |

**APÊNDICE E: MATRIZES DE CONFUSÃO REFERENTES AO CAPÍTULO 4  
(LEGENDA COM NOVE CLASSES)**

**Tabela A1.** Matriz de confusão da classificação obtida pelo algoritmo polarimétrico Wishart (PW).

|                                 |       | Referência ( <i>pixels</i> ) |               |               |              |                           |            |          |              |             | Total          |
|---------------------------------|-------|------------------------------|---------------|---------------|--------------|---------------------------|------------|----------|--------------|-------------|----------------|
|                                 |       | A                            | B             | C             | D            | E                         | F          | G        | H            | I           |                |
| Classificação ( <i>pixels</i> ) | A     | <b>251937</b>                | 18302         | 85939         | 43728        | 84648                     | 288        | 0        | 191          | 13387       | 498420         |
|                                 | B     | 6337                         | <b>377772</b> | 34210         | 1473         | 1112                      | 805        | 0        | 7911         | 849         | 430469         |
|                                 | C     | 72769                        | 276129        | <b>274777</b> | 137719       | 21135                     | 5208       | 0        | 3098         | 4901        | 795736         |
|                                 | D     | 524                          | 6816          | 1583          | <b>27017</b> | 188                       | 1191       | 0        | 394          | 0           | 37713          |
|                                 | E     | 16747                        | 11            | 162           | 1176         | <b>62841</b>              | 0          | 0        | 0            | 113         | 81050          |
|                                 | F     | 62                           | 30            | 351           | 7982         | 55                        | <b>874</b> | 0        | 50           | 0           | 9404           |
|                                 | G     | 237                          | 188           | 42            | 87           | 10847                     | 0          | <b>0</b> | 0            | 0           | 11401          |
|                                 | H     | 1356                         | 24695         | 8386          | 4095         | 303                       | 5          | 0        | <b>12596</b> | 454         | 51890          |
|                                 | I     | 91182                        | 1506          | 1758          | 5573         | 86180                     | 21         | 0        | 0            | <b>3826</b> | 190046         |
|                                 | Total | 441151                       | 705449        | 407208        | 228850       | 267309                    | 8392       | 0        | 24240        | 23530       | <b>2106129</b> |
|                                 |       | <b>Acurácia Global: 48%</b>  |               |               |              | <b>Índice Kappa: 0,35</b> |            |          |              |             |                |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

**Tabela A2.** Matriz de confusão da classificação obtida pelo algoritmo *Naive Bayes* (NB).

|                                   |       | Referência (segmentos/objetos) |     |     |    |    |                    |   |    |    | Total |
|-----------------------------------|-------|--------------------------------|-----|-----|----|----|--------------------|---|----|----|-------|
|                                   |       | A                              | B   | C   | D  | E  | F                  | G | H  | I  |       |
| Classificação (segmentos/objetos) | A     | 107                            | 0   | 38  | 18 | 17 | 1                  | 0 | 0  | 47 | 228   |
|                                   | B     | 1                              | 207 | 34  | 1  | 0  | 23                 | 0 | 18 | 0  | 284   |
|                                   | C     | 7                              | 5   | 170 | 15 | 0  | 20                 | 0 | 21 | 9  | 247   |
|                                   | D     | 1                              | 0   | 3   | 42 | 4  | 2                  | 0 | 0  | 5  | 57    |
|                                   | E     | 26                             | 0   | 2   | 7  | 52 | 0                  | 2 | 0  | 16 | 105   |
|                                   | F     | 0                              | 0   | 1   | 0  | 0  | 2                  | 0 | 0  | 0  | 3     |
|                                   | G     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 0  | 0  | 0     |
|                                   | H     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 13 | 0  | 13    |
|                                   | I     | 2                              | 0   | 1   | 2  | 0  | 1                  | 0 | 0  | 16 | 22    |
|                                   | Total | 144                            | 212 | 249 | 85 | 73 | 49                 | 2 | 52 | 93 | 959   |
|                                   |       | Acurácia Global: 64%           |     |     |    |    | Índice Kappa: 0,55 |   |    |    |       |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

**Tabela A3.** Matriz de confusão da classificação obtida pelo algoritmo J48.

|                                   |       | Referência (segmentos/objetos) |     |     |    |     |                    |   |    |    | Total |
|-----------------------------------|-------|--------------------------------|-----|-----|----|-----|--------------------|---|----|----|-------|
|                                   |       | A                              | B   | C   | D  | E   | F                  | G | H  | I  |       |
| Classificação (segmentos/objetos) | A     | 122                            | 0   | 32  | 13 | 39  | 0                  | 0 | 0  | 22 | 228   |
|                                   | B     | 1                              | 217 | 54  | 8  | 0   | 4                  | 0 | 0  | 0  | 284   |
|                                   | C     | 20                             | 12  | 191 | 15 | 3   | 4                  | 0 | 1  | 1  | 247   |
|                                   | D     | 3                              | 0   | 2   | 45 | 5   | 0                  | 0 | 0  | 2  | 57    |
|                                   | E     | 31                             | 0   | 3   | 5  | 53  | 0                  | 2 | 0  | 11 | 105   |
|                                   | F     | 0                              | 0   | 2   | 0  | 0   | 1                  | 0 | 0  | 0  | 3     |
|                                   | G     | 0                              | 0   | 0   | 0  | 0   | 0                  | 0 | 0  | 0  | 0     |
|                                   | H     | 0                              | 0   | 0   | 0  | 0   | 0                  | 0 | 13 | 0  | 13    |
|                                   | I     | 5                              | 0   | 8   | 0  | 3   | 0                  | 0 | 0  | 6  | 22    |
|                                   | Total | 182                            | 229 | 292 | 86 | 103 | 9                  | 2 | 14 | 42 | 959   |
|                                   |       | Acurácia Global: 68%           |     |     |    |     | Índice Kappa: 0,59 |   |    |    |       |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

**Tabela A4.** Matriz de confusão da classificação obtida pelo algoritmo *Random Forest* (RF).

|                                   |       | Referência (segmentos/objetos) |     |     |    |    |                    |   |    |    | Total |
|-----------------------------------|-------|--------------------------------|-----|-----|----|----|--------------------|---|----|----|-------|
|                                   |       | A                              | B   | C   | D  | E  | F                  | G | H  | I  |       |
| Classificação (segmentos/objetos) | A     | 136                            | 0   | 36  | 21 | 27 | 0                  | 0 | 0  | 8  | 228   |
|                                   | B     | 0                              | 232 | 47  | 2  | 0  | 0                  | 0 | 3  | 0  | 284   |
|                                   | C     | 8                              | 16  | 210 | 8  | 2  | 0                  | 0 | 2  | 1  | 247   |
|                                   | D     | 2                              | 0   | 3   | 46 | 3  | 0                  | 0 | 0  | 3  | 57    |
|                                   | E     | 35                             | 0   | 2   | 7  | 58 | 0                  | 0 | 0  | 3  | 105   |
|                                   | F     | 0                              | 0   | 2   | 0  | 0  | 1                  | 0 | 0  | 0  | 3     |
|                                   | G     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 0  | 0  | 0     |
|                                   | H     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 13 | 0  | 13    |
|                                   | I     | 6                              | 0   | 7   | 0  | 3  | 0                  | 0 | 0  | 6  | 22    |
|                                   | Total | 187                            | 248 | 307 | 84 | 93 | 1                  | 0 | 18 | 21 | 959   |
|                                   |       | Acurácia Global: 73%           |     |     |    |    | Índice Kappa: 0,66 |   |    |    |       |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

**Tabela A5.** Matriz de confusão da classificação obtida pelo algoritmo *Multilayer Perceptron* (MLP).

|                                   |       | Referência (segmentos/objetos) |     |     |    |    |                    |   |    |    | Total |
|-----------------------------------|-------|--------------------------------|-----|-----|----|----|--------------------|---|----|----|-------|
|                                   |       | A                              | B   | C   | D  | E  | F                  | G | H  | I  |       |
| Classificação (segmentos/objetos) | A     | 134                            | 0   | 31  | 6  | 31 | 1                  | 0 | 1  | 24 | 228   |
|                                   | B     | 0                              | 223 | 47  | 1  | 1  | 4                  | 0 | 8  | 0  | 284   |
|                                   | C     | 3                              | 8   | 215 | 2  | 1  | 7                  | 0 | 4  | 7  | 247   |
|                                   | D     | 1                              | 0   | 5   | 44 | 6  | 0                  | 0 | 0  | 1  | 57    |
|                                   | E     | 29                             | 0   | 3   | 4  | 58 | 0                  | 0 | 1  | 10 | 105   |
|                                   | F     | 0                              | 0   | 2   | 0  | 0  | 1                  | 0 | 0  | 0  | 3     |
|                                   | G     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 0  | 0  | 0     |
|                                   | H     | 0                              | 0   | 0   | 0  | 0  | 0                  | 0 | 13 | 0  | 13    |
|                                   | I     | 4                              | 0   | 5   | 0  | 1  | 0                  | 0 | 0  | 12 | 22    |
|                                   | Total | 171                            | 231 | 308 | 57 | 98 | 13                 | 0 | 27 | 54 | 959   |
|                                   |       | Acurácia Global: 73%           |     |     |    |    | Índice Kappa: 0,66 |   |    |    |       |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.



**Tabela A6.** Matriz de confusão da classificação obtida pelo algoritmo *Support Vector Machine* (SVM).

|                                   |       | Referência (segmentos/objetos) |     |     |    |     |                           |   |    |    | Total |
|-----------------------------------|-------|--------------------------------|-----|-----|----|-----|---------------------------|---|----|----|-------|
|                                   |       | A                              | B   | C   | D  | E   | F                         | G | H  | I  |       |
| Classificação (segmentos/objetos) | A     | 136                            | 0   | 39  | 7  | 39  | 0                         | 0 | 0  | 7  | 228   |
|                                   | B     | 1                              | 237 | 42  | 1  | 0   | 1                         | 0 | 2  | 0  | 284   |
|                                   | C     | 7                              | 12  | 214 | 4  | 1   | 2                         | 0 | 3  | 4  | 247   |
|                                   | D     | 3                              | 0   | 3   | 44 | 7   | 0                         | 0 | 0  | 0  | 57    |
|                                   | E     | 33                             | 0   | 1   | 3  | 66  | 0                         | 0 | 0  | 2  | 105   |
|                                   | F     | 0                              | 0   | 2   | 0  | 0   | 1                         | 0 | 0  | 0  | 3     |
|                                   | G     | 0                              | 0   | 0   | 0  | 0   | 0                         | 0 | 0  | 0  | 0     |
|                                   | H     | 0                              | 0   | 0   | 0  | 0   | 0                         | 0 | 13 | 0  | 13    |
|                                   | I     | 3                              | 0   | 4   | 0  | 5   | 0                         | 0 | 0  | 10 | 22    |
|                                   | Total | 183                            | 249 | 305 | 59 | 118 | 4                         | 0 | 18 | 23 | 959   |
|                                   |       | <b>Acurácia Global: 75%</b>    |     |     |    |     | <b>Índice Kappa: 0,68</b> |   |    |    |       |

A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo.

**APÊNDICE F: MATRIZES DE CONFUSÃO REFERENTES AO CAPÍTULO 4 (LEGENDA COM CINCO CLASSES)**

**Tabela A1.** Matriz de confusão da classificação obtida pelo algoritmo polarimétrico Wishart (PW).

| Classificação (pixels)   | Referência (pixels)     |            |                     |             |   |       |         |
|--|-------------------------|------------|---------------------|-------------|---|-------|---------|
|  |                         | A+D        | B+C+F               | E+I         | G | H     | TOTAL   |
|  | A+D                     | 323206     | 114119              | 98223       | 0 | 585   | 536133  |
|  | B+C+F                   | 226342     | 970156              | 28052       | 0 | 11059 | 1235609 |
|  | E+I                     | 114678     | 3458                | 152960      | 0 | 0     | 271096  |
|  | G                       | 324        | 230                 | 10847       | 0 | 0     | 11401   |
|  | H                       | 5451       | 33086               | 757         | 0 | 12596 | 51890   |
|  | TOTAL                   | 670001     | 1121049             | 290839      | 0 | 24240 | 2106129 |
|  | <b>Acurácia Global:</b> | <b>69%</b> | <b>Índice Kappa</b> | <b>0,48</b> |   |       |         |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                         |            |                     |             |   |       |         |

**Tabela A2.** Matriz de confusão da classificação obtida pelo algoritmo Naive Bayes (NB).

| Classificação (segmentos/objetos)  | Referência (segmentos/objetos) |            |                     |             |   |    |       |
|--|--------------------------------|------------|---------------------|-------------|---|----|-------|
|  |                                | A+D        | B+C+F               | E+I         | G | H  | TOTAL |
|  | A+D                            | 168        | 44                  | 73          | 0 | 0  | 285   |
|  | B+C+F                          | 24         | 462                 | 9           | 0 | 39 | 534   |
|  | E+I                            | 37         | 4                   | 84          | 2 | 0  | 127   |
|  | G                              | 0          | 0                   | 0           | 0 | 0  | 0     |
|  | H                              | 0          | 0                   | 0           | 0 | 13 | 13    |
|  | TOTAL                          | 229        | 510                 | 166         | 2 | 52 | 959   |
|  | <b>Acurácia Global:</b>        | <b>76%</b> | <b>Índice Kappa</b> | <b>0,60</b> |   |    |       |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                                |            |                     |             |   |    |       |

**Tabela A3.** Matriz de confusão da classificação obtida pelo algoritmo J48.

| Classificação (segmentos/objetos)  | Referência (segmentos/objetos) |            |                     |     |             |    |       |
|--|--------------------------------|------------|---------------------|-----|-------------|----|-------|
|  |                                | A+D        | B+C+F               | E+I | G           | H  | TOTAL |
|  | A+D                            | 183        | 34                  | 68  | 0           | 0  | 285   |
|  | B+C+F                          | 44         | 485                 | 4   | 0           | 1  | 534   |
|  | E+I                            | 41         | 11                  | 73  | 2           | 0  | 127   |
|  | G                              | 0          | 0                   | 0   | 0           | 0  | 0     |
|  | H                              | 0          | 0                   | 0   | 0           | 13 | 13    |
|  | TOTAL                          | 268        | 530                 | 145 | 2           | 14 | 959   |
| <b>Acurácia Global:</b>  |                                | <b>79%</b> | <b>Índice Kappa</b> |     | <b>0,64</b> |    |       |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                                |            |                     |     |             |    |       |

**Tabela A4.** Matriz de confusão da classificação obtida pelo algoritmo *Random Forest* (RF).

| Classificação (segmentos/objetos)  | Referência (segmentos/objetos) |            |                     |     |             |    |       |
|--|--------------------------------|------------|---------------------|-----|-------------|----|-------|
|  |                                | A+D        | B+C+F               | E+I | G           | H  | TOTAL |
|  | A+D                            | 205        | 39                  | 41  | 0           | 0  | 285   |
|  | B+C+F                          | 18         | 508                 | 3   | 0           | 5  | 534   |
|  | E+I                            | 48         | 9                   | 70  | 0           | 0  | 127   |
|  | G                              | 0          | 0                   | 0   | 0           | 0  | 0     |
|  | H                              | 0          | 0                   | 0   | 0           | 13 | 13    |
|  | TOTAL                          | 271        | 556                 | 114 | 0           | 18 | 959   |
| <b>Acurácia Global:</b>  |                                | <b>83%</b> | <b>Índice Kappa</b> |     | <b>0,71</b> |    |       |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                                |            |                     |     |             |    |       |

**Tabela A5.** Matriz de confusão da classificação obtida pelo algoritmo *Multilayer Perceptron* (MLP).

| Classificação (segmentos/objetos)  | Referência (segmentos/objetos) |            |                     |     |             |    |       |
|--|--------------------------------|------------|---------------------|-----|-------------|----|-------|
|  |                                | A+D        | B+C+F               | E+I | G           | H  | TOTAL |
|  | A+D                            | 185        | 37                  | 62  | 0           | 1  | 285   |
|  | B+C+F                          | 6          | 507                 | 9   | 0           | 12 | 534   |
|  | E+I                            | 37         | 8                   | 81  | 0           | 1  | 127   |
|  | G                              | 0          | 0                   | 0   | 0           | 0  | 0     |
|  | H                              | 0          | 0                   | 0   | 0           | 13 | 13    |
|  | TOTAL                          | 228        | 552                 | 152 | 0           | 27 | 959   |
| <b>Acurácia Global:</b>  |                                | <b>82%</b> | <b>Índice Kappa</b> |     | <b>0,69</b> |    |       |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                                |            |                     |     |             |    |       |

**Tabela A6.** Matriz de confusão da classificação obtida pelo algoritmo *Support Vector Machine* (SVM).

| Classificação (segmentos/objetos)  | Referência (segmentos/objetos) |            |                     |     |             |    |       |
|--|--------------------------------|------------|---------------------|-----|-------------|----|-------|
|  |                                | A+D        | B+C+F               | E+I | G           | H  | TOTAL |
|  | A+D                            | 190        | 42                  | 53  | 0           | 0  | 285   |
|  | B+C+F                          | 13         | 511                 | 5   | 0           | 5  | 534   |
|  | E+I                            | 39         | 5                   | 83  | 0           | 0  | 127   |
|  | G                              | 0          | 0                   | 0   | 0           | 0  | 0     |
|  | H                              | 0          | 0                   | 0   | 0           | 13 | 13    |
|  | TOTAL                          | 242        | 558                 | 141 | 0           | 18 | 959   |
| <b>Acurácia Global:</b>  |                                | <b>83%</b> | <b>Índice Kappa</b> |     | <b>0,71</b> |    |       |
| A = Pastagens; B = Formações florestais; C = Cerrado Típico antropizado/Cerrado Ralo/Campo Cerrado Sujo; D = Culturas temporárias; E = Solo exposto/palhada; F = Silvicultura; G = Corpo d'água continental; H = Áreas com edificações; I = Campo Limpo. |                                |            |                     |     |             |    |       |

**APÊNDICE G: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4: SIGNIFICÂNCIAS DOS ÍNDICES KAPPA (LEGENDA COM NOVE CLASSES)**

**Quadro A1.** Teste de hipótese da classificação obtida pelo algoritmo polarimétrico Wishart (PW).

H0:  $Kappa\ PW = 0$   
 H1:  $Kappa\ PW > 0$   
 $Z = 841,46$ ; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita  $H_0$  (Kappa PW é significativamente maior que zero)

**Quadro A2.** Teste de hipótese da classificação obtida pelo algoritmo *Naive Bayes* (NB).

H0:  $Kappa\ NB = 0$   
 H1:  $Kappa\ NB > 0$   
 $Z = 30,35$ ; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita  $H_0$  (Kappa NB é significativamente maior que zero)

**Quadro A3.** Teste de hipótese da classificação obtida pelo algoritmo J48.

H0:  $Kappa\ J48 = 0$   
 H1:  $Kappa\ J48 > 0$   
 $Z = 31,46$ ; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita  $H_0$  (Kappa J48 é significativamente maior que zero)

**Quadro A4.** Teste de hipótese da classificação obtida pelo algoritmo *Random Forest* (RF).

H0:  $Kappa\ RF = 0$   
 H1:  $Kappa\ RF > 0$   
 $Z = 36,43$ ; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita  $H_0$  (Kappa RF é significativamente maior que zero)

**Quadro A5.** Teste de hipótese da classificação obtida pelo algoritmo *Multilayer Perceptron* (MLP).

H0: Kappa MLP = 0

H1: Kappa MLP > 0

Z = 37,09; P-valor = 0,0000;  $\alpha = 0,05$

Conclusão: Rejeita H<sub>0</sub> (Kappa MLP é significativamente maior que zero)

**Quadro A6.** Teste de hipótese da classificação obtida pelo algoritmo *Support Vector Machine* (SVM).

H0: Kappa SVM = 0

H1: Kappa SVM > 0

Z = 38,65; P-valor = 0,0000;  $\alpha = 0,05$

Conclusão: Rejeita H<sub>0</sub> (Kappa SVM é significativamente maior que zero)

**APÊNDICE H: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4: SIGNIFICÂNCIAS DOS ÍNDICES KAPPA (LEGENDA COM CINCO CLASSES)**

**Quadro A1.** Teste de hipótese da classificação obtida pelo algoritmo polarimétrico Wishart (PW).

H0: Kappa PW = 0  
 H1: Kappa PW > 0  
 Z = 943,20; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita H<sub>0</sub> (Kappa PW é significativamente maior que zero)

**Quadro A2.** Teste de hipótese da classificação obtida pelo algoritmo *Naive Bayes* (NB).

H0: Kappa NB = 0  
 H1: Kappa NB > 0  
 Z = 53,90; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita H<sub>0</sub> (Kappa NB é significativamente maior que zero)

**Quadro A3.** Teste de hipótese da classificação obtida pelo algoritmo J48.

H0: Kappa J48 = 0  
 H1: Kappa J48 > 0  
 Z = 54,73; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita H<sub>0</sub> (Kappa J48 é significativamente maior que zero)

**Quadro A4.** Teste de hipótese da classificação obtida pelo algoritmo *Random Forest* (RF).

H0: Kappa RF = 0  
 H1: Kappa RF > 0  
 Z = 54,37; P-valor = 0,0000;  $\alpha = 0,05$   
 Conclusão: Rejeita H<sub>0</sub> (Kappa RF é significativamente maior que zero)

**Quadro A5.** Teste de hipótese da classificação obtida pelo algoritmo *Multilayer Perceptron* (MLP).

H0: Kappa MLP = 0

H1: Kappa MLP > 0

Z = 55,03; P-valor = 0,0000;  $\alpha = 0,05$

Conclusão: Rejeita H<sub>0</sub> (Kappa MLP é significativamente maior que zero)

**Quadro A6.** Teste de hipótese da classificação obtida pelo algoritmo *Support Vector Machine* (SVM).

H0: Kappa SVM = 0

H1: Kappa SVM > 0

Z = 54,24; P-valor = 0,0000;  $\alpha = 0,05$

Conclusão: Rejeita H<sub>0</sub> (Kappa SVM é significativamente maior que zero)



**APÊNDICE I: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4:  
COMPARAÇÕES DE ÍNDICES KAPPA (LEGENDA COM NOVE CLASSES)**

**Quadro A1.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Naive Bayes* (NB).

H<sub>0</sub>: Kappa PW – Kappa NB = 0  
 H<sub>1</sub>: Kappa PW – Kappa NB < 0  
 Z = -11,06; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa NB).

**Quadro A2.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e J48.

H<sub>0</sub>: Kappa PW – Kappa J48 = 0  
 H<sub>1</sub>: Kappa PW – Kappa J48 < 0  
 Z = -12,68; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa J48).

**Quadro A3.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Random Forest* (RF).

H<sub>0</sub>: Kappa PW – Kappa RF = 0  
 H<sub>1</sub>: Kappa PW – Kappa RF < 0  
 Z = -16,92 ; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa RF).

**Quadro A4.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Multilayer Perceptron* (MLP).

H<sub>0</sub>: Kappa PW – Kappa MLP = 0  
 H<sub>1</sub>: Kappa PW – Kappa MLP < 0  
 Z = -17,26; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa MLP).

**Quadro A5.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Support Vector Machine* (SVM).

$H_0$ : Kappa PW – Kappa SVM = 0  
 $H_1$ : Kappa PW – Kappa SVM < 0  
 $Z = -18,72$ ; P-valor = 0,0000;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa PW é significativamente menor que o Kappa SVM).

**Quadro A6.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e J48.

$H_0$ : Kappa NB – Kappa J48 = 0  
 $H_1$ : Kappa NB – Kappa J48 < 0  
 $Z = -1,38$ ; P-valor = 0,0837;  $\alpha = 0,05$   
Conclusão: Aceita-se  $H_0$  (Os Kappas PW e NB não são significativamente diferentes).

**Quadro A7.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Random Forest* (RF).

$H_0$ : Kappa NB – Kappa RF = 0  
 $H_1$ : Kappa NB – Kappa RF < 0  
 $Z = -4,06$ ; P-valor = 0,0000;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa NB é significativamente menor que o Kappa RF).

**Quadro A8.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Multilayer Perceptron* (MLP).

$H_0$ : Kappa NB – Kappa MLP = 0  
 $H_1$ : Kappa NB – Kappa MLP < 0  
 $Z = -4,14$ ; P-valor = 0,0000;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa NB é significativamente menor que o Kappa MLP).

**Quadro A9.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Support Vector Machine* (SVM).

$H_0$ : Kappa NB – Kappa SVM = 0  
 $H_1$ : Kappa NB – Kappa SVM < 0  
 $Z = -5,11$ ; P-valor = 0,0000;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa NB é significativamente menor que o Kappa SVM).

**Quadro A10.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Random Forest* (RF).

$H_0$ : Kappa J48 – Kappa RF = 0  
 $H_1$ : Kappa J48 – Kappa RF < 0  
 $Z = -2,60$ ; P-valor = 0,0046;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa RF).

**Quadro A11.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Multilayer Perceptron* (MLP).

$H_0$ : Kappa J48 – Kappa MLP = 0  
 $H_1$ : Kappa J48 – Kappa MLP < 0  
 $Z = -2,67$ ; P-valor = 0,0038;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa MLP).

**Quadro A12.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Support Vector Machine* (SVM).

$H_0$ : Kappa J48 – Kappa SVM = 0  
 $H_1$ : Kappa J48 – Kappa SVM < 0  
 $Z = -3,62$ ; P-valor = 0,0001;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa SVM).

**Quadro A13.** Teste de hipótese comparando o desempenho dos classificadores *Random Forest* (RF) e *Multilayer Perceptron* (MLP).

$H_0$ : Kappa RF – Kappa MLP = 0

$H_1$ : Kappa RF – Kappa MLP < 0

Z = -0,05; P-valor = 0,4790;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas RF e MLP não são significativamente diferentes).

**Quadro A14.** Teste de hipótese comparando o desempenho dos classificadores *Random Forest* (RF) e *Support Vector Machine* (SVM).

$H_0$ : Kappa RF – Kappa SVM = 0

$H_1$ : Kappa RF – Kappa SVM < 0

Z = -1,01; P-valor = 0,1565;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas RF e SVM não são significativamente diferentes).

**Quadro A15.** Teste de hipótese comparando o desempenho dos classificadores *Multilayer Perceptron* (MLP) e *Support Vector Machine* (SVM).

$H_0$ : Kappa MLP – Kappa SVM = 0

$H_1$ : Kappa MLP – Kappa SVM < 0

Z = -0,96; P-valor = 0,1673;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas MLP e SVM não são significativamente diferentes).

**APÊNDICE J: TESTES DE HIPÓTESES REFERENTES AO CAPÍTULO 4: COMPARAÇÕES DE ÍNDICES KAPPA (LEGENDA COM CINCO CLASSES)**

**Quadro A1.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Naive Bayes* (NB).

H<sub>0</sub>: Kappa PW – Kappa NB = 0  
 H<sub>1</sub>: Kappa PW – Kappa NB < 0  
 Z = -11,16; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa NB).

**Quadro A2.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e J48.

H<sub>0</sub>: Kappa PW – Kappa J48 = 0  
 H<sub>1</sub>: Kappa PW – Kappa J48 < 0  
 Z = -13,65; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa J48).

**Quadro A3.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Random Forest* (RF).

H<sub>0</sub>: Kappa PW – Kappa RF = 0  
 H<sub>1</sub>: Kappa PW – Kappa RF < 0  
 Z = -17,52; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa RF).

**Quadro A4.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Multilayer Perceptron* (MLP).

H<sub>0</sub>: Kappa PW – Kappa MLP = 0  
 H<sub>1</sub>: Kappa PW – Kappa MLP < 0  
 Z = -17,06; P-valor = 0,0000; α = 0,05  
 Conclusão: Rejeita-se H<sub>0</sub> (Kappa PW é significativamente menor que o Kappa MLP).

**Quadro A5.** Teste de hipótese comparando o desempenho dos classificadores polarimétricos Wishart (PW) e *Support Vector Machine* (SVM).

|   |
|---|
| H <sub>0</sub> : Kappa PW – Kappa SVM = 0   |
| H <sub>1</sub> : Kappa PW – Kappa SVM < 0   |
| Z = -17,68; P-valor = 0,0000; α = 0,05  |
| Conclusão: Rejeita-se H <sub>0</sub> (Kappa PW é significativamente menor que o Kappa SVM). |

**Quadro A6.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e J48.

|  |
|--|
| H <sub>0</sub> : Kappa NB – Kappa J48 = 0  |
| H <sub>1</sub> : Kappa NB – Kappa J48 < 0  |
| Z = -1,15; P-valor = 0,1242; α = 0,05  |
| Conclusão: Aceita-se H <sub>0</sub> (Os Kappas PW e NB não são significativamente diferentes). |

**Quadro A7.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Random Forest* (RF).

|  |
|--|
| H <sub>0</sub> : Kappa NB – Kappa RF = 0   |
| H <sub>1</sub> : Kappa NB – Kappa RF < 0   |
| Z = -3,58; P-valor = 0,0002; α = 0,05  |
| Conclusão: Rejeita-se H <sub>0</sub> (Kappa NB é significativamente menor que o Kappa RF). |

**Quadro A8.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Multilayer Perceptron* (MLP).

|   |
|---|
| H <sub>0</sub> : Kappa NB – Kappa MLP = 0   |
| H <sub>1</sub> : Kappa NB – Kappa MLP < 0   |
| Z = -3,15; P-valor = 0,0008; α = 0,05   |
| Conclusão: Rejeita-se H <sub>0</sub> (Kappa NB é significativamente menor que o Kappa MLP). |

**Quadro A9.** Teste de hipótese comparando o desempenho dos classificadores *Naive Bayes* (NB) e *Support Vector Machine* (SVM).

$H_0$ : Kappa NB – Kappa SVM = 0  
 $H_1$ : Kappa NB – Kappa SVM < 0  
 $Z = -3,73$ ; P-valor = 0,0001;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa NB é significativamente menor que o Kappa SVM).

**Quadro A10.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Random Forest* (RF).

$H_0$ : Kappa J48 – Kappa RF = 0  
 $H_1$ : Kappa J48 – Kappa RF < 0  
 $Z = -2,38$ ; P-valor = 0,0087;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa RF).

**Quadro A11.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Multilayer Perceptron* (MLP).

$H_0$ : Kappa J48 – Kappa MLP = 0  
 $H_1$ : Kappa J48 – Kappa MLP < 0  
 $Z = -1,94$ ; P-valor = 0,0260;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa MLP).

**Quadro A12.** Teste de hipótese comparando o desempenho dos classificadores J48 e *Support Vector Machine* (SVM).

$H_0$ : Kappa J48 – Kappa SVM = 0  
 $H_1$ : Kappa J48 – Kappa SVM < 0  
 $Z = -2,52$ ; P-valor = 0,0058;  $\alpha = 0,05$   
Conclusão: Rejeita-se  $H_0$  (Kappa J48 é significativamente menor que o Kappa SVM).

**Quadro A13.** Teste de hipótese comparando o desempenho dos classificadores *Random Forest* (RF) e *Multilayer Perceptron* (MLP).

$H_0$ : Kappa RF – Kappa MLP = 0

$H_1$ : Kappa RF – Kappa MLP < 0

Z = 0,45; P-valor = 0,3265;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas RF e MLP não são significativamente diferentes).

**Quadro A14.** Teste de hipótese comparando o desempenho dos classificadores *Random Forest* (RF) e *Support Vector Machine* (SVM).

$H_0$ : Kappa RF – Kappa SVM = 0

$H_1$ : Kappa RF – Kappa SVM < 0

Z = -0,14; P-valor = 0,4449;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas RF e SVM não são significativamente diferentes).

**Quadro A15.** Teste de hipótese comparando o desempenho dos classificadores *Multilayer Perceptron* (MLP) e *Support Vector Machine* (SVM).

$H_0$ : Kappa MLP – Kappa SVM = 0

$H_1$ : Kappa MLP – Kappa SVM < 0

Z = -0,59; P-valor = 0,2766;  $\alpha$  = 0,05

Conclusão: Aceita-se  $H_0$  (Os Kappas MLP e SVM não são significativamente diferentes).



## ANEXO A: COMPROVANTE DE SUBMISSÃO E SITUAÇÃO DO ARTIGO # 1



Flávio Fortes Camargo &lt;flaviofortescamargo@gmail.com&gt;

---

**Your manuscript, CAGEO\_2018\_421, has not been accepted**

---

Derek Karssenberg (Computers and Geosciences) &lt;Evisesupport@elsevier.com&gt;

1 de novembro de 2018 às  
07:27

Responder a: d.karssenberg.editor.cageo@uu.nl

Para: flavio.fortes.camargo@gmail.com

Ref: CAGEO\_2018\_421

Title: Data mining techniques applied to ALOS-2/PALSAR-2 satellite imagery for land use and land cover classification

Journal: Computers and Geosciences

Dear Mr. Camargo,

Thank you for submitting your manuscript to Computers and Geosciences. I regret to inform you that the Associate Editor and reviewers have advised against publishing your manuscript, and I must therefore reject it.

Please refer to the comments listed at the end of this letter for details of why I reached this decision.

We appreciate your submitting your manuscript to this journal and for giving us the opportunity to consider your work.

Kind regards,

Dr Derek Karssenberg  
Editor-in-Chief  
Computers and Geosciences

**Comments from the editors and reviewers:****-Associate Editor**

All the three reviewers raised some serious criticisms on the novelty, experimental setting and English writing. So I must reject this manuscript.

**-Reviewer 1**

This paper proposes a process workflow for dual-pol SAR data classification. The content is easily followed and understood. But, the reviewer could not find novel point from this paper. After over-segmentation, the object-based classification has been widely reported on PolSAR data process. The authors have to explain their contribution on remote sensing society and address the novelty when comparing with previous researches. Therefore, I do not recommend to publish this manuscript. The detail comments are followed.

1) The proposed workflow and the comparison DT, RF and MLP ANN should be widely tested in other data sets. The conclusion is unreliable from a single data experiment.

2) In Fig. 5, it seems that the number of validation blocks is small. The authors should consider to exploit the whole image extent to validate their workflow.

3) The experiment should be repeated dozens of times to avoid the influence of randomly picking training samples.

4) The speckle noise always impairs the segmentation and classification process. How author address this problem.

5) Again, this research is a repeat and combination of current techniques. The novel point is not clear.

**-Reviewer 2**

A few major technical problems in this manuscript can be summarized as below.

1) According to my understanding, all the three classification methods (J48 DT, RF, MLP ANN) used in this study cannot be categorized as data mining techniques. They can only be called as machine learning algorithms. Therefore, the paper title actually does not agree with the content.

2) The study carried out in this manuscript just focused on LULC classification of a single image acquired by ALOS-2/PALSAR-2 using traditional machine learning algorithms. This topic is quite old in the field of SAR remote sensing, while the authors failed to put forward any new problem or develop any new method.

3) In the experiments, the authors simply used the established WEKA software to carry out the whole chain of data processing, and in my opinion, they did not make any methodological innovation.

4) It is well known that speckle suppression is a vital preprocessing step for thematic information extraction from satellite SAR imagery. However, in this study the authors did not consider taking any measure to perform speckle suppression before image segmentation and classification, which might partially account for the relatively low accuracy of classification results.

5) The content organization and structure of this manuscript are rather awful. The authors spent a lot of texts on irrelevant stuff, e.g. the whole section 2. Furthermore, there are two "section 4" in the manuscript, which is misleading.

### **-Reviewer 3**

Due to the low novelty, poor organizing and English writing, I must reject this paper.

### **Have questions or need assistance?**

For further assistance, please visit our [Customer Support](#) site. Here you can search for solutions on a range of topics, find answers to frequently asked questions, and learn more about EVISE® via interactive tutorials. You can also talk 24/5 to our customer support team by phone and 24/7 by live chat and email.

---

Copyright © 2018 Elsevier B.V. | [Privacy Policy](#)

Elsevier B.V., Radarweg 29, 1043 NX Amsterdam, The Netherlands, Reg. No. 33156677.

**ANEXO B: DADOS DE PRECIPITAÇÃO PLUVIOMÉTRICA DA ESTAÇÃO METEOROLÓGICA AUTOMÁTICA DO INMET (01/05/2016 - 14/05/2016) – ESTAÇÃO ECOLÓGICA ÁGUAS EMENDADAS (BRASÍLIA/DF).**



MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO-MAPA

INSTITUTO NACIONAL DE METEOROLOGIA - INMET

ESTAÇÃO METEOROLÓGICA AUTOMÁTICA DE AGUAS EMENDADAS/DF

Alt. 1200m  
Lat. 15°35'S  
Lon. 47°37'W

| HORA UTC    | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>(mm) | PRECIPITAÇÃO<br>ACUMULADA (mm) |
|-------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|--------------------------------|
| 0000        |                      |                      |                      |                      |                      |                      |                      |                      |                                |
| 01-mai-2016 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | 0,0                            |
| 02-mai-2016 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | 0,0                            |
| 03-mai-2016 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | NULL                 | 0,0                            |
| 04-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 05-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 06-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 07-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 08-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 09-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 10-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 11-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 12-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 13-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |
| 14-mai-2016 | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                  | 0,0                            |