

**UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL
UNIDADE UNIVERSITÁRIA DE CASSILÂNDIA
PÓS-GRADUAÇÃO EM AGRONOMIA**

**POTENCIAL DE UTILIZAÇÃO DE ATRIBUTOS ESPACIAIS NO PROCESSO
DE CLASSIFICAÇÃO DIGITAL DE PLANTIOS DE SERINGUEIRA NO MATO
GROSSO DO SUL**

JÚLIA VAZ TOSTES MILUZZI DE OLIVEIRA

Dissertação apresentada à Universidade Estadual de Mato Grosso do Sul, Unidade Universitária de Cassilândia, como parte das exigências para a obtenção do título de Mestre em Agronomia - Sustentabilidade na Agricultura.

**CASSILÂNDIA – MS
Agosto/2018**

UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL
UNIDADE UNIVERSITÁRIA DE CASSILÂNDIA
PÓS-GRADUAÇÃO EM AGRONOMIA

**POTENCIAL DE UTILIZAÇÃO DE ATRIBUTOS ESPACIAIS NO PROCESSO
DE CLASSIFICAÇÃO DIGITAL DE PLANTIOS DE SERINGUEIRA NO MATO
GROSSO DO SUL**

JÚLIA VAZ TOSTES MILUZZI DE OLIVEIRA

Orientador: **Prof. Dr. Erivaldo José Scaloppi Junior**

Co-orientador: **Prof. Dr. Fausto Weimar Acerbi Junior**

Dissertação apresentada à Universidade Estadual de Mato Grosso do Sul, Unidade Universitária de Cassilândia, como parte das exigências para a obtenção do título de Mestre em Agronomia - Sustentabilidade na Agricultura.

CASSILÂNDIA – MS
Agosto/2018

O47p Oliveira, Júlia Vaz Tostes Miluzzi de
Potencial de utilização de atributos espaciais no processo de
classificação digital de plantios de seringueira no Mato Grosso
do Sul informacional/ Júlia Vaz Tostes Miluzzi de Oliveira.
Cassilândia, MS: UEMS, 2018.
158p. ; 30cm

Dissertação (Mestrado) – Agronomia – Universidade
Estadual de Mato Grosso do Sul, 2018.
Orientador: Prof. Dr Erivaldo José Scaloppi Junior
Co-orientador: Dr. Fausto Weimar Acerbi Junior

1. Heveicultura 2. Sensoriamento Remoto 3. Geoestatística I.
Título CDD 23.ed. 633.8952



Universidade Estadual
de Mato Grosso do Sul

UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL
UNIDADE UNIVERSITÁRIA DE CASSILÂNDIA
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA
ÁREA DE CONCENTRAÇÃO EM SUSTENTABILIDADE NA AGRICULTURA-
PGAC



“POTENCIAL DE UTILIZAÇÃO DE ATRIBUTOS ESPACIAIS NO PROCESSO
DE CLASSIFICAÇÃO DIGITAL DE PLANTIOS DE SERINGUEIRA NO MATO
GROSSO DO SUL”

Acadêmica: Júlia Vaz Tostes Miluzzi de Oliveira

Orientador: Erivaldo José Scaloppi Junior

APROVADO: 16/08/2018

Erivaldo José Scaloppi Junior
Orientador

Alan de Brito

Flávio Jorge Ponzoni

“Somos a memória que temos e a responsabilidade que assumimos.

*Sem memória não existimos,
sem responsabilidade talvez não mereçamos existir”.*

José Saramago

DEDICATÓRIA

*Para minha mãe Celina,
e meus irmãos João Victor e Bruno,
vocês são o que eu tenho de mais valioso!*

AGRADECIMENTOS

Primeiramente gostaria de agradecer à minha ‘pãe’ Celina Vas Tostes por toda dedicação e amor incondicionais. Por todo colo e palavra reconfortante em momentos providenciais, mesmo eu sendo tão ‘marruá’ e péssima em falar a respeito do que me aflige. Por mesmo com toda dificuldade ter me proporcionado uma vida tão rica em aprendizados das mais diversas fontes, nossa riqueza sempre foi abstrata, mas não menos real e importante por isso. Por ter me ensinado a nunca me limitar a nada e ter me mostrado que eu posso tudo o que me propor desde que eu coloque meu coração no que estiver fazendo. Mãe, crescer com você é e sempre foi uma aventura, e a sra. sempre vai ser a minha melhor amiga.

Agradeço também aos meus irmãos João Victor e Bruno pelo amor e confiança incondicionais, mesmo quando eu não mereço. Por todo apoio e força e por serem motivo de tanto orgulho.

Devo a ‘dona’ Celina também a oportunidade de ter nascido e crescido nessa família, que foi onde conheci as mulheres mais fantásticas, as minhas mulheres preferidas, que são um pouco minhas mães também. Obrigada Vovó Ruth, Tia Célia, Tia Nini, Tia Cilene, Tia Dinha e Nany, se eu sou o que sou hoje devo cada pedacinho a vocês que sempre me instruíram, acreditaram em mim e que cada uma a sua forma, me ajudou a continuar, mesmo quando eu me senti sem forças e quis desistir.

Eu não me vejo em nenhuma outra família no mundo, por que foi com vocês que aprendi as lições mais valiosas e que vivi os momentos mais importantes. Obrigada ao Vovô Zeca por todo amor que recebi, mesmo num curto espaço de tempo, e por mesmo não estando mais conosco sempre ter estado presente. Ao Tio Vítor e ao Padrinho Paulo por terem sido tão presentes na minha vida. A meus primos Fábio, Maria Vitória, Enzo, Pedro Ricardo, Denner e Isabela por toda a diversão e amizade. Dizem que os primos são os primeiros melhores amigos... e eu não consigo discordar. Obrigada também a Tia Nelma que sempre participou da minha vida e mesmo insistindo que é agregada, é Vaz Tostes de coração, e às minhas cunhadas lindas Iara e Beatriz. Obrigada a todos os ‘parentes’ que estiveram comigo nessa luta, tudo isso é por vocês Vaz Tostes!

Aos meus amigos Bruna, Clara, Lívia, Sulipa, Alan, Bárbara, Samantha, Monique, Gui e Clarissa, por sempre acreditar em mim e por fazerem minha vida mais divertida e com um sentido mais amplo.

Aos meus orientadores Dr. Erivaldo Scaloppi Junior e Dr. Fausto Weimar Acerbi Junior por todo apoio, amizade, puxões de orelha e por terem confiado em mim no desenvolvimento deste trabalho.

À família 'De Pieri-Sepulveda' por ter me acolhido e feito com que eu me sentisse em casa. À Maria Paula, Fá, Pedrinho Pedro E Pedro Henrique que foram minha segunda família e nunca duvidaram que eu fosse conseguir.

Aos grandes amigos que fiz em Cassilândia Fernando, Gustavo, Verinha, Kelle, Vicente, Vick e Marquinhos. Vocês fizeram dessa época uma época maravilhosa da minha vida e sempre vão estar comigo.

A Turma do Mestrado 2015/1 por todo conhecimento e cervejas que dividimos.

Agradeço também aos professores da UEMS por toda dedicação e por terem me proporcionado uma visão mais ampla do ensino superior no Brasil.

Todos vocês contribuíram e/ou ainda contribuem para que minha caminhada seja feita com menos sobrecarga, de forma mais leve. Eu me considero abençoada sempre ter sido cercada por pessoas que dividem o peso da caminhada ao invés de somar. Dividir não deixa pesado pra ninguém e é sempre mais agradável do que somar, assim como encontrar soluções juntos é sempre a melhor do que reclamar. Obrigada por terem me proporcionado essa visão da vida, e por consequência essa visão de mim.

Então, finalmente, mas não menos importante agradeço a Deus pela saúde, pela inteligência, pela determinação e por ter me feito merecedora e rodeada de todas as oportunidades e pessoas ímpares que fizeram e/ou ainda fazem parte da minha caminhada.

SUMÁRIO

LISTA DE TABELAS	X
LISTA DE FIGURAS	XIII
LISTA DE ABREVIACÕES E SÍMBOLOS	XIV
INTRODUÇÃO	18
CAPÍTULO 1 – CONSIDERAÇÕES GERAIS	21
1.1. Heveicultura	21
1.1.1. Contexto histórico e ambiental da cultura.....	21
1.1.2. Balança comercial: oferta vs. demanda	23
1.1.3. A heveicultura no Mato Grosso do Sul: Caracterização e Perspectivas.....	25
1.1.4. Sazonalidade da Seringueira	26
1.2. Técnicas de Sensoriamento Remoto no monitoramento da vegetação	27
1.2.1. Índices de Vegetação.....	29
1.2.2. Segmentação: do <i>pixel</i> ao objeto.....	31
1.2.3. Sensoriamento Remoto na heveicultura	33
1.2.4. Reconhecimento de padrões: Espectral vs. Espacial	35
1.2.4.1. Geoestatística	36
1.2.4.2. Caracterização do semivariograma	39
1.2.4.3. Aplicação de semivariogramas para a classificação.....	40
1.3. Random Forest	45
1.3.1. De uma <i>Random Tree</i> à uma <i>Random Forest</i>	49
1.3.2. O <i>Random Forest</i> na classificação	50
REFERÊNCIAS	53
CAPÍTULO 2. ATRIBUTOS ESPECTRAIS E ESPACIAIS NO MAPEAMENTO DE SERINGAIS – ESTUDO DE CASO DE UMA ÁREA NO MATO GROSSO DO SUL	65
2.1. Introdução.....	67
2.2. Material e métodos	69
2.2.1. Área de Estudo	69
2.2.2. Banco de dados (Imagens e dados)	70

2.2.3.	Metodologia	72
2.2.3.1.	Segmentação.....	73
2.2.3.2.	Amostragem	74
2.2.3.3.	Extração de atributos	75
2.2.3.3.1.	Cálculo dos semivariogramas e extração de atributos espaciais por objeto	75
2.2.3.3.2.	Extração de atributos espectrais	77
2.2.3.4.	Exclusão de atributos espaciais	78
2.2.3.5.	Classificação.....	78
2.2.3.6.	Avaliação da classificação.....	78
2.2.4.	Análise exploratória dos dados.....	79
2.1.	Resultados e discussão	86
2.1.1.	Segmentação.....	86
2.1.2.	Semivariogramas para as classes em estudo	88
2.1.3.	Testes.....	89
2.1.4.	Comparações entre os testes.....	91
2.1.4.1.	Comparação 1.....	91
2.1.4.2.	Comparação 2.....	94
2.1.4.3.	Comparação 3.....	103
2.1.4.4.	Comparação 4.....	110
2.2.	Conclusões	118
REFERÊNCIAS		119

CAPÍTULO 3. AVALIAÇÃO DE GRUPOS DE ATRIBUTOS ESPACIAIS INSERIDOS NA CLASSIFICAÇÃO DIGITAL DE BASE ESPECTRAL DE PLANTIOS DE SERINGUEIRA NO MATO GROSSO DO SUL 122

3.1.	Introdução.....	124
3.2.	Material e métodos	126
3.2.1.	Área de Estudo	126
3.2.2.	Banco de dados (Imagens e dados)	128
3.2.3.	Metodologia.....	129
3.2.3.1.	Segmentação.....	130
3.2.3.2.	Amostragem	131
3.2.3.3.	Extração de atributos	131
3.2.3.3.1.	Cálculo dos semivariogramas e extração de atributos espaciais por objeto	132
3.2.3.3.2.	Extração de atributos espectrais	133

2.2.4.1. Contribuição dos atributos	134
3.2.4. Classificação.....	134
3.2.5. Avaliação da classificação.....	137
3.3. Resultados e discussão	138
3.3.1. Classificações	138
3.3.1.1. Atributos espaciais extraídos da fonte de informação R	139
3.3.1.2. Atributos espaciais extraídos da fonte de informação NIR	142
3.3.1.3. Atributos espaciais extraídos da fonte de informação NDVI.....	146
3.3.1.4. Melhores classificações.....	151
3.4. Conclusões	155
REFEREÊNCIAS.....	156

LISTA DE TABELAS

Tabela 1	Produção de borracha natural em 2016	24
Tabela 2	Atributos espaciais descritos em Balaguer et al. (2010) divididas em: (A) Atributos extraídos próximas à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos	76
Tabela 3	Detalhamento dos atributos utilizados nos testes e resumo dos testes realizados, com seus respectivos objetivos principais e tipos de atributos utilizados. Linha tracejada divide os três grupos de testes (puramente espectrais, puramente espaciais e espectrais-espaciais)	84
Tabela 4	Atributos utilizados nos testes com suas respectivas divisões: (A) Atributos extraídos próximas à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos (S=Sim, a fonte de informação foi utilizada e N=Não, a fonte de informação não foi utilizada)	91
Tabela 5	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados nessa comparação 1	92
Tabela 6	Matrizes de confusão dos testes 1 e 2 (acurácias do produtor nas colunas e acurácias do usuário nas linhas)	92
Tabela 7	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados nessa comparação 2	95
Tabela 8	Matrizes de confusão dos testes 2, 3, 4, 5 e 6 (acurácias do produtor nas colunas e acurácias do usuário nas linhas)	95
Tabela 9	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados na Comparação 3	103
Tabela 10	Matrizes de confusão dos testes 7, 8, 9 e 10 (acurácias do produtor nas colunas e acurácias do usuário nas linhas)	103
Tabela 11	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados na Comparação 4	110
Tabela 12	Matrizes de confusão dos testes 11, 12, 13 e 14 (acurácias do produtor nas colunas e acurácias do usuário nas linhas)	111
Tabela 13	Atributos espaciais descritos em Balaguer <i>et al.</i> (2010) divididas em: (A) Atributos extraídos próximas à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos	133
Tabela 14	Resumo dos atributos utilizados nas classificações realizadas e o número de atributos contidos em cada classificação (S= Sim, foi utilizada a fonte de informação marcada na classificação)	137
Tabela 15	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação R	139

Tabela 16	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 1 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	140
Tabela 17	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 2 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	141
Tabela 18	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 2 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	142
Tabela 19	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação NIR	143
Tabela 20	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 4 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	144
Tabela 21	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 5 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	145
Tabela 22	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 6 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	146
Tabela 23	Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação NDVI	147
Tabela 24	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 7 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	148
Tabela 25	<i>Ranking</i> da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 8 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise)	149

Tabela 26 *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 9 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise) 150

LISTA DE FIGURAS

Figura 1	Plantios de Seringueira enfolhados e desfolhados, em comparação com floresta nativa	34
Figura 2	Semivariograma – Distância x Semivariância	37
Figura 3	Semivariograma e seus parâmetros Patamar (C), Alcance (A) e Efeito Pepita (Co)	40
Figura 4	Atributos próximos a origem	42
Figura 5	Atributos até o primeiro máximo e entre o primeiro e segundo máximos	45
Figura 6	Esquema de uma classificação do algoritmo <i>Random Forest</i>	48
Figura 7	Cena utilizada para cobrir a área de estudo, localizada no Mato Grosso do Sul	70
Figura 8	Dados de precipitação (mm) diária da estação meteorológica de Paranaíba – MS, fornecidos pelo INMET para o período de 01/08/2016 a 31/08/2017	71
Figura 9	Fluxograma da metodologia utilizada no Capítulo 2	73
Figura 10	Segmentação gerada a partir da configuração – Escala: 120, Forma: 0,2 e Compacidade: 0,5 - do algoritmo <i>Multiresolution Segmentation</i> implementada no <i>software Ecognition</i>	87
Figura 11	Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação Vermelho	88
Figura 12	Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação NIR	88
Figura 13	Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação NDVI	89
Figura 14	Cenas utilizadas para cobrir área de estudo, localizadas no Mato Grosso do Sul	127
Figura 15	Precipitação (mm) diária da estação meteorológica de Paranaíba – MS, para o período de 01/08/2016 a 31/08/2017	129
Figura 16	Fluxograma da metodologia utilizada no Capítulo 3	130
Figura 17	Classificação 3 e tabela com a área ocupada por cada uma das classes, nesta classificação	152
Figura 18	Classificação 5 e tabela com a área ocupada por cada uma das classes, nesta classificação	152

LISTA DE ABREVIACÕES E SÍMBOLOS

ANN	Artificial Neural Network
AG	Acurácia Global
AP	Acurácia do Produtor
B	Faixa espectral do Azul (Blue)
AU	Acurácia do Usuário
CO ₂	Gás Carbônico
DT	Decision Tree
EVI	Enhanced Vegetation Index
G	Faixa espectral do Verde (Green)
GLCM	Gray Level Co-occurrence Matrix
IAC	Instituto Agronômico de Campinas
IBGE	Instituto Brasileiro de Geografia e Estatística
IFMG	Inventário Florestal de Minas Gerais
INMET	Instituto Nacional de Meteorologia
INPE	Instituto Nacional de Pesquisas Espaciais
Kg	Quilograma
LM	Learning Machine
LSWI	Land Surface Water Index
MIR	Middle Infrared
NDVI	Normalized Difference Vegetation Index
NIR	Faixa espectral do Infravermelho próximo (Near Infrared)
OBIA	Object-Based Image Analysis
OOB	Out Of Bag
OLI	Operational Terra Imager
PRODES	Projeto de Estimativa do Desflorestamento da Amazônia
R	Faixa espectral do Vermelho (Red)
REDD+	Reducing Emissions from Deforestation and Forest Degradation
REM	Radiação Eletromagnética
RF	Random Forest
RVI	Ratio Vegetation Index
SAM	Spectral Angle Mapper

SAVI	Soil Adjusted Vegetation Index
SIG	Sistema de Informações Geográficas
SR	Sensoriamento Remoto
SVM	Support Vector Machine
SWIR	Faixa espectral do Infravermelho médio (Short Wave-length Infrared)
Tg	Teragrama
TIRS	Thermal Infrared Sensor
UNFCCC	United Nations Framework Convention on Climate Change
UA	Unidade Amostral
VHR	Very High Resolution
%	Porcentagem

RESUMO: A Seringueira apresenta deciduidade na época seca, e essa característica pode tanto atrapalhar quanto auxiliar na identificação e separabilidade da classe em aplicações de sensoriamento remoto. Assim, o objetivo da presente dissertação foi explorar o potencial da introdução de atributos espaciais à classificação de base puramente espectral e apresentar uma metodologia, apoiada em uma abordagem orientada à objeto e técnicas automáticas de classificação digital, para identificar e caracterizar quatro classes de estudo: Seringueira, Eucalipto, Nativa e Outros Usos de uma região do Mato Grosso do Sul, com ênfase à Seringueira. O estudo foi feito em três capítulos. O Capítulo 1 foi destinado à revisão de literatura. No Capítulo 2 encontra-se uma minuciosa análise exploratória dos dados para avaliar a sensibilidade do classificador às Unidades Amostrais (UAs) coletadas e entender a interação entre os atributos e as características da área para escolher a melhor estratégia a ser utilizada. No Capítulo 3 foi explorada a capacidade de descrição de dois grupos de atributos espaciais e a utilização do *ranking* de importância dos atributos na análise dos atributos, buscando aumentar o poder de discriminação das quatro classes para geração de informações a respeito da heveicultura sul-mato-grossense. A metodologia dos dois artigos é similar. De imagens Landsat 8 foram gerados objetos utilizando o algoritmo *Multiresolution Segmentation*. A amostragem foi manual, 66% das UAs foram utilizadas para treinar o classificador *Random Forest* e as outras 34% para avaliar a acurácia. As características espectrais foram geradas através das médias da reflectância dos *pixels* por polígono, e as espaciais através dos máximos e mínimos das funções de semivariância. Os principais resultados do Capítulo 2 mostram que a inclusão da componente temporal a classificações puramente espectrais diminuiu a confusão da classe Seringueira com a classe Outros Usos, e que utilizar atributos espectrais e espaciais em conjunto permite melhor separação das classes em estudo. Já no Capítulo 3 foi observado que de maneira geral, os atributos espaciais mais interessantes a serem empregados em classificações espectrais-espaciais são os atributos do primeiro grupo, e que o grupo a qual os atributos espaciais utilizados pertencem é mais importante do que a quantidade de atributos espaciais e do que a fonte de informação utilizada.

Palavras-chave: sensoriamento remoto, random forest, *hevea brasiliensis*, landsat 8, semivariograma, heveicultura.

ABSTRACT: The Rubber Tree presents defoliation in the dry season, and this characteristic can both disturb and support class identification and separability in remote sensing applications. Thus, the aim of this work was to explore the potential of spatial attributes in purely spectral classification and to present a methodology based on an object oriented approach and automatic digital classification techniques, to identify and characterize four classes of study: Rubber Tree, Eucalyptus, Native and Other Uses present in a region of Mato Grosso do Sul, with emphasis on the Rubber Tree. The study was made in three chapters. Chapter 1 was destined to literature review. In Chapter 2 a data exploratory analysis was performed to evaluate the sensitivity of the classifier to the sample units (SUs) and to understand the interaction between the attributes and the features of the area in order to choose the best strategy to be applied. In Chapter 3 we explored the capacity to describe the classes of two groups of spatial attributes and the use of the importance ranking of attributes in the analysis of attributes, in order to increase the discrimination power of the four classes to generate information about the sul-mato-grossense heveiculture. The methodology of the articles is similar. Objects were generated from Landsat 8 images using the Multiresolution Segmentation algorithm. The sampling was manual, 66% of the SUs were used to train the Random Forest classifier and other 34% to evaluate the accuracy. The spectral characteristics were generated through the means of the reflectance of the pixels by polygon, and the spatial ones through the maximum and minimum of the semivariance functions. The main results of Chapter 2 have shown that inclusion of temporal component to purely spectral classifications has reduced the confusion between Rubber Tree and Other Uses classes, and additionally joint spectral and spatial analysis achieves higher separability for the studied classes. Finally in Chapter 3 was observed that in general the most interesting spatial attributes to be employed in spectral-spatial classifications are the attributes of the first group, and also that the group to which the spatial attributes used belong is more important than the number of spatial attributes and the source used.

Key words: remote sensing, random forest, *hevea brasiliensis*, landsat 8, semivariogram, heveiculture.

INTRODUÇÃO

Segundo Leal (2017) a borracha é considerada uma *commodity* estratégica para os mercados globais, e embora o Brasil seja o centro de origem da Seringueira atualmente o país apresenta pouca expressão no mercado mundial. Isso faz com que o país tenha pouca expressão nos fóruns de discussão sobre as tendências e acordos envolvendo o mercado mundial de borracha, que é dominado pelos países asiáticos. Esses países detêm mais de 80% da produção mundial de borracha natural, enquanto o Brasil, de acordo com Scaloppi *et al.* (2017), produz apenas cerca de 30% da demanda interna, tornando necessário uma importação média anual de mais de 500 milhões de dólares.

Entretanto, devido ao aumento da produção de borracha natural identificado no país na última década, o déficit interno entre oferta e demanda de borracha natural está prestes a ser modificado, ou ao menos demonstrar redução. De acordo com Nogueira (2015) o aumento na produção aconteceu devido a crescente demanda global por produtos que utilizam borracha natural em suas composições.

Para compreender o balanço entre oferta e demanda da borracha natural nacional é necessário entender a expansão dos plantios de Seringueira no país. Isso possibilita, dentre outras aplicações, melhor controle das previsões e metas de produção, assim como assistir e direcionar os produtores desde programas municipais até federais. No entanto, informações precisas a respeito dos plantios não estão disponíveis, o que limita substancialmente a compreensão da expansão das plantações de Seringueira.

Quando se trata de grandes áreas a aquisição de informações detalhadas e atualizadas é um problema através de levantamentos em campo. Esforços como esses demandam muito tempo, mão de obra e investimento. Assim, informações sobre usos do solo ganharam um novo enfoque, onde é atribuída grande importância ao mapeamento e monitoramento remoto dessas áreas. Li e Fox (2011) acreditam que a classificação de imagens de satélite é um dos principais meios de extração de informações sobre uso e cobertura do solo, podendo auxiliar e modificar o paradigma da coleta de dados e extração de informações.

No entanto, a classificação de imagens de satélite traz consigo alguns problemas que influenciam diretamente a identificação específica das classes. Dentre esses problemas o principal é a natureza ambígua de dados oriundos de sensores remotos. Objetos de naturezas diferentes podem apresentar *pixels* com valores idênticos

ou muito parecidos, e essa ambiguidade causa o aumento da complexidade da identificação, que pode ser maior ou menor, dependendo dos cenários analisados.

A Seringueira possui deciduidade, apresentando plantios completamente desfolhados em determinada época do ano. Essa dinâmica espectral é responsável pela diversificação de classes com que esses plantios se confundem espectralmente. Quando enfolhados se assemelham espectralmente a vegetação nativa e alguns outros usos do solo de cobertura arbórea. Já quando desfolhados a confusão espectral acontece em maior escala com áreas de vegetação rasteira como pastos e campos, por exemplo. Por isso separar plantios de Seringueira de objetos de vegetação arbórea e rasteira em imagens de satélite é um grande desafio.

Muitas técnicas de classificação foram desenvolvidas visando, sobretudo, a identificação de classes de uso do solo de forma automática na tentativa de diminuir a subjetividade causada pela natureza ambígua dos dados e pela interpretação, intrínseca à processos manuais de classificação. Dentre essas técnicas podemos destacar a incorporação de novos atributos na identificação, como os atributos espaciais que podem aumentar a acurácia de classificações de espécies florestais em imagens de satélite, e a utilização de algoritmos especializados como o *Random Forest* (RF), que constrói muitas árvores de decisão para classificação.

Para tanto se trabalhou com quatro classes de cobertura do solo: Eucalipto, Nativa, Seringueira e Outros usos. Para identificar essas classes assumiu-se que as faixas espectrais fornecidas pelo sensor OLI/Landsat 8 e o índice de vegetação NDVI (do inglês *The Normalized Difference Vegetation Index*, traduzido como Índice de Vegetação da Diferença Normalizada) podem ser utilizados como fontes de informação para extração de valores e geração de atributos espectrais e espaciais.

A hipótese testada é que a introdução de atributos espaciais ao processo de classificação de base espectral é capaz de aumentar a separabilidade da classe Seringueira das demais. Dessa forma, o presente estudo objetiva explorar o potencial da introdução de atributos espaciais em classificações de base espectral e apresentar uma metodologia, utilizando uma abordagem orientada à objeto e técnicas automáticas de classificação digital, para identificar seringais.

Como objetivos específicos têm-se:

- (a) Estudar a influência das fontes de informação propostas para extração de atributos espaciais (valores de semivariância) e espectrais à serem utilizados nas classificações;
- (b) Avaliar a eficiência da utilização conjunta e isolada de atributos espaciais e espectrais no processo de classificação de imagens digitais de média resolução espacial em um classificador não paramétrico;
- (c) Estudar a incorporação do componente temporal às análises por meio da utilização de imagens em duas datas distintas, representando os plantios de Seringueira enfolhados e desfolhados;

O documento foi estruturado em três seções apresentadas na forma de capítulos.

O Capítulo 1 trata-se de uma revisão de literatura na qual são encontradas informações relevantes a respeito da heveicultura e das técnicas utilizadas.

No Capítulo 2, intitulado “Atributos espectrais e espaciais no mapeamento de seringais – estudo de caso de uma área no Mato Grosso do Sul”, encontra-se uma minuciosa análise exploratória dos dados, realizada para selecionar a melhor estratégia e atributos a serem utilizados, visando atingir os objetivos aqui propostos.

No terceiro e último capítulo intitulado “Avaliação de grupos de atributos espaciais inseridos na classificação digital de base espectral de plantios de Seringueira no Mato Grosso do Sul”, encontra-se um estudo da aplicação das conclusões alcançadas no artigo anterior aliado à avaliação dos grupos de atributos espaciais propostos por Balaguer *et al.* (2010).

Espera-se que os resultados possam ser utilizados na espacialização dos dados a respeito da heveicultura sul-mato-grossense, visando contribuir com as informações a respeito da cultura, imprescindível para o planejamento estratégico da mesma.

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

1.1. Heveicultura

1.1.1. Contexto histórico e ambiental da cultura

Botanicamente a seringueira é uma dicotiledônea monóica do gênero *Hevea*, família Euphorbiaceae. No Brasil existem dez espécies das onze conhecidas, sendo todas as espécies lenhosas e arbóreas, com exceção da *Hevea camargoana*, que é arbustiva. Segundo Scaloppi Junior *et al.* (2017), a *Hevea brasiliensis* (Willd. ex ADR. de Juss.) Muell.-Arg., é a espécie mais importante do gênero do ponto de vista comercial. Originária da região amazônica é a principal fonte de borracha natural do mundo. Segundo Li e Fox (2012), a espécie se desenvolve bem em locais com chuvas contínuas durante os 12 meses do ano, dentro da zona equatorial entre as coordenadas 10° N e 10° S. No sudeste asiático, áreas com essas características incluem parte do sul de Mianmar, do Sul da Tailândia e sudeste do Vietnã.

Segundo Alvarenga e Do Carmo (2008) a planta inicia sua produção no 7º ano após o plantio em média. O declínio da produção acontece aproximadamente aos 40 anos de idade dos plantios e após esse período a madeira pode ser usada, dentre outros usos, como combustível para fornos e caldeiras. Pode ser agregado valor à madeira de Seringueira ao final do ciclo produtivo de látex através da produção de serrados em geral e outros produtos à base de madeira (CONEGLIAN; SEVERO, 2014).

A história do desenvolvimento econômico do Brasil aponta que entre os séculos XVI e XX (até 1930), a economia do país seguia um modelo produtivo em que a geração de renda era alicerçada na produção e exportação de algumas *commodities* agrícolas comercializadas no mercado internacional, o que a caracterizava como uma economia primário-exportadora (BAER, 1996). O ciclo da borracha aconteceu de 1666 a 1913, quando o país chegou a suprir 90% da demanda de borracha do mundo. Em 1910 a *commodity* contribuía com 40% das exportações brasileiras.

O país desfrutou da condição de principal produtor e exportador mundial até metade do século XX, passando a ser importador desta matéria-prima a partir de 1951 (GONÇALVES, 2013), que aconteceu devido ao aumento da oferta de látex impulsionado pela produção dos seringais introduzidos no oriente. A queda nos preços

internacionais da borracha desestimulou a produção interna e foi determinante para a drástica redução da participação brasileira na exportação desta *commodity*.

Segundo Scaloppi Junior *et al.* (2017), no Brasil o plantio da Seringueira fora do habitat natural se iniciou no estado de São Paulo em 1917. Na década de 1940 pesquisadores do Instituto Agrônomo de Campinas (IAC) observaram o bom desenvolvimento da Seringueira e iniciaram os primeiros estudos da espécie fora da Amazônia. Décadas mais tarde, confirmou-se o sucesso dessa cultura no estado de São Paulo, que se tornou e ainda é o principal produtor de borracha no país.

Na China, de acordo com Li e Fox (2012), o investimento em pesquisas com enfoque no crescimento da Seringueira em ambientes fora da faixa ótima, com temperaturas mais amenas e sazonalidade, começou no início dos anos 50, como forma de garantir o desenvolvimento econômico. Isso permitiu que a cultura se estabelecesse em áreas que se estendem até a faixa 22° N e foi um importante passo para o monopólio da *commodity* no mercado mundial pela Ásia por que, segundo Qiu (2009), além de plantios em locais com estações secas distintas, permitiu também o plantio da cultura em locais com elevação superior a 1000 metros, condição essa que ocorre na maior parte do sudeste asiático, onde a cultura não era encontrada naturalmente.

A expansão impulsionou nessa região muitos estudos (EKADINATA *et al.*, 2004; LI *et al.*, 2007; HU *et al.* 2008) enfocados na temática ambiental, por que segundo Li e Fox (2012) mais de um milhão de hectares foram convertidos para a cultura nas últimas décadas no sudeste asiático. De acordo com estudo realizado por Fox *et al.* (2009) a área de plantios de Seringueira na Ásia pode chegar a quadruplicar até 2050, e a transformação de florestas naturais com grande importância ecológica e áreas agricultáveis em extensos plantios de Seringueira pode afetar os processos ecológicos (SHRESHTA *et al.*, 2010), fluxos locais de energia (HU *et al.*, 2008), água (GUARDIOLA-CLARAMONTE *et al.*, 2010) e a dinâmica do carbono (LI *et al.*, 2008).

Leal (2017) acredita que a cultura da Seringueira contribui à fixação de CO₂, minimizando problemas como o aumento da emissão de gases que intensificam o efeito estufa. Para Scaloppi Junior (2014) os seringais são mais eficientes nessa prática do que as florestas naturais, e ainda segundo o autor, a madeira é matéria prima renovável evitando a degradação de florestas nativas, além dos recursos financeiros provenientes da venda da madeira que podem custear novos plantios no final do ciclo produtivo.

No Brasil produtores de borracha natural do Noroeste Paulista estão apostando no fortalecimento do mercado de crédito de carbono para essa cultura, a fim de conseguir uma fonte extra de renda. O comércio de créditos de carbono para a Seringueira ainda não se consolidou no país, mas futuramente pode ser uma importante fonte de renda para os produtores rurais, já que a cultura tem a capacidade de fixar grande quantidade de carbono atmosférico devido à produção de borracha natural que utiliza o carbono fixado como principal componente.

A cultura enfrenta muitos desafios no Brasil e para Scaloppi Junior (2014), a autossuficiência do país na produção de borracha natural requer políticas de crédito, preços, entre outros incentivos. Essas medidas garantiriam mais que o fornecimento de borracha natural, mas a sobrevivência de milhares de pequenas propriedades rurais que tem como base o trabalho familiar, por que para Alvarenga e Do Carmo (2006) a Seringueira proporciona uma rentabilidade atrativa ao agricultor, sendo adequada à pequena produção e à agricultura familiar, fixando populações no meio rural.

1.1.2. Balança comercial: oferta vs. demanda

De acordo com Pennacchio (2014) a produção mundial de borracha no começo de 2014 cresceu 1,2 %, enquanto o consumo teve um aumento de 4,0%, isso demonstra um aumento da demanda pela matéria prima, e segundo Esperante (2015), as previsões apontam que esse padrão continuará acontecendo, com a oferta mundial abaixo da demanda até 2018, voltando a se equilibrar nos anos entre 2019 e 2022.

Em 2015 de acordo com o *Malaysian Rubber Statistics* (2016), a produção mundial de borracha natural foi de 12,31 milhões de toneladas, valor menor que a produção da borracha sintética de 14,46 milhões de toneladas. A principal região produtora continua sendo a Ásia, com 95% da produção mundial e a China o principal mercado consumidor, mas esse país vem diminuindo suas encomendas.

Segundo dados do ABRABOR (2017) a produção mundial de borracha natural em 2016 foi de 12.081 mil toneladas, onde mais de 8 mil toneladas foram originadas do Sudeste Asiático, envolvendo países como a Tailândia (34,51%), Indonésia (26,55%) e Vietnã (8,54%). No mesmo ano o Brasil produziu 191 mil toneladas, que corresponde a cerca de 1,58% da produção mundial, ocupando a 7ª posição e sendo considerado um país com baixa expressão no mercado mundial, conforme consta na Tabela 1.

Tabela 1. Produção de borracha natural em 2016.

Posição	País	Produção (mil toneladas)	Percentual (%)
1	Tailândia	3.169	34,5
2	Indonésia	3.208	26,6
3	Vietnã	1.032	8,5
4	China	774	6,4
5	Malásia	674	5,6
6	Índia	624	5,2
7	Brasil	191	1,6
8	Cambodia	145	1,2
9	Filipinas	91	0,8
10	Sri Lanka	79	0,7
	Outros países	1.095	9,1
	Mundo	12.081	100

Fonte: Adaptado ABRABOR (2017).

No Brasil, 54% da borracha natural produzida é oriunda dos seringais do estado de São Paulo (IAC, 2015), e segundo dados do Levantamento Censitário de Unidades de Produção Agropecuárias (SÃO PAULO, 2015) na Regional Agrícola de General Salgado foi identificado um incremento de 6.000 ha na área cultivada com seringais entre 1995-1996 e 2007-2008.

Um levantamento da oferta e demanda da borracha natural seca no Brasil no período de 2005 a 2015 foi feito pelo IBGE (2017) e observou-se aumento de 61,23% na produção interna nos últimos 10 anos, que segundo Nogueira (2015) é justificada por programas de financiamento do governo com taxas de juros baixas para plantio de Seringueira, aumento da produção de pneus em fábricas brasileiras devido, principalmente, ao marketing ambiental utilizado por empresas do setor, objetivando a comparação da borracha natural com a sintética. Já a importação cresceu apenas 18,38%, no mesmo período. A demanda total aumentou 33%, mostrando que a produção brasileira em 2015 atendeu a aproximadamente 41,44% da demanda nacional.

Para Nogueira *et al.* (2015) o desafio é que a produção atinja pelo menos o suficiente para suprir as necessidades do mercado interno. De acordo com a produção agrícola do IBGE (2017), no período de 2005 a 2015 houve aumento de 28% na área cultivada com Seringueira no Brasil. Nesse período apenas três estados apresentaram aumento na produção de borracha seca, Minas Gerais, Goiás e Mato Grosso do Sul.

1.1.3. A heveicultura no Mato Grosso do Sul: Caracterização e Perspectivas

Durante as últimas décadas a vegetação nativa do estado do Mato Grosso do Sul foi gradualmente substituída por usos antrópicos, principalmente nas áreas de ocorrência do Cerrado e da Mata Atlântica. Segundo Sano (2008) restam apenas 32% de cobertura vegetal natural no estado. As coberturas vegetais antrópicas do Mato Grosso do Sul são representadas principalmente pelas pastagens cultivadas, e por isso segundo Sano *et al.* (1999), é o estado que apresenta o maior contingente bovino do Brasil.

Nos últimos anos houve aumento da agricultura semiperene da cana-de-açúcar para a produção de etanol e também o plantio de Eucalipto para a produção de celulose, que se intensificou a partir de 2007 (SILVA *et al.*, 2011), principalmente no leste do estado (ALMEIDA, 2012). A heveicultura também vem crescendo de modo expressivo no estado, e mesmo que a Seringueira esteja fora de seu habitat, o clima e a topografia da região são considerados adequados à cultura, apresentando enorme vantagem em relação ao estado de São Paulo por possuir menores custos de aquisição de terras. Outra vantagem é a proximidade da região Noroeste do estado de São Paulo, onde se encontram grande parte das indústrias de processamento de borracha natural.

Colombo (2009), em estudo técnico e econômico, considera que a cultura se estabelece como uma atividade lucrativa e sustentável, com boas perspectivas de crescimento da produção para o país e perspectivas para que num futuro próximo o país possa suprir a demanda interna. Nesse estudo fica evidente que a heveicultura na região é recente, mas também que o complexo produtivo da cultura na região se encontra em crescimento, em especial a partir dos anos 2009 e 2010, quando se iniciou forte campanha de divulgação de grupos empresariais interessados em se estabelecer na região leste do estado, devido às promissoras perspectivas identificadas.

Segundo Leal (2017) no estado de Mato Grosso do Sul a área colhida com látex coagulado cresceu 47% entre 2005 e 2015, sendo o município de Paraíso das Águas o maior representante, com uma participação de área total cultivada com Seringueira de quase 59%. Para a autora, o município de Cassilândia permaneceu com 25 ha, mas esse cenário se encontra prestes a ser modificado uma vez que existe no município um projeto informalmente batizado de “Cidade da Borracha”, que conta com um bairro planejado, oito milhões de pés plantados e uma usina de beneficiamento de borracha.

Foi da posição promissora da expansão da Seringueira no Estado do Mato Grosso do Sul, mesmo que inicial, a motivação para a elaboração da presente dissertação, que pretende fornecer informações-chave a respeito da distribuição dos plantios, informações essas que podem ser de grande valia para nortear tomadas de decisão e impulsionar a competitividade da cultura no Estado de Mato Grosso do Sul.

1.1.4. Sazonalidade da Seringueira

A Seringueira possui hábito decíduifólio, perdendo suas folhas periodicamente. Esse hábito é uma característica importante para a sobrevivência da espécie por que o desfolhamento natural faz com que haja a redução de inóculos de patógenos. Após o desfolhamento ocorre a renovação foliar, fase também conhecida como reenfolhamento. Não são todas as espécies que apresentam esse hábito e as três espécies que despertam maior interesse econômico, segundo Furtado (1996), diferem bastante quanto a troca de folhas. A *Hevea benthamiana* é parcialmente caduca, a *Hevea Pauciflora* tem folhas persistentes o ano todo e a espécie *Hevea Brasiliensis*, que é a representante mais importante do ponto de vista comercial, é considerada caduca.

A desfolha natural das espécies que possuem essa característica pode ser orientada tanto por fatores climáticos (precipitação fluvial, estação climática, fotoperíodo) como por fatores inerentes aos plantios (nutrição vegetal, idade da planta, material genético, estado sanitário...), mas a precipitação permanece como principal fator influenciando os padrões fenológicos. A desfolha segundo Pezzoni Filho *et al.* (2015a) acontece no final do mês de abril – maio, e a reenfolha segundo Pezzoni Filho *et al.* (2015b) ocorre no final de setembro – outubro.

De acordo com Gasparotto *et al.* (1997) para a *H. brasiliensis*, a periodicidade passa a acontecer após a formação da copa, que acontece a partir do terceiro ano de idade. A queda na maioria das vezes ocorre no início da estação seca e quanto mais rápido e pronunciado o decréscimo das chuvas, mais rápida e uniforme será a senescência e a queda das folhas. A duração desse período, intensidade e velocidade da troca de folhas varia de espécie para espécie e de clone para clone, sendo encontrados alguns híbridos com senescência tardia e mais escalonada.

1.2. Técnicas de Sensoriamento Remoto no monitoramento da vegetação

De uma forma resumida o Sensoriamento Remoto (SR) se baseia na interação da radiação eletromagnética (REM) proveniente do Sol, com os alvos terrestres. Segundo Jensen (2009) os dados são obtidos através de sensores remotos que registram a intensidade da energia eletromagnética refletida ou emitida pela superfície em determinados intervalos de comprimento de onda ou de frequência da REM. Esses intervalos de comprimentos de onda ou de frequência são denominados bandas espectrais. Uma banda espectral, ou faixa espectral, é um intervalo específico entre dois comprimentos de onda dentro do espectro eletromagnético solar (SPRING, 2006).

A vegetação, a água, o solo e outros tipos de cobertura da superfície terrestre refletem a REM incidente de maneira particular em diferentes faixas espectrais. Essas diferenças, por sua vez, resultam em diferentes padrões de cor, de tonalidade e de textura em imagens orbitais.

O aumento no número de sensores orbitais de observação da terra verificado nas últimas décadas facilitou o monitoramento de grandes áreas da superfície terrestre em intervalos de tempo relativamente curtos e de forma frequente, e as áreas do conhecimento que utilizam dados de SR sobre a vegetação foram beneficiadas por que com esse aumento houve crescimento exponencial de dados disponíveis e potenciais para o monitoramento e para a gestão de extensas áreas.

Muitos projetos e iniciativas foram iniciados em escalas locais, regionais e globais, com o objetivo de monitorar a vegetação. Em escalas locais foram desenvolvidos inúmeros trabalhos com objetivos, espécies e classes de estudo extremamente diversificados (WANG *et al.*, 2004; JUSOFF, 2007; JUSOFF, 2008; SANTORO *et al.*, 2010; WALKER *et al.*, 2010; LONGÉPÉ *et al.* 2011; MIETTINEN; LIEW, 2011; VICENTE *et al.*, 2012; DONG *et al.*, 2012a; DONG *et al.*, 2012b, GHOSH; JOSHI, 2014; SILVEIRA *et al.*, 2017).

Em escala regional o IFMG (Inventário Florestal do Estado de Minas Gerais) é um exemplo de como as imagens de sensores remotos podem auxiliar no monitoramento e na tomada de decisão. Segundo o IFMG (2008) o objetivo do projeto foi captar um conjunto de informações relacionadas aos recursos florestais para complementar, de forma essencial, o conjunto de dados fundamentais para o planejamento da gestão ambiental no estado. Através desse instrumento foi possível

identificar e determinar os principais atores e as principais causas que levam a alteração do uso do solo mineiro e ao empobrecimento de sua diversidade biológica.

Em escala nacional o projeto PRODES realiza o monitoramento por satélite do desmatamento por corte raso na Amazônia Legal, produzindo as taxas anuais de desmatamento na região, que são usadas pelo governo brasileiro, dentre outros usos, para o estabelecimento de políticas públicas. É com base nos dados do PRODES que o governo a cada ano define os municípios inseridos no Bioma Amazônia que são regidos pelas normas do Decreto 6.321 de 21/12/2007, que institui regras de acesso a créditos federais e impõe medidas de regularização fundiária e de redução de desmatamento para o restauro do acesso às linhas de crédito embargadas (INPE, 2013).

Abordando um conjunto de países, podemos citar a REDD+ (do inglês *Reducing Emissions from Deforestation and Forest Degradation*, que significa Redução de Emissões Decorrentes do Desmatamento e da Degradação de Florestas em Países em Desenvolvimento), que segundo o MMA (Ministério do Meio Ambiente) (2016) é um incentivo desenvolvido no âmbito da Convenção-Quadro das Nações Unidas sobre Mudança do Clima (*UNFCCC, do inglês United Nations Framework Convention on Climate Change*) para recompensar financeiramente países em desenvolvimento por resultados na redução de emissões de gases do efeito estufa provenientes do desmatamento e da degradação florestal, considerando o papel da conservação de estoques de carbono florestal, manejo sustentável de florestas e aumento de estoques de carbono florestal. Dentre os elementos para o reconhecimento de resultados de REDD+ se encontra “a criação de um sistema nacional robusto e transparente para o monitoramento florestal e o relato das atividades de REDD+”. A partir dessa exigência, é possível imaginar a importância de ferramentas que possibilitem a quantificação e o monitoramento dessas áreas, já que qualquer atividade de monitoramento de apoio a REDD+ exige, dentre outros fatores, a capacidade de estimar as mudanças em toda a área florestal de um país, para que a alocação de esforços seja otimizada, permitindo assim a quantificação dos resultados e o consequente pagamento por esses esforços.

A partir do exposto percebe-se que existem inúmeras aplicações e possibilidades de utilização para imagens de satélite no monitoramento de áreas de vegetação, sejam elas utilizadas para fins comerciais, tanto para espécies perenes como para espécies anuais, para fins de preservação e até mesmo para fins de organização da informação territorial e tomada de decisão.

Dessa forma, são necessárias também diferentes técnicas e recursos capazes de trabalhar com esses dados para extrair as informações necessárias. Dentre essas técnicas se destacou na literatura o uso de índices de vegetação, que segundo Kim *et al.* (2010) são largamente utilizados em estudos de detecção de mudanças na vegetação, incluindo entre essas mudanças não apenas as ocasionadas por pressões antrópicas, mas também as decorrentes de atividade sazonal da vegetação.

1.2.1. Índices de Vegetação

Segundo Epiphanyo *et al.* (1996) índices de vegetação são combinações entre dados provenientes de diferentes faixas espectrais que servem para realçar as feições de vegetação através da avaliação da refletância do dossel.

Para Ponzoni (2001) a resposta espectral da vegetação em determinado produto de SR é fruto de um processo complexo que envolve muitos parâmetros e fatores ambientais. O que é efetivamente medido de determinada vegetação por um sensor remoto não pode ser explicado somente pelas características intrínsecas dessa vegetação inclui também a interferência de vários outros fatores como a fonte de radiação, o espalhamento atmosférico, as características foliares e do dossel, os teores de umidade, a interferência da refletância do solo, sombra, entre outros.

Então, para extrair informações a respeito das superfícies a refletância tem sido transformada e combinada de várias formas possíveis para a geração de índices de vegetação. Esses índices podem também ser utilizados para correlacionar alguns parâmetros da vegetação, como: fitomassa, área foliar, produtividade, atividade fotossintética, fenologia da planta, entre outros (ELVIDGE; CHEN, 1995).

Segundo Rosa (2003), as faixas mais utilizadas nesses índices são a do vermelho e do NIR (do inglês *Near Infra Red*, que é traduzido como Infravermelho próximo), por conter mais de 90% da variação da resposta espectral da vegetação.

Existem muitos índices de vegetação, mas podemos destacar os principais segundo Eastman (1998) ao fazer uma revisão de literatura de vários autores. O autor destaca o EVI (do inglês *Enhanced Vegetation Index*, que significa Índice de Vegetação Melhorado), o SAVI (do inglês *Soil Adjusted Vegetation Index*, que significa Índice de Vegetação de Solo ajustado), o RVI (do inglês *Ratio Vegetation Index*, que significa Índice de Vegetação por Proporção), e o NDVI (do inglês *Normalized Difference*

Vegetation Index, que é traduzido como Índice de Vegetação da Diferença Normalizada), que dentre todos é o mais utilizado para aplicações na vegetação.

De acordo com Eastman (1998) o NDVI foi criado para separar a vegetação verde do solo de fundo. Minimiza efeitos topográficos ao produzir uma escala linear de medida. Seus valores variam de -1 a $+1$, quanto mais próximo de 1 maior a densidade da vegetação, onde 0 representa ausência de vegetação. A princípio não se espera valores negativos, a não ser para água, pois a refletância na região do vermelho é maior do que na região do NIR. Esse índice tem se mostrado bastante útil na estimativa de parâmetros biofísicos da vegetação e possui rápida saturação, o que o torna insensível ao aumento da densidade do dossel, fazendo com que os valores se estabilizem em um patamar independente do aumento da densidade.

De modo geral os índices buscam, além de identificar diferentes fitofisionomias, avaliar a resposta sazonal da vegetação. Demarchi *et al.* (2011) comparando o NDVI e SAVI no município de Santa Cruz do Rio Pardo-SP e utilizando imagens do sensor OLI/Landsat 5, concluíram que o NDVI e o SAVI se mostraram semelhantes e adequados na determinação das diferenças de vegetação da área de estudo quanto à densidade de cobertura vegetal e vigor vegetativo.

Li e Fox (2012) utilizaram o NDVI de imagens do sensor MODIS para monitorar a distribuição e o crescimento da Seringueira no sudeste asiático, e obtiveram que a utilização do NDVI pode auxiliar na superação do problema de superestimação de seringais, que havia sido detectado em estudos prévios.

Dong *et al.* (2013) utilizaram os índices de vegetação NDVI, EVI e LSWI para mapear seringais em imagens do sensor TM/Landsat 7. Os resultados mostraram que a sazonalidade típica da espécie em conjunto com os índices de vegetação possibilitou o delineamento de seringais dentro de paisagens florestais. Chen *et al.* (2016), também utilizaram os índices LSWI e NDVI para mapear plantios de Seringueira. O mapa resultante obteve acurácias, tanto do produtor quanto do usuário, maiores que 86%.

Chen *et al.* (2012a) utilizaram as faixas espectrais do sensor TM/Landsat 5, o NDVI e o RVI para desenvolver modelos de estimativa de idade de plantios de Seringueira em áreas que sofreram alguma perturbação natural. Como resultado obtiveram modelos com coeficientes de correlação variando de 0,74-0,82. Todos os modelos de regressão demonstraram superestimação de povoamentos jovens (com

menos de 25 anos) e subestimação de povoamentos antigos, devido às características de crescimento da cultura e à perturbação.

Li e Fox (2011) utilizaram o NDVI de imagens do sensor MODIS/Terra como uma das faixas espectrais em uma rede neural para a classificação de seringais, e os resultados mostraram que as precisões do produtor e o coeficiente Kappa para a classe Seringueira aumentaram quando foi introduzida a imagem NDVI na análise.

1.2.2. Segmentação: do *pixel* ao objeto

Análises orientadas a objeto, ou OBIA (do inglês *Object-Based Image Analysis*) têm sido estudadas e utilizadas desde a década de 70, mas para Hussain *et al.* (2013) foram os avanços nas tecnologias da informação geográfica alcançados nos últimos anos e o desenvolvimento de sistemas dedicados a OBIA, como o e-Cognition e o ENVI FX por exemplo, os responsáveis por trazer a tona essa abordagem no cenário de processamento de imagens.

Chen *et al.* (2012b) fizeram uma revisão sobre as principais questões relacionadas à detecção de mudanças da cobertura da terra, abordando as principais motivações para a utilização de abordagens orientadas a objeto. Já Hussain *et al.* (2013) discutem a OBIA de uma forma mais global, a técnica é comparada às tradicionais baseadas no *pixel* e as implicações de sua utilização ressaltadas, principalmente no que concerne ao contexto espacial que é ignorado na abordagem *pixel a pixel*.

Trabalhando com a Seringueira e comparando a análise padrão puramente baseada em *pixels* a OBIA, Ekadinata *et al.* (2004) obtiveram como principal resultado que uma classificação orientada a objetos permitiu maior acurácia e conseguiu extrair melhor características específicas relacionadas à complexidade da classe e à idade.

Para Fisher (1997), o *pixel* não pode ser considerado um objeto geográfico e sim uma representação de valores espectrais em uma grade cujos limites não possuem correspondência nenhuma com o mundo real. Em outras palavras, o *pixel* pode ser considerado uma medida do sensor imageador e não da área imageada. O agrupamento de *pixels* em uma determinada imagem cria porções com significados reais no campo.

Quando falamos de OBIA a principal hipótese é que os objetos originados na segmentação correspondam aos objetos na paisagem. Essa correspondência utópica perfeita depende de muitos fatores, mas principalmente da relação entre o objetivo do

trabalho, o tamanho dos objetos na imagem e a resolução espacial do sensor, e da correta configuração do algoritmo de segmentação. Esses fatores tornam praticamente impossível a obtenção de uma segmentação perfeita, visto que a segmentação funciona como uma representação da imagem, e não como a imagem em si.

A OBIA funciona mais próxima de como um analista de imagens trabalha, identificando padrões de objetos em uma imagem e não de *pixels* individuais. Possibilita às análises considerar as diferentes propriedades de um objeto, como tamanho, forma e a disposição espacial dos *pixels*, permitindo a exploração do contexto espacial.

Para Silveira *et al.* (2017) os objetos têm informações espectrais adicionais, quando comparados com *pixels* isolados, como média, valores mínimos e máximos, desvio padrão, etc, mas a maior vantagem é a informação espacial adicional, como por exemplo, características que são derivadas de funções geoestatísticas.

Assim, o primeiro passo é a determinação dos parâmetros adequados de segmentação, capazes de originar objetos com tamanhos e formas que melhor descrevam os objetos na imagem. Geralmente existem três parâmetros configuráveis, Forma, Compacidade e Escala, e dentre os três o mais crítico e delicado de configurar é o de Escala, que controla o tamanho dos objetos.

Segundo Benz *et al.* (2004), esse parâmetro estabelece um limiar de homogeneidade que determina o número de *pixels* vizinhos que podem ser unidos para formar um objeto. Para Dragut *et al.* (2010) o grau de heterogeneidade dentro de cada objeto é controlado por este parâmetro, mas para Arbiol *et al.* (2006) regiões semanticamente significativas são encontradas em diferentes escalas, salientando a ligação do parâmetro de escala da segmentação com os objetivos e objetos de estudo.

Smith (2010) destaca que em alguns casos pode não haver uma única configuração de segmentação ótima que seja ideal para identificar todas as classes de uma imagem, sendo a tentativa e erro a abordagem mais utilizada para a escolha dos parâmetros de segmentação, por que para Hussain *et al.* (2013) mesmo que diferentes metodologias tenham sido desenvolvidas e apresentadas para tratar dessa questão, uma metodologia consolidada para a determinação do parâmetro de escala ainda é um importante ponto a ser considerado em OBIA.

1.2.3. Sensoriamento Remoto na heveicultura

A caracterização e o mapeamento de plantios de Seringueira enfrentam muitos desafios, mas dentre eles Li e Fox (2011) destacaram os quatro principais:

1) a pequena área ocupada por seringais em comparação às outras classes mais comumente encontradas como, por exemplo, florestas nativas e plantios agrícolas;

2) a semelhança entre as características de reflectância de plantios maduros e de florestas nativas;

3) a mistura espectral detectada nos plantios jovens, podendo ser ocasionado pela mistura com o solo, ou ainda por misturas mais complexas envolvendo outras culturas econômicas de crescimento rápido que podem eventualmente, ser consorciadas com os plantios de Seringueira. Para os autores a espécie demora tempo considerável para fechar o dossel e conseqüente diminuir a mistura espectral. Plantios com 3-4 anos conseguem cobrir uma pequena parte do solo, fazendo com que os *pixels* de áreas com plantios jovens sejam frequentemente confundidos com outras classes;

4) a alta variabilidade intraclasse ocasionada por plantios de Seringueira com diferentes idades. A demora no fechamento do dossel resulta em diferenças estruturais entre os polígonos que são frequentemente classificadas como espécies diferentes.

Mas, a Seringueira possui uma característica de extrema importância para aplicações de SR, que é a deciduidade. Segundo Dong *et al.* (2013) espectralmente as espécies de Seringueira que apresentam deciduidade são distinguíveis de florestas nativas perenes nas duas fases fenológicas distintas, e a dinâmica espectral é ilustrada na Figura 1. Nesse estudo a correta seleção de imagens em fase fenológica apropriada para a classificação, especialmente para o delineamento da vegetação decídua, que é o caso da Seringueira, permitiu a geração de mapas com acurácias de 96%.

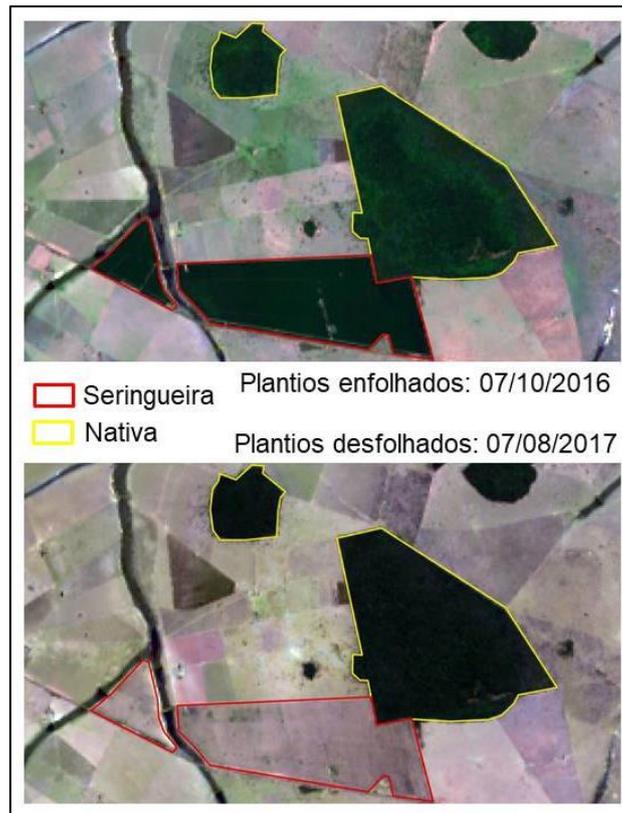


Figura 1. Plantios de Seringueira enfolhados e desfolhados, em comparação com floresta nativa.

Fonte: Dados da autora.

A idade dos plantios é um fator importante e por isso Chen *et al.* (2012) avaliaram a detecção de seringais com diferentes idades usando os dados do satélite *Landsat 5* em uma área frequentemente perturbada por furacões e baixas temperaturas. Os resultados indicaram que a correlação da banda NIR com a idade dos plantios diminuiu drasticamente com as perturbações ocasionadas pelos furacões.

Jusoff e Yusoff (2009) mapearam seringais na Malásia usando imagens de um sensor hiperespectral aerotransportado com resolução espacial de 1 metro e as correlações entre as áreas de copa das seringueiras e o diâmetro à altura do peito, e foi possível mapear copas individuais das seringueiras com uma precisão de 89,84%.

Para Lu e Weng (2007) o SR tem provado sua utilidade em várias aplicações, mas o sucesso de uma classificação satisfatória depende de vários fatores, incluindo entre os principais a escolha do procedimento de classificação adequado ao objetivo e dados disponíveis, assim como a época de aquisição das imagens, que segundo Moreira (2004) influencia muito nos resultados sobre os alvos na superfície terrestre.

Para Ponzoni *et al.* (2012) há ainda o processo de interação da REM com os objetos, que é normalmente expresso por apenas uma variável radiométrica. Esses valores são explicados pela ação de diferentes fatores como os físico-químicos dos objetos, e os geométricos de iluminação (fonte de radiação, normalmente o Sol) e visada. Esse caráter ambíguo da intensidade da REM refletida pelos objetos fragiliza as caracterizações espectrais em razão do acréscimo do grau de incerteza sobre os fatores que, efetivamente, explicam os valores dos fatores de reflectância medidos nos diferentes níveis de coleta possíveis, quer seja em laboratório, em campo, em nível de aeronave ou orbital (PONZONI *et al.*, 2012).

Mesmo com os inconvenientes e a ambiguidade, imagens de satélite têm sido amplamente utilizadas para o mapeamento de plantios de Seringueira nos últimos anos, mas poucos estudos se preocuparam com a caracterização de parâmetros de suporte a decisão e de planejamento da atividade, assim como com a distribuição e mapeamentos em escalas regionais que foi pouco estudada, principalmente para a região de estudo.

1.2.4. Reconhecimento de padrões: Espectral vs. Espacial

Ao tentarmos identificar objetos através dos padrões reconhecidos em imagens orbitais, cognitivamente são utilizadas características desse objeto que viabilizam a interpretação de modo mais ágil e acessível possível. Tradicionalmente, os processos de reconhecimento de padrões, em dados de sensores orbitais ou aerotransportados, utilizam essencialmente atributos com natureza espectral/radiométrica para formação de padrões estatísticos ou determinísticos no processo de identificação.

Mas, Para Chica-Olmo e Abarca-Hernández (2000) existem objetos que são espectralmente similares, mas que apresentam composições vegetais diferentes, ou ainda áreas contínuas com variação na densidade de árvores, por exemplo, que dificulta a diferenciação das classes vegetais e aumenta consideravelmente as incertezas associadas à classificação. Segundo os autores a maioria dos algoritmos utilizados na classificação de imagens não considera a dependência espacial entre os objetos e seus vizinhos, fazendo com que seja necessário complementar as informações fornecidas pelas bandas espectrais com outras características que permitam melhor separação entre as classes de cobertura do solo espectralmente semelhantes.

Informações conhecidas a respeito do alvo, como características contextuais, temporais e espaciais, por exemplo, podem ser agregadas ao processo para amparar a interpretação. Segundo Balaguer *et al.* (2010) informações espectrais utilizadas em conjunto com informações espaciais aumentam as acurácias de classificações de classes muito homogêneas devido ao efeito sinérgico exercido por um grupo de informação sobre o outro. Colaborando com esse raciocínio, resultados frequentemente encontrados na literatura, como por exemplo, em Berberoglu *et al.* (2007) e em Silveira *et al.* (2017), mostram que a utilização conjunta desses grupos realmente contribui de modo significativo no processo de reconhecimento de padrões.

O semivariograma é a principal ferramenta da geoestatística e sua utilização tem sido registrada no SR há anos, principalmente pela capacidade de extrair informações da organização espacial dos *pixels* na imagem, e por isso a maior parte dos trabalhos desenvolvidos se concentram na extração de atributos adicionais para a classificação, sendo utilizados os mais diferentes tipos de imagens e aplicações (MIRANDA *et al.*, 1992; EMRAN *et al.*, 1996; BISHOP *et al.*, 1998; BRUNIQUEL-PINEL; GASTELLU-ETCHGORRY, 1998; CARR; MIRANDA, 1998; MIRANDA *et al.* (1998); MIRANDA; FONSECA, 1996; BERBEROGLU *et al.*, 2007; BALAGUER *et al.*, 2010; PU; LANDRY, 2012; WU *et al.*, 2015; SILVEIRA *et al.*, 2017).

1.2.4.1. Geoestatística

A geoestatística surgiu na África do Sul, quando Krige (1951), trabalhando com dados de concentração de ouro, verificou que para encontrar sentido nas variâncias das Unidades Amostrais (UAs) em estudo era preciso considerar as distâncias entre elas, levando em consideração a localização geográfica e a dependência espacial. Para Matheron (1963) a geoestatística surgiu da inabilidade da estatística clássica em trabalhar com o aspecto espacial de um fenômeno, e baseado nas observações de Krige, desenvolveu uma teoria conhecida como Teoria das Variáveis Regionalizadas.

Assim, para a geoestatística cada ponto no espaço não apresenta um único valor, e sim uma distribuição de probabilidade de ocorrência de valores. Dessa forma, segundo Chilés e Delfiner (1999) a geoestatística objetiva fornecer informações quantitativas a respeito de variáveis naturais distribuídas no espaço, ou no tempo e

espaço. Em Van der Meer (2012) encontra-se um texto abordando o estado da arte da utilização da geoestatística no SR na literatura publicada entre os anos 2000 e 2010.

Na Teoria das Variáveis Regionalizadas a diferença nos valores de uma variável tomada em dois pontos diferentes depende da distância entre eles, então quanto maior a distância entre dois pontos, maior a diferença dos valores. Cada valor carrega uma forte interferência nos valores da vizinhança, ilustrando uma continuidade espacial. Segundo Matheron (1963) uma variável regionalizada é uma função espacial numérica que varia de um local para outro, com continuidade aparente e cuja variação não pode ser representada por uma função matemática simples, mas sim por um semivariograma.

O semivariograma é o gráfico que expressa a variação de alguma variável em função da distância. A variância total de um atributo é dividida em diferentes distâncias amostradas, conhecida como distância *lag*, objetivando encontrar a estrutura da variância. Se a variância aumenta com o aumento da distância, significa que a variância e a distância co-variam, indicando que são variáveis dependentes (SILVA, 1988).

Segundo Curran (1988) o semivariograma relaciona valores de semivariância no plano y, com a separação espacial entre os pontos, conhecida como distância *lag*, no plano x, conforme ilustrado na Figura 2.

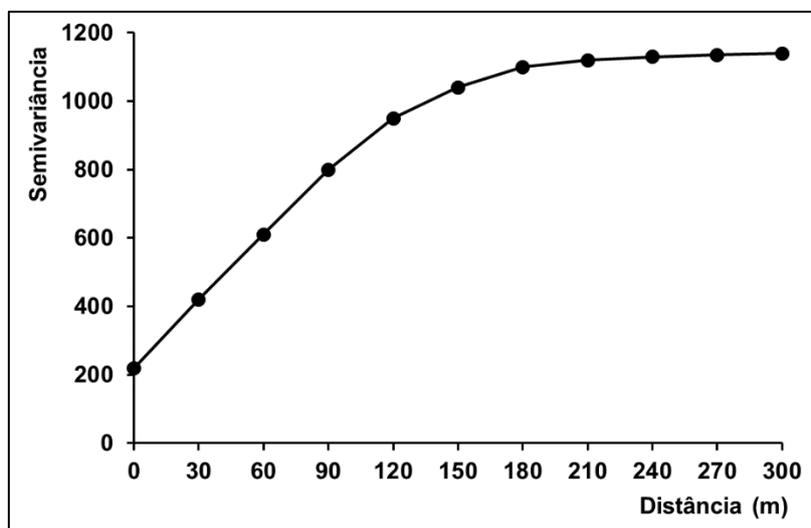


Figura 2. Semivariograma – Distância x Semivariância.
Fonte: Dados da autora.

A distância *lag* é a distância mínima necessária entre duas UAs para que sejam consideradas um par de observações aptas a participar da geração do semivariograma. Em OBIA o parâmetro de escala da segmentação determina o número de *pixels* vizinhos

que podem ser unidos para formar um objeto, ou seja, determina o tamanho dos objetos. Para a geração do semivariograma é necessário que os objetos tenham um tamanho mínimo, que é definido pela distância *lag*. Por tanto, essa distância é um critério crucial para a formação de um semivariograma e por isso é definido previamente.

O uso da função semivariograma no processamento de imagens tem sido amplamente aceito porque é considerada uma poderosa ferramenta para a análise da auto-correlação espacial de dados radiométricos e, mais especificamente da estrutura da variabilidade espacial (RAMSTEIN; RAFFY, 1989).

Segundo Bachmaier e Backes (2011) a definição do variograma teórico, γ , é baseada nas variáveis aleatórias regionalizadas $Z(x)$ e $Z(x + h)$, ou seja, dois pontos distintos no espaço, onde x e $x + h$ são as posições espaciais separadas pelo vetor h , a distância *lag*, com $\gamma(h)$ dependendo apenas dessa distância, e não da localização x .

Para Wu *et al.* (2015) a extração de atributos espaciais de semivariogramas pode ser feita de duas maneiras: Modelando o semivariograma através do ajuste de funções matemáticas (como por exemplo, modelos exponenciais, modelos Gaussianos e modelos esféricos) nos quais parâmetros do semivariograma são adotados como atributos espaciais. Esse método frequentemente é afetado pela escolha de uma função apropriada, por que funções simples não são suficientemente perceptíveis e complexas, podendo estar sujeitos a um superdimensionamento, além de modelos paramétricos serem sempre específicos; e utilizando o semivariogramas em vários *lags* (distâncias), não sendo afetado pelo problema causado pela modelagem, e por isso, é mais utilizado para descrever propriedades espaciais em imagens de sensores remotos.

Segundo Atkinson e Lewis (2000), um semivariograma experimental é definido como a metade da diferença média ao quadrado entre valores separados por uma dada distância, que é um vetor em ambas distâncias e direções, como expresso na Equação 1, que descreve um semivariograma para variáveis contínuas.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad [\text{Equação 1}]$$

,onde $Z(x_i)$ representa o valor da variável na localização x_i (no *pixel* i), h é a distância entre os elementos em uma dada direção (distância entre dois *pixels*) e $N(h)$ o número possível de pares de dados nas localizações x_i e $x_i + h$. Quando tratamos de imagens de satélite, os *pixels* são igualmente espaçados, então na equação acima

convencionamos que x_i representa o centro do *pixel* i , $Z(x_i + h)$ é igual a $Z(x_k)$, e x_k é localizado no centro do *pixel* mais próximo a localização $x_i + h$.

Nessa fórmula h é um vetor que indica se a função de semivariância é ou não anisotrópica, ou seja, se ela além da distância depende também da direção do vetor. Para tratar a direção do vetor h em dados de SR, Balaguer *et al.* (2010) propuseram a alternativa de calcular semivariogramas em 6 direções, variando de 0° a 180° , em intervalos de 30° . O semivariograma experimental é calculado através da média desses seis semivariogramas, e é conhecido por omnidirecional, como exposto na Equação 2:

$$\gamma(h_d) = \frac{1}{6} \sum_{i=0}^5 \gamma^{30i}(h_d) \quad [\text{Equação 2}]$$

, onde $\gamma^{30i}(h_d)$ é a representação do semivariograma definido na Equação 1, $30i$ indica a direção de início, em intervalos de 30° , enquanto h_d é a distância *lag* correspondente.

1.2.4.2. Caracterização do semivariograma

O semivariograma é caracterizado basicamente por três parâmetros, que auxiliam de maneira essencial à identificação dos elementos espaciais na imagem: Efeito Pepita (C0), Patamar (P) e Alcance (A), e podem explicar essa ciclicidade. Esses parâmetros estão exemplificados na Figura 3:

Patamar: valor onde o semivariograma tende a se estabilizar. Segunda Silva (1988) o patamar é atingido quando a variância dos dados se torna constante com a distância entre as UAs. Mostra quanto da variação total pode ser explicada pela organização dos valores na imagem, sendo o valor de semivariância nesse ponto aproximadamente igual à variância total dos dados. O patamar é dependente da classe de cobertura do solo, podendo fornecer informações úteis da estrutura da vegetação.

Alcance: Distância em *pixels* em que os dados são correlacionados. É a distância necessária para atingir o Patamar. Segundo Souza (1992), UAs separadas por distâncias menores que o alcance são espacialmente dependentes.

Efeito pepita: combinação dos erros de amostragem e das variações que ocorrem em escalas menores que a distância *lag*, indicando segundo Vieira (1997) uma variabilidade entre dois pontos não explicada pelo semivariograma. De acordo com Trangmar *et al.* (1985), um efeito pepita de 0% indica que o erro experimental é praticamente nulo e que não há variação significativa em distâncias menores que a

amostrada. Quanto menor a proporção do efeito pepita com relação ao patamar, maior é a semelhança entre os valores vizinhos e a continuidade do fenômeno, e menor a variância da estimativa e, portanto, maior a confiança nas estimativas (VIEIRA, 1998).

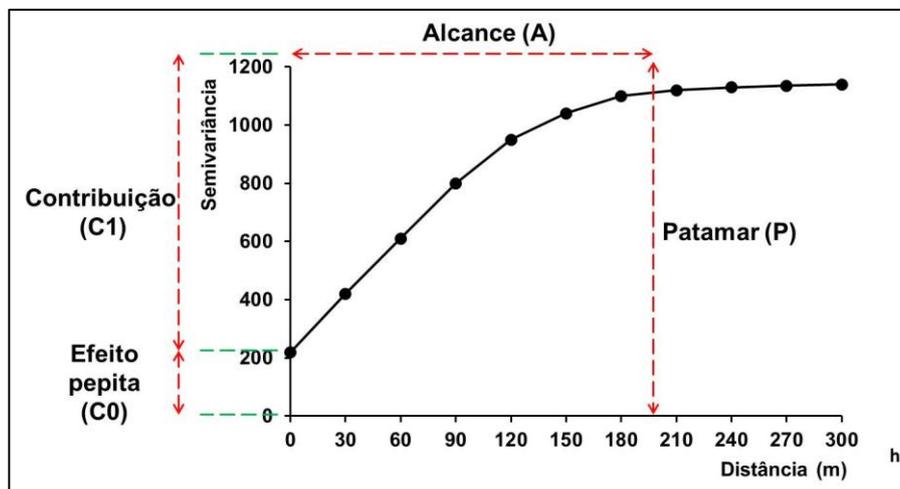


Figura 3. Semivariograma e seus parâmetros Patamar (C), Alcance (A) e Efeito Pepita (Co).

Fonte: Dados da autora.

Para o cálculo de um semivariograma por objeto o limite é a segmentação feita anteriormente, então dentro de um objeto é calculada uma sequência de valores de semivariância para diferentes distâncias *lag* (distâncias de *pixels*), e o conjunto desses valores de semivariância define a curva do semivariograma para cada objeto.

Segundo Ruiz *et al.* (2002) quando trabalhamos com OBIA cada objeto é caracterizado por um semivariograma omnidirecional, portanto, apenas os *pixels* totalmente inseridos no polígono são considerados no cálculo. A maior vantagem desse método é que o efeito de borda não afeta o resultado da caracterização dos objetos.

1.2.4.3. Aplicação de semivariogramas para a classificação

Atributos espaciais utilizam a organização espacial dos *pixels* para a geração de seus valores, por isso alguns estudos se concentraram na criação de metodologias e ferramentas para geração dos semivariogramas partindo de diversas abordagens e para diferentes aplicações (CHEN; GONG, 2004; SERTEL *et al.*, 2007; BALAGUER *et al.*, 2010; YUE *et al.*, 2013; WU *et al.*, 2015).

Dentre essas abordagens o grupo de atributos espaciais proposto por Balaguer *et al.* (2010) ganharam considerável destaque. Segundo os autores os parâmetros básicos do semivariograma são muito utilizados na classificação de imagens, mas eles têm difícil interpretação e aumentam muito a redundância dos dados. Eles propuseram 14 atributos com o objetivo de representar a variação espacial dos objetos em diferentes distâncias, utilizando os máximos e mínimos das funções.

Para Balaguer *et al.* (2010) uma das vantagens do conjunto de atributos proposto é que ao contrário dos valores de semivariância brutos as informações mais importantes a respeito da forma dos semivariogramas são resumidas em poucas características, que são capazes de identificar pontos importantes e realçar a informação contida nos primeiros *lags*, onde a correlação espacial é maior. Os autores dividiram os atributos propostos de acordo com a posição dos *lags* utilizados em sua definição como: Próximos a origem, Até o primeiro máximo e Entre o primeiro e segundo máximos. A definição dos atributos a seguir foi retirada do trabalho de Balaguer *et al.* (2010).

- Atributos próximos à origem: Resume alguns atributos do comportamento da curva próximo à origem do gráfico, como taxa de mudança no crescimento inicial da curva, inclinação, concavidade ou convexidade e nível de concavidade ou convexidade em curtas distâncias.

1. Razão entre variância total e a semivariância no primeiro *lag*:

$$RVF = \frac{\text{Variância}}{\gamma_i}$$

onde Variância é a variância total dos *pixels* do objeto. O semivariograma tende alcançar o patamar próximo a variância como mostrado na Figura 4, então esse parâmetro pode ser um indicador da relação entre a correlação espacial a longas e curtas distâncias. Esse valor tende aumentar quando ocorre alta variabilidade a longas distâncias, e baixa variabilidade a curtas distâncias. Essa característica apresenta altos valores para imagens com padrões periódicos.

2. Razão entre semivariância no segundo e primeiro *lags*:

$$RSF = \frac{\gamma_2}{\gamma_1}$$

representa a proporção do valor do segundo *lag* em relação ao primeiro *lag*. Dá informação a respeito das mudanças na variabilidade dos dados em curtas distâncias (apenas entre os dois primeiros *lags*).

3. Primeira derivada próxima à origem:

$$FDO = \frac{\gamma_2 - \gamma_1}{h}$$

representa a inclinação do semivariograma entre os dois primeiros *lags*, aproximando a primeira derivada da origem, conforme mostrado na Figura 4. Dá ideia da mudança na variabilidade dos dados em curtas distâncias.

4. Segunda derivada no terceiro *lag*:

$$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$$

é a aproximação da segunda derivada ao terceiro *lag* (Figura 4). Quantifica o nível de concavidade ou convexidade do semivariograma a curtas distâncias (entre 2° e 4° *lags*, quando $h = 1$), que corresponde à heterogeneidade dos objetos na imagem. Valores positivos indicam concavidade da curva, que em outras palavras significa que a imagem é homogênea a curtas distâncias. Valores negativos indicam convexidade da curva, significando que a imagem é heterogênea a curtas distâncias.

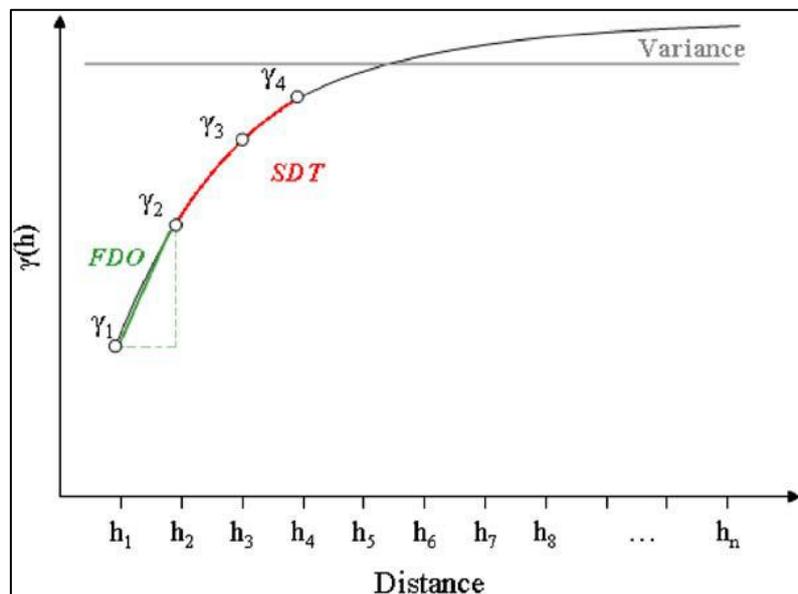


Figura 4. Atributos próximos a origem.

Fonte: Balaguer *et al.* 2010.

A utilização dos atributos dos outros dois grupos de características em conjunto (Até o primeiro máximo e Entre o primeiro e segundo máximos), são utilizadas para distinguir um semivariograma monótono de um cíclico.

- Atributos até o primeiro máximo:

5. Valor *lag* do primeiro máximo: FML

representa o valor *lag* onde a curva $\gamma(h)$ atinge o primeiro máximo local. Pode ser considerada uma aproximação do alcance do semivariograma. Quando se trata de curvas monótonas, essa é a distância máxima de análise para a geração do semivariograma. Esse valor é capaz de definir a “granulosidade” da imagem e é influenciado pela relação entre o tamanho médio das estruturas principais ou padrões na imagem, e sua separação um do outro.

6. Média dos valores do semivariograma até o primeiro máximo:

$$\text{MFM} = \gamma_{\text{max}_1}^{\text{média}} = \frac{1}{\text{max}_1} \sum_{i=1}^{\text{max}_1} \gamma_i$$

é um indicador da média dos valores do semivariograma entre o primeiro *lag* e o primeiro máximo. Informa as mudanças na variabilidade dos dados, e é relacionado à concavidade ou convexidade da curva nesse intervalo.

7. Variância dos valores do semivariograma até o primeiro máximo:

$$\text{VFM} = \frac{1}{\text{max}_1} \sum_{i=1}^{\text{max}_1} (\gamma_i - \gamma_{\text{max}_1}^{\text{média}})^2$$

é a variância dos valores do semivariograma entre o primeiro *lag* e o primeiro máximo. Complementa a informação do MFM. Tanto a média quanto a variância do semivariograma são diretamente ligadas à homogeneidade dos valores na imagem.

8. Diferença entre a média dos valores do semivariograma até o primeiro máximo (MFM) e a semivariância no primeiro *lag*:

$$\text{DFM} = \gamma_{\text{max}_1}^{\text{média}} - \gamma_1$$

mostra a taxa de diminuição da correlação espacial na imagem até os *lags* onde o semivariograma teoricamente tende a se estabilizar. Complementa a informação do RVF (atributo 1).

9. Relação entre a semivariância no primeiro máximo local e a média dos valores do semivariograma até esse máximo:

$$\text{RMM} = \frac{\gamma_{\text{max}_1}}{\gamma_{\text{max}_1}^{\text{média}}}$$

relação que também complementa a informação do RVF. Considera a influência da variabilidade total dos dados.

10. Diferença de segunda ordem entre o primeiro *lag* e o primeiro máximo:

$$\text{SDF} = \gamma_{\text{max}_1} - 2\frac{\gamma_{\text{max}_1}}{2} + \gamma_1$$

, se \max_1 é um número ímpar, a parte inteira de $\frac{\max_1}{2}$ é utilizada para calcular $\gamma_{\frac{\max_1}{2}}$.

Esse atributo dá informações a respeito da curvatura do semivariograma nesse intervalo, podendo representar também os valores com baixa frequência na imagem.

11. Área entre os valores do semivariograma no primeiro *lag* e a função do semivariograma até o primeiro máximo.

$$AFM = \frac{h}{2} (\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) + \gamma_{\max_1}) - (\gamma_1 (h_{\max_1} - h_1))$$

dá informações da curvatura do semivariograma e também é relacionado a variabilidade dos dados (Figura 5).

- Atributos entre o primeiro e segundo máximos:

Os atributos a seguir são úteis apenas quando as imagens possuem algum padrão periódico, por que esses atributos são tidas como nulos quando o semivariograma possui uma curva monótona. Quando existe apenas um máximo local na curva do semivariograma, então temos que: $h_{\max_1} = h_{\max_2}$, $\gamma_{\max_1} = \gamma_{\max_2}$

12. Distância entre a localização do primeiro máximo local e o segundo máximo local:

$$DMS = h_{\max_2} - h_{\max_1}$$

representa o tamanho da regularidade ou padrão estrutural da textura na imagem.

13. Distância entre o primeiro máximo e o primeiro mínimo:

$$DMM = h_{\min_1} - h_{\max_1}$$

complementa o DMS (atributo 12), adicionando informação a respeito da distância entre estruturas repetidas. A DMS e a DMM podem ser úteis na identificação da regularidade de curvas cíclicas.

14. Área da abertura:

$$HA = \frac{\frac{h}{2} (\gamma_{\max_1} + 2 (\sum_{i=2}^{\max_2-1} \gamma_i) + \gamma_{\max_2})}{\frac{1}{2} (h_{\max_2} - h_{\max_1}) (\gamma_{\max_2} + \gamma_{\max_1})}$$

aproxima a integral do semivariograma do h_{\max_1} para h_{\max_2} , que coincide com os *lags* onde o semivariograma atinge o primeiro e segundo máximos locais, respectivamente. Esse valor é dividido pela área total delimitada pela linha que conecta os dois máximos, para quantificar o peso da área da abertura proporcionalmente (região marcada na Figura 5). Dá informação a respeito do aumento/diminuição da depressão

entre os máximos, e quantifica o *hole effect*, que se relaciona diretamente à variabilidade, ou contraste, dos padrões de regularidade.

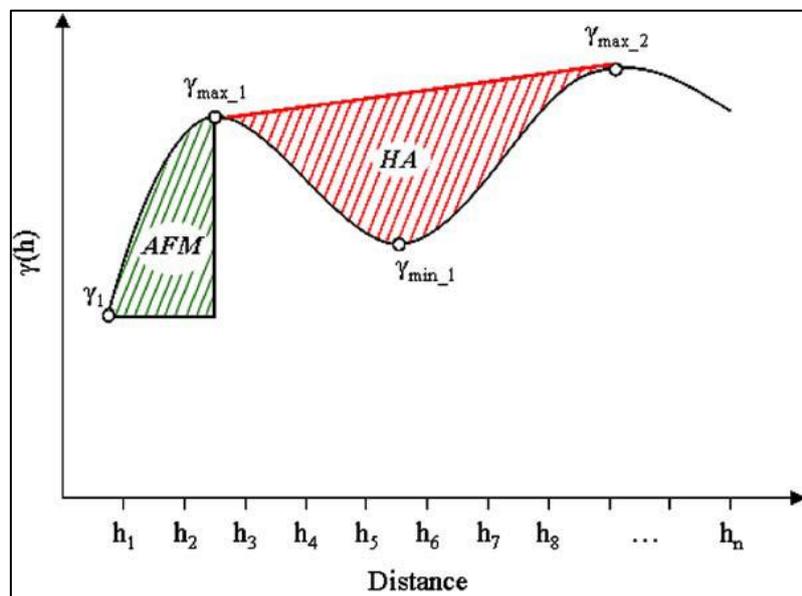


Figura 5. Atributos até o primeiro máximo e entre o primeiro e segundo máximos.
Fonte: Balaguer *et al.* 2010.

1.3. Random Forest

Uma grande variedade de métodos de classificação foi usada para mapear a cobertura da terra utilizando SR para diversas aplicações. Esses métodos variam de algoritmos mais simples até complexas máquinas com um grande número de parâmetros a serem configurados. São utilizados algoritmos não supervisionados, como por exemplo, o ISODATA ou K-means, algoritmos paramétricos supervisionados, como por exemplo, a máxima verossimilhança, conforme abordado por Jensen (2005).

Nesse contexto, LMs (do inglês *Learning Machines*, que significa máquinas de aprendizado) também se tornaram bem populares especialmente na última década, e o motivo segundo Rodriguez-Galiano *et al.* (2012) é a maior precisão e eficiência quando colocadas diante de grandes e complexos bancos de dados, assim como de grandes áreas. Para o autor esses classificadores apresentam superioridade de resultados, quando comparados a algoritmos paramétricos convencionais, principalmente por não dependerem de exigências relacionadas à distribuição dos dados, como por exemplo, a normalidade. Mas, algumas LMs são muito difíceis de configurar, necessitando de

ajuste em muitos parâmetros e tornando o processo de automatização complexo, como é o exemplo das ANNs e SVMs.

Dentre as principais LMs utilizadas podemos citar as ANNs, que foram extensamente discutidas e analisadas por Mas e Flores (2008), as DTs (do inglês *Decision Trees*, que significa árvores de decisão) desenvolvidas e discutidas por Breiman (1984), SVMs (do inglês *Support Vector Machines*, que significa máquinas de apoio a decisão) analisadas por Mountrakis *et al.* (2011) e a utilização de classificadores em conjunto, prática iniciada por Breiman (1996).

Rogan *et al.* (2008) comparando o desempenho da classificação de mudanças na cobertura terrestre de três algoritmos de LM: duas DTs e uma ANN, encontraram que para esse fim, a ANN foi mais apropriada, se mostrando mais robusta e precisa para o monitoramento de grandes áreas. Já Lippitt *et al.* (2008) obtiveram resultados opostos concluindo que para conjuntos de amostras de treinamento deficientes, as DTs se mostraram mais robustas e precisas, independente da regra de divisão utilizada.

Esse detalhe é de extrema importância para análises florestais, já que na maioria das vezes, conjuntos de dados de treinamento para classificação de florestas são muito pequenos ou não representam toda a variabilidade espectral de classes com grande variabilidade natural e/ou que agregam grande número de distúrbios antrópicos (LIPPITT *et al.*, 2008). Para Hansen e Reed (2000) as LMs, no entanto são capazes de lidar com áreas espectralmente complexas, com a mínima intervenção humana possível e com tempo de processamento reduzido drasticamente, quando comparadas a classificadores paramétricos tradicionais.

Dentre os sistemas de LMs conhecidos, recentemente houve grande aumento na utilização de conjuntos de aprendizagem conhecidos como *ensemble learning* (que significa conjuntos de aprendizagem), e segundo Zhou (2012) são considerados um paradigma de LM, onde várias unidades de aprendizado são treinadas para resolver o mesmo problema, diferente de abordagens de aprendizado clássicas, que objetivam aprender uma hipótese a partir dos dados de treinamento.

Conjuntos de aprendizado tentam construir um conjunto de hipóteses que depois são combinadas, por que para Oshiro (2013) uma combinação de opiniões leva, na maior parte das vezes, a uma decisão melhor do que uma decisão tomada por um único indivíduo. Um exemplo desses conjuntos de classificadores é encontrado no trabalho de Sesnie *et al.* (2008) que utilizaram DTs para mapear 32 classes de

vegetação, e Chan e Paelinckx (2008) que investigaram a possibilidade de usar imagens aéreas hiperspectrais para a classificação de ecótipos utilizando algumas DTs.

Os conjuntos para geração de DTs mais conhecidos e utilizados são o *Boosting*, desenvolvido por Freund e Schapire (1996), o *Bagging*, criado por Breiman (1996) e o mais recentemente de todos, e que tem ganho muita notoriedade na comunidade científica, é o RF (do inglês *Random Forest*, que significa Floresta aleatória) que também foi desenvolvido por Breiman (2001). Segundo o autor, é um algoritmo para classificação e regressão que combina os resultados de um grande número de árvores para atingir a classificação final de um conjunto de dados. No algoritmo RF, cada árvore é gerada a partir de uma amostra *bootstrap*.

Amostragem *bootstrap* é uma técnica de amostragem com reposição a partir da amostra inicial, onde são selecionadas, aleatória e independentemente, observações para um novo subconjunto de treinamento, a amostra *bootstrap*. Dessa forma, cada árvore cresce a partir de uma subamostra diferente da amostra de treinamento, e no final é feita uma votação simples por maioria para a previsão da classificação. O conjunto de dados não utilizados na amostra *bootstrap* é conhecido como amostra OOB (do inglês *Out Of Bag*, que significa fora do conjunto) e é utilizado para testar o modelo da árvore através da geração de uma estimativa da taxa de erro. Em média cada dado do conjunto total fica fora da amostra *bootstrap* 36% das vezes, e com base nessa informação a estimativa da taxa de erro OOB pode ser calculada.

No RF as árvores são construídas selecionando aleatoriamente ‘m’ atributos em cada nó de uma árvore. Os nós são os pontos de decisão das árvores, e os atributos são permutados entre si nesses nós. Dessa forma centenas, ou até milhares, de árvores são geradas formando assim uma ‘floresta’, que é o conjunto de classificadores.

As árvores são construídas usando um novo subconjunto amostral e uma seleção totalmente aleatória de atributos, e para Smith (2010) essa rotina de dupla aleatorização origina várias características desejáveis como resistência ao super ajuste e a capacidade de estimar internamente a acurácia potencial da classificação, sem a necessidade de uma amostra adicional para esse fim. A acurácia interna estimada é gerada através da aplicação das regras de decisão definidas pela amostra *bootstrap* à sub amostra OOB para cada árvore, e a classificação final de um objeto é baseada na classificação da maioria das árvores geradas para esse objeto, conforme ilustrado na Figura 6.

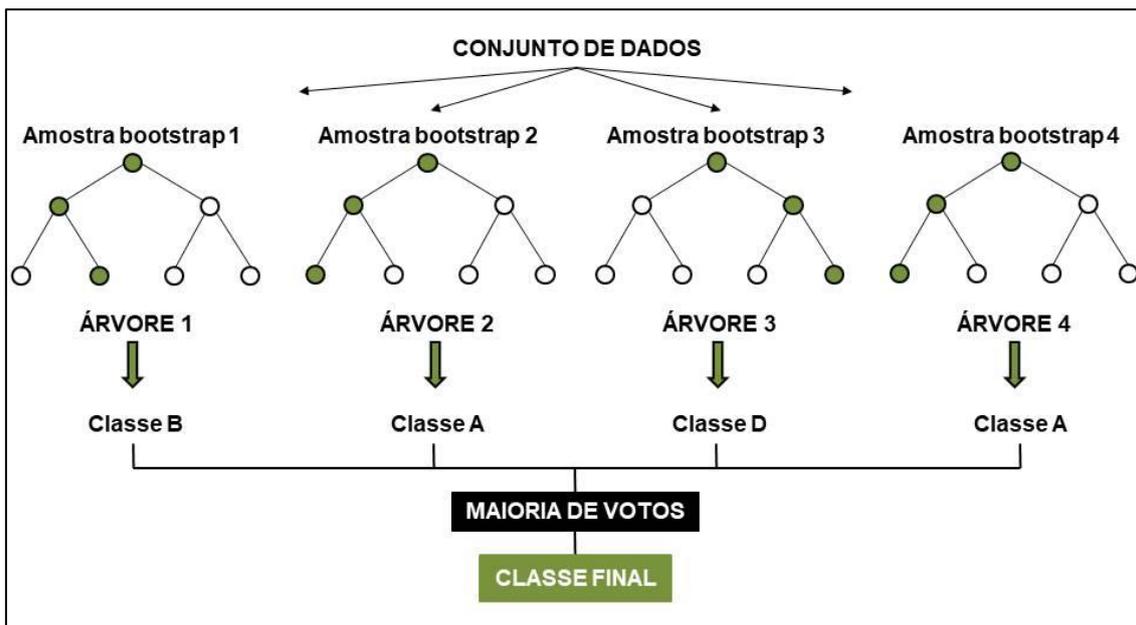


Figura 6. Esquema de uma classificação do algoritmo *Random Forest*.
Fonte: Dados da autora.

O RF é uma LM supervisionada, e formalmente nesse tipo de aprendizado é fornecida uma característica do objeto a ser alcançada, ou seja, o algoritmo recebe o valor da saída desejada para a entrada apresentada (OSHIRO, 2013). Em outras palavras, nesse tipo de LM é fornecida uma amostra (entrada) e as classes a que os objetos dessas classes pertencem (saída), para que o algoritmo possa aprender a classificar o caso e as classes em questão.

Belgiu e Dragu (2016), fizeram uma revisão a respeito da utilização do RF no SR. Esse estudo revelou que tal classificador pode lidar com grande sucesso com a alta dimensionalidade e multicolinearidade dos dados de entrada, de uma maneira rápida e sensível ao superajuste.

Devries *et al.* (2016) acreditam que as principais vantagens de uma RF são a excelente acurácia da classificação, a eficiente implementação em grandes conjuntos de dados e uma estrutura que permite o uso de árvores geradas em posteriores aplicações.

Segundo Breiman (2001) o algoritmo é fácil de implementar pois possui apenas dois parâmetros a serem configurados, o número de variáveis a serem testadas e utilizadas na amostra *bootstrap* para a divisão de cada nó, o *Mtry*, e o número de árvores na floresta, o *Ntree*. O algoritmo não se mostra muito sensível a esses valores, mas dentre os dois parâmetros o *Ntree* é o que desperta menor sensibilidade.

1.3.1. De uma *Random Tree* à uma *Random Forest*

Considerando um conjunto de treinamento ‘T’ com ‘a’ atributos e ‘n’ observações, seja T_k uma amostra *bootstrap* extraída a partir de T, contendo n exemplos e usando m atributos aleatórios ($m \leq a$) em cada nó das árvores. Uma *Random Tree* é uma árvore produzida aleatoriamente a partir de um conjunto de árvores possíveis, usando ‘m’ atributos aleatórios em cada nó, onde o termo aleatório quer dizer que cada árvore tem uma chance igual de ser amostrada.

Random Trees podem ser geradas eficientemente e a combinação de grandes conjuntos geralmente leva a modelos precisos (OSHIRO, 2013). A cada nó da árvore um subconjunto de ‘m’ atributos é selecionado aleatoriamente e avaliado. O melhor atributo é então escolhido para dividir o nó. O valor ‘m’ é fixado para todos os nós, e nesse classificador, as árvores crescem sem poda.

Ao contextualizar na aplicação de classificação orientada a objeto as árvores ‘votam’ na classe a qual o objeto pertence e em seguida a classe com o maior número de votos recebe o objeto em questão, e assim acontece para todos os objetos.

O resultado de uma RF são muitas DTs, e segundo Oshiro *et al.* (2012) conforme o número de árvores geradas vai crescendo, a floresta tende a usar todos os atributos da base de dados. Porém isso não é algo útil, uma vez que nem todos os atributos existentes são realmente importantes e capazes de agregar acurácia à classificação. Por esse motivo Liaw e Wiener (2002) desenvolveram uma métrica para identificar os melhores atributos e conseqüentemente, o atributo que seria escolhido em um nó para compor a árvore final. De acordo com os autores o RF produz outras duas informações adicionais, uma medida de importância dos atributos de predição e uma medida da estrutura interna dos dados. A importância dos atributos é um conceito com definição difícil, pois a importância de forma isolada não pode ser expressa, devido a sua interação com os outros atributos que acontece de forma complexa.

Em uma DT cada nó contém um teste para algum atributo e cada ramo descendente corresponde a um possível valor deste atributo. O conjunto de ramos é distinto e cada folha está associada a uma classe, sendo cada percurso da árvore, da raiz à folha, correspondente a uma regra de classificação (GAMA, 2004). O critério utilizado para as partições é o da utilização do atributo para a classificação. Aplica-se por este critério um determinado ganho de informação a cada atributo. O atributo

escolhido como atributo teste para o corrente nó é aquele que possuir maior ganho de informação, onde a partir deste inicia-se um novo processo de partição (SILVA, 2005).

Segundo Onoda (2001), quando a árvore é utilizada para classificação os critérios de partição mais conhecidos são baseados na entropia e índice de Gini, mas o RF utiliza a entropia, que segundo Mitchell (1997) é o cálculo do ganho de informação baseado em uma medida utilizada na teoria da informação. A entropia caracteriza a impureza dos dados e em um conjunto de dados é uma medida da falta da homogeneidade dos dados de entrada em relação a sua classificação.

Assim, o RF estima a importância de um atributo observando quanto o erro de previsão aumenta quando os dados OOB desse atributo são permutados, mantendo inalterados todos os outros. Se o resultado não é afetado através da distorção forçada da relação entre o atributo e sua resposta o atributo é considerado de baixa importância. Os cálculos são realizados árvore por árvore à medida que a floresta cresce, e uma árvore com taxa de erro baixa é considerada um classificador forte, possibilitando que a taxa de erro da floresta diminua, devido ao entendimento da força das árvores individuais.

A medida de proximidade utiliza a matriz de proximidade produzida pelo RF, onde (i, j) é a fração de árvores onde os elementos i e j aparecem ao mesmo tempo no mesmo nó terminal. Observações similares devem por lógica aparecer nos mesmos nós terminais com maior frequência do que observações díspares, e quanto maior a disparidade entre duas observações mais raro é o aparecimento das duas observações no mesmo nó terminal. A matriz de proximidade pode ser usada para identificar a estrutura da organização dos dados conforme observado em Breiman (2002).

Então, conforme o exposto o RF é um classificador menos sensível a qualidade das UAs de treinamento e ao superajuste do que outras LM simplificadas, devido ao grande número de DTs que são produzidas a partir da seleção aleatória do conjunto de UAs de treinamento e do conjunto de UAs dos atributos que serão utilizados para dividir as árvores em cada nó (BELGIU; DRAGU, 2016).

1.3.2. O *Random Forest* na classificação

O RF está sendo cada vez mais utilizado na classificação do uso do solo de imagens de sensores remotos por produzir classificações excelentes e apresentar tempos de processamento interessantes. Ghimire *et al.* (2010) utilizaram o RF para classificar

uma área heterogênea composta por 8 classes, dentre elas 4 de vegetação. O objetivo foi testar um novo método para incorporar a dependência espacial local em análises de dados espectrais e para os autores o algoritmo tem demonstrado aumentar a acurácia de classificações de uso do solo por que o erro de classificação de uma árvore pode ser superado pelo conjunto de árvores.

Wessels *et al.* (2016) avaliaram o LALCUM, um sistema rápido de ‘update’ da cobertura da terra para um ano específico, que utiliza o RF como classificador, e tem como base um mapa de referência de cobertura da terra preexistente.

Rodriguez-Galiano *et al.* (2012) exploraram o desempenho do RF na classificação de 14 diferentes classes de cobertura do solo de uma área complexa da Espanha e avaliaram a sensibilidade do algoritmo à redução do tamanho do conjunto de dados e ao ruído. Obtiveram-se classificações com acurácia global de até 92%, e os testes demonstraram que o RF é resistente à redução de dados e ao ruído, pois só foram observadas diferenças consideráveis nos valores Kappa quando a redução de dados chegou a 50% e a adição de ruído a 20%. Outro resultado foi que os atributos identificados pelo RF como mais importantes para a classificação da cobertura do solo da região coincidiram com o esperado.

O RF é considerado o estado da arte no campo da classificação de uso do solo, mas como a maioria dos métodos de classificação, não utiliza a dependência espacial entre objetos vizinhos. Essa dependência pode ser utilizada para aumentar as acurácias das classificações de uso do solo, minimizando assim a variação interna das classes.

Na maioria das aplicações a incorporação da dependência espacial às análises é encontrada na literatura através de bandas espectrais transformadas em informações espaciais de organização dos *pixels* para classificação de imagens, com o objetivo de capturar a variabilidade local dos números digitais em uma vizinhança utilizando-se da geoestatística, conforme observado em Atkinson e Lewis (2000), e auto-correlações espaciais conforme em Myint *et al.* (2007), ou através de filtros de suavização por maioria aplicada a imagens já classificadas, como em Mather (2004). Ghimire *et al.* (2010) estudaram uma nova maneira de utilizar a dependência espacial em regiões heterogêneas utilizando o RF e uma medida de dependência espacial conhecida por estatística ‘Getis’. Já Vogels *et al.* (2017), apresentaram um método semiautomático para mapeamento de áreas cultivadas em fotografias em preto e branco.

Métodos de classificação da cobertura do solo foram exaustivamente estudados ao longo dos anos e são amplamente utilizados, mas há algumas ressalvas quanto à utilização de imagens com média resolução espacial (resolução espacial de 15 a 30 m). Para Blaschke *et al.* (2006) em pequenas escalas de análises frequentemente podem ser encontrados *pixels* maiores do que a classe de cobertura, o que causa mistura espectral das classes. Rogan e Miller (2006) alertam ainda que há também problemas relacionados a grandes áreas, por que quando os alvos são extensas áreas, os produtos cartográficos podem ter problemas de precisão espacial devido às características de paisagens heterogêneas, como deslocamento topográfico, gradientes de umidade e distúrbios passados, sejam eles naturais ou antrópicos. O RF ajuda contornar esses problemas originando classificações em escala regional com maior estabilidade frente a variações na reflectância das classes, causadas por gradientes ou perturbações, e alta variabilidade intraclasses, que pode ser resultado de paisagens heterogêneas.

Orientado ao estudo da Seringueira, segundo Li e Fox (2012) quando são utilizados classificadores rígidos para identificar classes de vegetação há uma tendência à superestimação de áreas plantadas com Seringueira, quando enfrentam as realidades de paisagens heterogêneas com alta variação intraclasses. Isso faz com que o RF seja uma interessante opção para estudos voltados à identificação dessa classe.

REFEREÊNCIAS

- ABRABOR – ASSOCIAÇÃO BRASILEIRA DE PRODUTORES E BENEFICIADORES DE BORRACHA NATURAL. **Estatísticas e Tendências da Borracha Natural**. Brasília-DF: Camara Setorial da Cadeia Produtiva da Borracha Natural. 2017. Disponível em: < <http://www.agricultura.gov.br/assuntos/camaras-setoriais-tematicas/documentos/camaras-setoriais/borracha-natural/2017/38a-ro/app-abrabor-38ro-borracha.pdf>> . Acesso em 29 jul. 2018.
- ALMEIDA, R. A. Territorialização complexo eucalipto-celulose-papel em Mato Grosso do Sul. In: Encontro Nacional de Geografia Agrária, 11., 2012, Uberlândia-MG. **Anais...**: Universidade Federal de Uberlândia, 2012.
- ALVARENGA, A. P.; DO CARMO, C. A. F. S. **Sequestro de Carbono**: Quantificação em Seringais de Cultivo e na Vegetação Natural. Viçosa-MG: UFV, 2006. 338 p.
- ALVARENGA, A. P.; DO CARMO, C. A. F. S. **Seringueira**. Viçosa-MG: EPAMIG, 2008. p. 893 p.
- ARBIOL, R.; ZHANG, Y.; PALÀ, V. Advanced classification techniques: a review. In: ISPRS COMMISSION, 7., MID-TERM SYMPOSIUM REMOTE SENSING: FROM PIXELS TO PROCESSES, 2006, Enschede-NL. **Issues...** Enschede-Overijssel: ISPRS – International Society for Photogrammetry and Remote Sensing, 2006, p. 292–296.
- ATKINSON, P. M.; LEWIS, P. Geostatistical classification for remote sensing: na introduction. **Computers & Geoscience**, Houston-TX, v. 26, p. 361–371, 2000.
- BACHMAIER, M.; BACKES, M. Variogram or Semivariogram? Variance or Semivariance? Allan Variance or Introducing a New Term? **Mathematical Geoscience**, Limerick-Munster v. 43, p. 735–740, 2011.
- BAER, W. **A Economia Brasileira**. São Paulo-SP: Nobel, 1996. 416 p.
- BALAGUER, A.; RUIZ, L. A.; HERMOSILLA, T.; RECIO, J. A. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers & Geoscience**, Houston-TX, v. 36, p. 231-240, 2010.
- BATISTA, M. H. **Classificação hierárquica orientada a objeto em imagens de alta resolução especial empregando atributos espaciais e espectrais**. 2006. 199 p. Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre-RS, 2006.
- BELGIU, M.; DRAGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 114, p. 24–31, 2016.
- BENZ, U. C.; HOFMANN, P.; WILLHAUCK, G.; LINGENFELDER, I.; HEYNEN, K. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready

information. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 58, p. 239–258, 2004.

BERBEROGLU, S.; CURRAN, P. J.; LLOYD, C. D.; ATKINSON, P. M. Texture Classification of Mediterranean Land Cover. **IEEE International Journal of Remote Sensing**, London-UK, v. 9 n. 3, p. 322–334, 2007.

BLASCHKE, T.; BURNETT, C.; PEKKARINEN, A. Image segmentation methods for object-based analysis and classification. In: JONG, S. M. D., MEER, F. D. V. **Remote Sensing Image Analysis: Including the Spatial Domain**. Dordrecht-NL: Springer, p. 211–236, 2006.

BISHOP, M. P.; SHRODER JR, J. F.; HICKMAN, B. L.; COPLAND, L. Scale dependent analysis of satellite imagery for characterization of glacier surfaces in the Karakoram Himalaya. **Geomorphology**, Xangai-CHN, v. 21, p. 217–232, 1998.

BREIMAN, L. Bagging predictors. **Machine Learning**, Bristol-UK, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and Regression Trees**. Belmont-CA: Chapman and Hall/CRC, 1984. 358 p.

BREIMAN, L. Manual on setting up, using, and understanding random forests v3.1, 2002.

BREIMAN, L. Random forests. **Machine Learning**, Bristol-UK, v. 45, n. 1, p. 5–32, 2001.

BRUNIQUEL-PINEL, V.; GASTELLU-ETCHEGORRY, J. P. Sensitivity of texture of high resolution images of forest to biophysical and acquisition parameters. **Remote Sensing of Environment**. Toronto-ON, v. 65, p. 61–85, 1988.

CARR, J. R., MIRANDA, F. P. The semivariogram in comparison to the co- occurrence matrix for classification of image texture. **IEEE Transactions on Geoscience and Remote Sensing**, Pasadena-CA, v. 36, p. 1945–1952, 1998.

CARR, J. R. Spectral and textural classification of single and multiple band digital images. **Computers & Geoscience**, Houston-TX, v. 22, p. 849–865, 1996.

CHAN, J. C.-W.; PAELINCKX, D. Evaluation of Random Forest and Adaboost treebased ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. **Remote Sensing of Environment**. Toronto-ON, v. 112, n. 6, p. 2999–3011, 2008.

CHEN, B.; CAO, J.; WANG, J.; WU, Z.; TAO, Z.; CHEN, J.; YANG, C.; XIE, G. Estimation of rubber stand age in typhoon and chilling injury afflicted area with Landsat TM data: A case study in Hainan Island, China. **Forest Ecology and Management**, Flagstaff-AZ, v. 274, p. 222–230, 2012a.

CHEN, B.; LI, X.; XIAOA, X.; ZHAO, B.; DONG, J.; KOUD, W.; QIN, Y.; YANG, C.; WU, Z.; SUN, R.; LAN, G.; XIE, G. Mapping tropical forests and deciduous rubber plantations in Hainan Island, China by integrating PALSAR 25-m and multi-temporal Landsat images. **International Journal of Applied Earth Observation and Geoinformation**, Enschede-Overijssel, v. 50, p. 117–130, 2016.

CHEN, G.; HAY, G. J.; CARVALHO, L. M. T.; WULDER, M. A. Object-based change detection. **International Journal of Remote Sensing**, London-UK, v. 33, p. 4434–4457, 2012b.

CHEN, Q.; GONG, P. Automatic Variogram Parameter Extraction for Textural Classification of the Panchromatic Ikonos Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, Pasadena-CA, v. 42, n. 5, p. 1106–1115, 2004.

CHICA-OLMO, M.; ABARCA-HERNÁNDEZ, F. Computing geostatistical image texture for remotely sensed data classification. **Computers & Geoscience**, Houston-TX, v. 26, n. 4, p. 373–383, 2000.

CHILÉS, J. P., DELFINER, P. **Geostatistics: Modeling Spatial Uncertainty**. New York-NY: Wiley, 1999. 694 p.

COLOMBO, A. S. **Caracterização Socioeconômica e Tecnológica dos Heveicultores Vinculados à Aprobon no Noroeste Paulista**. 2009. 88p. Dissertação (Mestrado) – Faculdade de Engenharia, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Ilha Solteira-SP, 2009.

CONEGLIAN, A.; SEVERO, E. T. D. Eficácia de produtos químicos no controle do fungo manchador *Botryodiplodia theobromae* em madeira de *Hevea brasiliensis*. **Revista Energia na agricultura**, Botucatu-SP, v. 29, n. 2, p. 95–101, 2014.

CURRAN, P. The semivariograma in remote sensing: an introduction. **Remote Sensing of Environment**. Toronto-ON, v. 24, p. 493–507, 1988.

DEMARCHI, J. C.; PIROLI, E. L.; ZIMBACK, C. R. L. Análise temporal do uso do solo e comparação entre os índices de vegetação NDVI e SAVI no município de Santa Cruz do Rio Pardo-SP usando imagens Landsat 5. **RA'EGA**, Curitiba-PR, v. 21, p. 234–271, 2011.

DEVRIES, B.; PRATIHAST, A. K.; VERBESSELT J.; KOOISTRA L.; HEROLD, M. Characterizing forest change using community-based monitoring data and Landsat time series. **Plos One**, College Park-MD, v. 11, n. 3, 2016.

DONG, J.; XIAO, X.; SHELDON, S.; BIRADAR, C.; DUONG, N. D.; HAZARIKA, M. A comparison of forest cover maps in Mainland Southeast Asia from multiple sources: PALSAR, MERIS, MODIS and FRA. **Remote Sensing of Environment**, New York-NY, v. 127, p. 60–73, 2012a.

DONG, J.; XIAO, X.; SHELDON, S.; BIRADAR, C.; XIE, G. Mapping tropical forests and rubber plantations in complex landscapes by integrating PALSAR and MODIS

imagery. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 74, p.20–33, 2012b.

DONG, J.; XIAO, X.; CHEN, B.; TORBICK, N.; JIN, C.; ZHANG, G.; BIRADAR, C. Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. **Remote Sensing of Environment**, New York-NY, v. 134, p. 392–402, 2013.

DRAGUT, L.; TIEDEC, D.; Levick, S. R. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. **International Journal of Geographical Information Science**, London-UK, v. 24, n. 6, p. 859-871, 2010.

EASTMAN, J. R. **Idrisi for Windows**. Porto Alegre: UFRGS, Centro de Recursos Idrisi, Brasil, 1998.

ELVIDGE, C. D.; CHEN, Z. Comparison of broadband and narrow-band red and near-infrared vegetation indices. **Remote Sensing of Environment**, New York-USA. v. 54, p. 38-48, 1995.

EKADINATA, A.; WIDAYATI, A.; VINCENT, G. Rubber agroforest identification using object-based classification in Bungo District, Jambi, Indonesia. In: **ASIAN CONFERENCE ON REMOTE SENSING**, 25., 2004, Chiang Mai-Thailand. **Anais...** Chiang Mai-Thailand: ACRS, 2004. p. 551-556.

EMRAN, A.; HAKDAOUI, M.; CHOROWICZ, J. Anomalies on geologic maps from multispectral and textural classification: the Bleida mining district (Morocco). **Remote Sensing of Environment**, New York-USA, v. 57, p. 13–21, 1996.

EPIPHANIO, J. C. N.; GLERIANI, J. M.; FORMAGGIO, A. R.; RUDORFF, B. F. T. Índices de vegetação no sensoriamento remoto da cultura do feijão. **Pesquisa Agropecuária Brasileira**, Brasília-DF, v. 31, n. 6, p. 445-454, 1996.

ESPERANTE, D. **O que esperar dos preços da borracha natural para 2016?** São Paulo-SP: Seringueira, 2015. Disponível em: <http://seringueira.com.br/noticia_seringueira_140116.php> Acesso em: 18 fev. 2016.

FISHER, P. The pixel: a snare and a delusion. **International Journal of Remote Sensing**, London-UK, v. 18, p. 679–685, 1997.

FOX, J.; VOGLER, J.; SEN, O. L.; ZIEGLER, A. L.; GIAMBELLUCA, T. W. Simulating land-cover change in Montane Mainland Southeast Asia. **Environmental Managment**, Wako-TX, v. 49, p. 968-979, 2009.

FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: **Proceedings of the International Conference on Machine Learning**, 13. 1996, **Anais...** Desenzano Sul Garda-Bréscia: Universidade de Torino 'Departamento de informática, 1996. p. 148-156.

FURTADO, E. L. Comportamento de cultivares de seringueira (*Hevea spp.*) frente ao Mal das folhas na região do Vale do Ribeira – SP. 1996. 92 p. Tese (Doutorado em Agronomia)- Universidade de São Paulo – USP, Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ, Piracicaba-SP, 1996.

GAMA, J. **Árvores de Decisão**. Porto: Universidade do Porto, 2004.

GASPAROTTO, L.; SANTOS, A. F.; PEREIRA, J. C. R.; FERREIRA, F. A. **Doenças da seringueira no Brasil**. Brasília-DF: Serviço de Produção de Informação - EMBRAPA, 1997. 168p.

GHIMIRE, B.; ROGAN, J.; MILLER, J. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, Nottingham: UK, v. 1, n. 1, p. 45–54, 2010.

GHOSH, A.; JOSHI, P. K. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. **International Journal of Applied Earth Observation Geoinformation**, Enschede-Overijssel, v. 26, p. 298–311, 2014.

GONÇALVES, P. S. **Novos clones de seringueira para o estado de São Paulo**. São Paulo-SP: Secretaria de Agricultura e Abastecimento do Estado de São Paulo, 2013. Disponível em: <<http://www.defesa.agricultura.sp.gov.br>> Acesso em: 28 janeiro 2018.

GUARDIOLA-CLARAMONTE, M.; TROCH, P. A.; ZIEGLER, A. D.; GIAMBELLUCA, T. W.; DURCIK, M.; VOGLER, J. B.; NULLET, M. A. Hydrologic effects of the expansion of rubber (*Hevea brasiliensis*) in a tropical catchment. **Ecohydrology**, Perth-WA v. 3, p. 306-314, 2010.

HANSEN, M. C.; REED, B. A comparison of the IGBP DISCover and University of Maryland 1 km global land-cover products. **International Journal of Remote Sensing**, London-UK, v. 21 n. 6-7, p. 1365–1373, 2000.

HU, H.; LIU, W.; CAO, M. Impact of land use and land cover changes on ecosystem services in Menglun, Xishuangbanna, Southwest China. **Environmental Monitoring and Assessment**, Orono-ME, v. 46, p. 147-156, 2008.

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

IAC - INSTITUTO AGRONÔMICO DE CAMPINAS. **A importância da borracha nacional**. 2015. Disponível em: <<http://www.iac.sp.gov.br/areasdepesquisa/seringueira/importancia.php>>. Acesso em: 18 jun. 2017.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção Agrícola Mundial**, 2017. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=ms>> . Acesso em: 28 jun. 2018.

IFMG – Inventário Florestal de Minas Gerais. **Objetivos**. 2008. Disponível em: <<http://www.inventarioflorestal.mg.gov.br/>> . Acesso em: 24 mai. 2018.

INMET – INSTITUTO NACIONAL DE METEOROLOGIA. **Série histórica** – Dados diários, 2018. Disponível em: <<http://www.inmet.gov.br/portal/>>. Acesso em: 01 jul. 2018.

INPE – Instituto Nacional de Pesquisas Espaciais. **Coordenação de Observação da Terra: PRODES** – Monitoramento da Floresta Amazônica Brasileira por Satélite. 2018. Disponível em: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>>. Acesso em 24 mai. 2018.

INPE – Instituto Nacional de Pesquisas Espaciais. **Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal**. 2013. Disponível em: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/metodologia_t_axaprodes.pdf>. Acesso em 24 mai. 2018.

JENSEN, J. R. **Introductory Digital Image Processing**. Upper Saddle River-NJ: Prentice Hall, 2005. 544 p.

JENSEN, J. R. **Sensoriamento remoto do ambiente**: uma perspectiva em recursos terrestres. São José dos Campos: Parêntese, 2009. 604 p.

JUSOFF, K. Advanced Processing of UPM-APSB's AISA Airborne Hyperspectral Images for Individual Timber Species Identification and Mapping. **International Journal of Systems Applications, Engineering and Development**, Victoria-British Columbia, v. 2, p. 21-26, 2007.

JUSOFF, K. UPM-APSB AISA Airborne Hyperspectral Technology for Managing Mangrove Forest in Malaysia. **Modern Applied Science**, North York-ON, v.2, n. 6, p. 90-96, 2008.

JUSOFF, K.; YUSOFF, M. H. M. New approaches in estimating rubberwood standing volume using airborne hyperspectral sensing. **Modern Applied Science**, North York-ON, v. 3, n. 4, 62-70, 2009.

KRIGE, D. G. A statistical approach to some basic mine evaluation problems on the witwatersrand. **Journal of the South African Institute of Mining and Metallurgy**, Johannesburg-Gauteng, v. 52, p. 119-139, 1951.

LEAL, S. T. **A heveicultura na mesorregião leste do Estado de Mato Grosso do Sul**: Aspectos técnicos e econômicos. 2017. 108 p. Tese (Mestrado em Agronomia) - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, Ilha Solteira, 2017.

LIAW, A.; WIENER, M. Classification and regression by random forest. *R News*, Vienna-Vienna, v. 2, n. 3, p. 18–22, 2002.

LI, H.; AIDE, T. M.; MA, Y.; LIU, W. Demand for rubber is causing the loss of high diversity rain forest in SW China. *Biodiversity Conservation*, Kew-Surrey, v. 16, p. 1731-1745, 2007.

LIPPITT, C. D.; ROGAN, J.; LI, Z.; EASTMAN, JR.; JONES, T.G. Mapping selective logging in mixed deciduous forest: a comparison of machine learning algorithms. *Photogrammetric Engineering & Remote Sensing*, v. 74, p. 1201–1211, 2008.

LI, Z.; FOX, J. M. Integrating Mahalanobis typicalities with a neural network for rubber distribution mapping. *Remote Sensing Letters*, Nottingham: UK, v. 2, p. 157-166, 2011.

LI, Z.; FOX, J. M. Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data. *Applied Geography*, Kent-OH, v. 32, p. 420-432, 2012.

LONGÉPÉ, N.; RAKWATIN, P.; ISOGUCHI, O.; SHIMADA, M.; URYU, Y.; YULIANTO, K. Assessment of ALOS PALSAR 50 m Orthorectified FBD Data for Regional Land Cover Classification by Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, Pasadena-CA, v. 49, n. 6, p. 2135-2150, 2011.

LU, D.; WENG, Q. A. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, Morgantown-WV, v. 28, p. 823–870, 2007.

MALAYSIAN RUBBER STATISTICS. Natural Rubber Statistics 2016. Malaysian Rubber Board, 2016. Disponível em: <<http://www.lgm.gov.my/nrstat/nrstats.pdf>>. Acesso em: 28 jun. 2018.

MAS, J. F.; FLORES, J. J. The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, Morgantown-WV, v. 29, n. 3, p. 617–663, 2008.

MATHER, P.M. *Computer Processing of Remotely-Sensed Images: An Introduction*. Chichester-UK: John Wiley & Sons, 2004. 442 p.

MATHERON, G. Principles of **geostatistics**. *Economic Geology*, Lancaster-CA, v. 58. p. 1246- 1266, 1963.

MIETTINEN, J.; LIEW, S. C. Separability of insular Southeast Asian woody plantation species in the 50 m resolution ALOS PALSAR mosaic product. *Remote Sensing Letters*, Nottingham: UK, v. 2, n. 4, 299–307, 2011.

MIRANDA, F. P., FONSECA, L. E. N. Application of the semivariogram textural classifier (STC) for vegetation discrimination using JERS-1 SAR data of the Uaupés

river (north western Brazil). In: Simpósio Brasileiro de Sensoriamento Remoto, 1996, 8. Salvador-BA. **Anais...** INPE, 1996. p. 527–532.

MIRANDA, F. P., FONSECA, L. E. N., CARR, J. R. Semivariogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest. **IEEE International Journal of Remote Sensing**, London-UK, v. 19, n. 3, p. 549–556, 1998.

MIRANDA, F. P., MACDONALD, J. A., CARR, J. R. Application of the semivariogram textural classifier (STC) for vegetation discrimination using SIR-B data of Borneo. **IEEE International Journal of Remote Sensing**, London-UK, v. 13, p. 549–552/349–23546, 1992.

MITCHELL, T. M. **Machine Learning**. USA: McGraw-Hill, USA, 1997. 432 p.

MMA – MINISTÉRIO DO MEIO AMBIENTE. **O que é REDD+**. 2016. Disponível em: <<http://redd.mma.gov.br/pt/pub-apresentacoes/item/82-o-que-e-redd>> Acesso em: 24 maio. 2018.

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 2. ed. Viçosa: Universidade Federal de Viçosa - UFV, 2004. 307 p.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: a review. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 66, m. 3, p. 247–259, 2011.

MYINT, S. W.; WENTZ, E. A.; PURKIS, S. J. Employing spatial metrics in urban land-use/ land-cover mapping: comparing the Getis and Geary indices. **Photogrammetric Engineering & Remote Sensing**, v. 73, p. 1403–1415, 2007.

NOGUEIRA, R. F.; CORDEIRAO, S. A.; LEITE, A. M. P.; BINOTI, M. L. M. S. Mercado de borracha natural e viabilidade econômica do cultivo da seringueira no Brasil. **Nativa**, Sinop-MT, v. 3, n. 2, p. 143-149, 2015.

ONODA, M. 2001. 110 p. **Estudo sobre um algoritmo de árvore de decisão acoplado a um sistema de banco de dados relacional**. Dissertação (Mestrado) Universidade Federal de São João Del Rei – UFRJ, São João Del Rei-MG, 2001.

OSHIRO, T. M.; PEREZ, P. S.; BARANAUSKAS, J. A. How many trees in a random forest? In: P. PERNER, P. Machine Learning and Data Mining in Pattern Recognition. **Lecture Notes in Computer Science**. Heidelberg-Berlin, v. 7376, p. 154-168, 2012.

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. 101 p. Dissertação (Mestrado em Bioinformática) – Universidade de São Paulo – USP, Ribeirão Preto-SP, 2013.

PENNACCHIO, H. L. **Borracha natural**. Brasília-DF: Companhia Nacional de Abastecimento - CONAB, 2014. Disponível em:

<http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_07_16_13_52_53_borrachajunho2014.pdf> Acesso em: 20 jun. 2018.

PEZZONI FILHO, J. C.; BENSO, L.; FURTADO, E. L.; ZIMBACK, C. L. R. Índice de vegetação em clone de seringueira antes da desfolha natural. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 10., 2015. **Anais...** Uvaranas-PR: Associação Brasileira de Agroinformática – SBIAGRO, 2005a.

PEZZONI FILHO, J. C.; BENSO, L.; FURTADO, E. L.; ZIMBACK, C. L. R. Índice de vegetação em clone de seringueira após a desfolha natural. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 10., 2015. **Anais...** Uvaranas-PR: Associação Brasileira de Agroinformática – SBIAGRO, 2005b.

PONZONI, F. J. Comportamento Espectral da Vegetação. In: MENESES, P. R., NETTO, J. S. M. **Sensoriamento remoto, reflectância dos alvos naturais**. Brasília-DF: Editora Universidade de Brasília - UNB, Embrapa Cerrados, p 157-199, 2001.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M. **Sensoriamento remoto da vegetação**. São José dos Campos: Oficina de Textos, 2012. 135 p.

PU, R.; LANDRY, S. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. **Remote Sensing of Environment**. Toronto-ON, v. 124, p. 516–533, 2012.

QIU, J. Where the rubber meets the garden. **Nature**, v. 457, p. 246-247, 2009.

RAMSTEIN, G.; RAFFY, M. Analysis of the structure of radiometric remotely-sensed images. **International Journal of Remote Sensing**, Basingstoke-Hampshire, v. 10, p. 1049-1073, 1989.

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

ROGAN, J.; MILLER, J. Integrating GIS and remotely sensed data for mapping forest disturbance and change. In: WULDER, M.; FRANKLIN, S. E. **Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches**. Boca Raton-FL: Taylor & Francis, 2006. p. 133–172.

ROGAN, J.; FRANKLIN, J.; STOW, D.; MILLER, J.; WOODCOCK, C.; ROBERTS, D. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. **Remote Sensing of Environment**, New York-NY, v. 112, n. 5, p. 2272–2283, 2008.

ROSA, R. **Introdução ao Sensoriamento Remoto**, Uberlândia: EDUFU, 2003.

RUIZ, L. A.; FERNÁNDEZ-SARRÍA; A., RECIO, J. Evaluation of texture analysis techniques to characterize vegetation. In: SOBRINO, J. A. **Recent Advances in Quantitative Remote Sensing**. Valencia-Valencia: Universitat de Valencia, 2002. p. 514–521.

SANO, E. E.; BARCELLOS, A. O.; BEZERRA, H. S. **Área e distribuição espacial de pastagens cultivadas no Cerrado brasileiro**. Planaltina, DF: Embrapa Cerrados, 1999. 21 p. (Embrapa Cerrados. Boletim de Pesquisa, n. 3).

SANO, E. E.; ROSA, R.; BRITO, J. L. S.; FERREIRA, L. G. **Mapeamento de Cobertura Vegetal do Bioma Cerrado. Boletim de pesquisa e Desenvolvimento**. Planaltina, DF: Embrapa Cerrados, 2008. 60 p.

SANTORO, M.; FRANSSON, J. E. S.; ERIKSSON, L. E. B.; ULANDER, L. M. H. Clear-Cut Detection in Swedish Boreal Forest Using Multi-Temporal ALOS PALSAR Backscatter Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, Starkville-MS, v. 3, n. 4, p. 618-931, 2010.

SÃO PAULO (Estado). Secretaria de Agricultura e Abastecimento. Coordenadoria de Assistência Técnica Integral. Instituto de Economia Agrícola. **Levantamento censitário de unidades de produção agrícola do Estado de São Paulo - LUPA 2007/2008**. São Paulo: SAA/CATI/IEA, 2008. Disponível em: Acesso em: 15 jun. 2015.

SCALOPPI JUNIOR, E. J. **Borracha Natural se Expande no Brasil**. AGRIANUAL 2014 - ANUÁRIO DA AGRICULTURA BRASILEIRA. São Paulo-SP: Informa Economics South America/FNP, 2014. 399 p.

SCALOPPI JUNIOR, E. J. FREITAS, R. S, GONÇALVES, P. S. **O Agrônomo - Boletim Técnico-Informativo do Instituto Agrônomo**, Campinas-SP, v. 69 p. 56 – 2017.

SERTEL, E.; KAYA, S.; CURRAN, P. J. Use of Semivariograms to Identify Earthquake Damage in an Urban Area. **IEEE Transactions on Geoscience and Remote Sensing**, Pasadena-CA, v. 45, n. 6, p. 1590–1594, 2007.

SESNIE, S.; GESSLER, P.; FINEGAN, B.; THESSLER, S. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. **Remote Sensing of Environment**. Toronto-ON, v. 112, n. 5, p. 2145–2159, 2008.

SHRESTHA, R. P.; SCHMIDT-VOGT, D.; GNANAVELRAJAH, N. Relating plant diversity to biomass and soil erosion in cultivated landscape of the eastern seaboard region of Thailand. **Applied Geography**, Kent-OH, v. 30, p. 606-617, 2010.

SILVA, A. P. **Variabilidade espacial de atributos físicos do solo**. 1988. 105 p. Tese (Doutorado em Agronomia)- Universidade de São Paulo – USP, Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ, Piracicaba-SP, 1988.

SILVA, J. S. V.; SPERANZA, E. A.; VENDRUSCULO, L. G.; ESQUERDO, J. C. D. M.; MAURO, R. A.; BIANCHINI, S. L.; FLORENCE, R. O. **Projeto GeoMS: melhorando o Sistema de Licenciamento Ambiental do Estado do Mato Grosso do Sul**. Campinas/SP: Embrapa Informática Agropecuária, 2011. 64 p.

SILVA, L. M. O. **Uma aplicação de árvores de decisão, redes neurais e KNN para identificação de modelos ARMA não-sazonais e sazonais**. 2005. 138 p. Tese (Doutorado em Engenharia Elétrica) – Pontífca Universidade Católica – PUC, Rio de Janeiro- RJ, 2005.

SILVEIRA, E. M. O.; MENEZES, M. M.; ACERBI JÚNIOR, F. W.; TERRA, M. C. N. S.; MELLO, J. M. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Orlando-FL, v. 11, n. 3, 2017.

SMITH, A. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. **Journal of Spatial Science**, Colchester-Essex, v. 55, n. 1, p. 69–79, 2010.

SPRING. Sistema de Aquisição de Imagens: O que é uma imagem digital. In: SPRING. **Tutorial de Geoprocessamento**. São José dos Campos-SP: DPI/INPE, 2006. Disponível em: < http://www.dpi.inpe.br/spring/portugues/tutorial/introducao_sen.html >. Acesso em: 24 jul. 2018.

SOUZA, L. S. **Variabilidade espacial do solo em sistemas de manejo**. Porto Alegre, 1992. 162 p. Tese (Doutorado)-Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre-RS, 1992.

TRANGMAR, B. B.; YOST, R. S.; UEHARA, G. Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy*, San Diego-CA, v. 38, p. 45-94, 1985.

VAN DER MEER, F. Remote-sensing image analysis and geostatistics. **IEEE International Journal of Remote Sensing**, London-UK, v. 33, p. 5644–5676, 2012.

VIEIRA, S. R. *Geoestatística em estudos de variabilidade espacial*. Campinas-SP: IAC, 1998.

VIEIRA, S. R. Variabilidade espacial de argila, silte e atributos químicos em uma parcela experimental de um Latossolo roxo de Campinas (SP). **Bragantia**, Campinas-SP, v. 56, n.1, p. 1- 13, 1997.

VICENTE, L. E.; GOMES, D.; VICTORIA, D. C.; GARÇON, E. A. M.; BOLFE, E. L.; ANDRADE, R. G.; SILVA, G. B. S. Séries temporais de NDVI do sensor SPOT Vegetation e algoritmo SAM aplicados ao mapeamento de cana-de-açúcar. **Pesquisa Agropecuária Brasileira**, v. 47, n. 9 Brasília-DF, p. 1337-1345, 2012.

VOGELS, M. F. A.; DE JONG, S. M.; STERK, G.; ADDINK, E. A. Agricultural cropland mapping using black-and-white aerial photography, object-based image

analysis and random forests. **IEEE International Journal of Applied Earth Observation and Geoinformation**, Enschede-Overijssel, v. 54, p. 114–123, 2017.

WALKER, W. S.; STICKLER, C. M.; KELLNDORFER, J. M.; KIRSCH, K. M.; NEPSTAD, D. C. Large-Area Classification and Mapping of Forest and Land Cover in the Brazilian Amazon: A Comparative Analysis of ALOS/PALSAR and Landsat Data Sources. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, Starkville-MS, v. 3, n. 4, p. 594-604, 2010.

WANG, L.; SOUSA, W.P.; GONG, P.; BIGING, G. S. Comparison of IKONOS and Quick-Bird images for mapping mangrove species on the Caribbean coast of Panama. **Remote Sensing of Environment**, New York-NY, v. 91, n. 3–4, p. 432–440, 2004.

WESSELS, K. J.; VAN-DEN-BERGH, F.; DAVID P. ROY, D. P.; SALMON, B. P.; STEENKAMP, K. C.; MACALISTER, B.; SWANEPOEL, D.; JEWITT, D. Rapid land cover map updates using change detection and robust random forest classifiers. **Remote Sensing**, Basileia-SWI, v. 8, n. 11, p. 888, 2016.

WU, X.; PENG, J.; SHAN, J.; CUI, W. Evaluation of semivariogram features for object based image classification. **Geo-spatial Information Science**, Londres-UK, v. 18, n. 4, p. 159-170, 2015.

YUE, A. Z.; ZHANG, C.; YANG, J. Y.; SU, W.; YUN, W. J.; ZHU, D. Texture Extraction for Object-oriented Classification of High Spatial Resolution Remotely Sensed Images Using a Semivariogram. **IEEE International Journal of Remote Sensing**, London-UK, v. 34 n. 11, p. 3736–3759, 2013.

ZHOU, Z. H. **Ensemble Methods: Foundations and Algorithms**. Boca Raton-FL: CRC Press, 2012. 234 p.

CAPÍTULO 2. ATRIBUTOS ESPECTRAIS E ESPACIAIS NO MAPEAMENTO DE SERINGAIS – ESTUDO DE CASO DE UMA ÁREA NO MATO GROSSO DO SUL

RESUMO: A borracha natural vem sendo cada vez mais utilizada na composição dos mais diversos tipos de produtos, o que desencadeou considerável incremento nas áreas plantadas com Seringueira. Assim, o mapeamento e monitoramento de seringais ganhou notoriedade, sendo imagens de satélite utilizadas para esse fim. Mesmo que a estrutura dos dosséis traga muitos desafios à identificação dos seringais, a deciduidade característica da espécie é fundamental para aplicações de sensoriamento remoto à heveicultura. Uma das formas de trabalhar com os inconvenientes estruturais dos dosséis é utilizando atributos espaciais gerados a partir de valores de semivariogramas em conjunto com atributos espectrais no processo de classificação. Assim, este trabalho objetivou investigar a contribuição de um conjunto de atributos espaciais à classificação puramente espectral na tentativa de identificar quatro classes de estudo (Seringueira, Eucalipto, Nativa e Outros Usos) em uma área no Mato Grosso do Sul. Primeiro, imagens *Landsat* 8 foram adquiridas e segmentadas. Em seguida foi realizada a amostragem e a extração dos atributos espectrais e espaciais. Para a classificação utilizou-se o algoritmo *Random Forest*, e a validação foi realizada através de matrizes de confusão. Os resultados mostraram que a inclusão da componente temporal a classificações puramente espectrais diminuiu a confusão da classe Seringueira com a classe Outros Usos. As classificações puramente espaciais simuladas não forneceram acurácias aceitáveis, desqualificando a utilização de forma isolada dos atributos espaciais; e que utilizar atributos espectrais e espaciais em conjunto permite melhor separação das classes em estudo.

Palavras-chave: caracterização espectral, random forest, *Hevea brasiliensis*, Landsat 8, semivariograma, geoestatística.

CHAPTER 2. SPECTRAL AND SPATIAL ATTRIBUTES FOR MAPPING RUBBER TREES – A CASE STUDY OF AN AREA LOCATED IN MATO GROSSO DO SUL

ABSTRACT: Natural rubber is used in the production of thousands of different articles and hevea cultures continue to grow. Satellite imagery is often used to monitor these cultures. Although the canopy structure presents challenges for identification, the defoliation process typical for this species remains fundamental in remote sensing applications. To deal with the difficulties caused by the canopy structure, one method consists of employing spatial attributes generated from semivariogram values and spectral attributes data for classification. This essay aimed to explore the contribution of a set spatial attributes to the classic spectral classification in an attempt to identify four classes (Rubber Tree, Eucalyptus, Native and Other Uses) in an area in Mato Grosso do Sul. First, Landsat 8 images were obtained and segmented, then the objects were sampled and both spectral and spatial attributes were extracted for training the algorithm. The algorithm used for classification was the Random Forest and the validation was performed through confusion matrices. The results shown that the inclusion of temporal component to purely spectral classifications has reduced the confusion between Ruber Tree and Other Uses classes. Purely spatial classifications did not provide acceptable accuracies, disqualifying the use of spatial attributes isolated; and additionally joint spectral and spatial analysis achieves higher separability for the studied classes.

Key words: random forest, spectral characterization, *Hevea brasiliensis*, Landsat 8, smivariogram, geostatistics.

2.1. Introdução

Nos últimos anos observou-se aumento do número de produtos que utilizam borracha natural em suas composições, devido principalmente ao apelo ecológico e sustentável que envolve a produção da borracha natural quando comparada à borracha sintética. Dessa forma, o aumento da demanda da *commodity* resultou em considerável incremento nas áreas plantadas com Seringueira. De acordo com Pennacchio (2014) a produção mundial de borracha em 2014 cresceu 1,2%, enquanto o consumo teve aumento de 4%, e segundo Esperante (2015) as previsões apontam que esse padrão continue acontecendo, com a oferta mundial abaixo da demanda até 2018, voltando a se equilibrar entre 2019 e 2022. No estado do Mato Grosso do Sul, segundo dados divulgados por Leal (2017), a área ocupada com seringais cresceu 47% entre 2005 a 2015.

A Seringueira possui hábito decíduo o que a faz perder as folhas periodicamente. Essa característica é importante para aplicações de Sensoriamento Remoto (SR), mas ao mesmo tempo suscita alguns desafios. Segundo Dong *et al.* (2013), a deciduidade permite a separação espectral de dosséis da espécie de dosséis de florestas perenes (nativas ou exóticas) nas duas fases fenológicas distintas, a desfolha e a reenfolha. Muitos autores mapearam plantios de Seringueira com os objetivos mais diversos (DONG *et al.*, 2013; CHEN *et al.*, 2012; JUSOFF; YUSOFF, 2009; EKADINATA *et al.* 2004), e mesmo que a característica decídua dessa cultura colabore de certa forma ao processo de identificação dos plantios, ainda assim para Li e Fox (2011) a classificação e o mapeamento da espécie enfrentam muitos desafios.

Como forma de minimizar o efeito da variação espectral ao longo de um ano causada pela deciduidade das folhas, atributos espaciais extraídos de valores de semivariogramas podem ser utilizados na identificação de classes de uso do solo. O cálculo do semivariograma normalmente é feito utilizando valores de uma fonte de informação (como por exemplo, uma banda espectral ou um índice de vegetação) por vez e a utilização de semivariogramas tem sido registrada no campo do processamento de imagens digitais há alguns anos, principalmente pela sua capacidade de extrair informações a respeito da organização espacial dos valores dos *pixels* de um determinado objeto, sendo utilizados os mais diferentes tipos de imagens e aplicações

(CURRAN, 1988; BALAGUER *et al.*, 2010; WU *et al.*, 2015; SILVEIRA *et al.*, 2017), assim como os mais diversos tipos e números de atributos utilizados.

Segundo Curran (1988) o semivariograma é uma função geoestatística capaz de descrever a variabilidade espacial dos valores de uma variável, fornecendo uma descrição da escala e do padrão da variabilidade, sendo úteis na identificação de padrões. Para Silva (2011a) os semivariogramas são empregados em SR a partir da utilização dos valores de reflectância dos *pixels* como uma variável regionalizada, que possui características como aleatoriedade e correlação espacial.

Em análises de imagens de satélite é utilizado o semivariograma omnidirecional, e para variáveis contínuas como a reflectância, o semivariograma experimental é definido a partir da variância espacial interna dos objetos. Isso quer dizer que se calcula a variância entre os valores dos *pixels* de cada objeto, e para trabalhar com essa variância é necessário que seja fixada uma distância mínima entre dois pontos para que seja possível extrair o semivariograma. Essa distância é conhecida como distância *lag* e corresponde ao eixo x do semivariograma. Para análises de imagens de satélite a distância *lag* mínima é a distância entre dois *pixels*.

A distância *lag* não deve ser maior do que o tamanho do objeto e nem demasiadamente pequena. O semivariograma possui três parâmetros básicos: Alcance, Patamar e Efeito Pepita e para Woodcock *et al.* (1988 a e b), é necessário que o objeto seja maior que o alcance (distância em *pixels* onde a curva tende a se estabilizar; indica a distância máxima em que os dados estão correlacionados) do semivariograma, para que a variabilidade seja caracterizada corretamente, e grande o suficiente para identificar a periodicidade em classes que apresentem tal comportamento.

Segundo Curran (1988) semivariogramas periódicos são frequentemente encontrados em áreas com padrão repetitivo sobre superfícies espacialmente variáveis, como por exemplo, superfícies modificadas pelo homem como áreas urbanas e agrícolas, ou ainda algumas formações florestais que apresentem deciduidade, que é o caso da Seringueira. Em um semivariograma, essa periodicidade é caracterizada por curvas com comportamento cíclico ou aproximadamente cíclico.

Em modelagens convencionais é possível obter informações descritivas do semivariograma a partir dos três parâmetros básicos, mas nesse tipo de aplicação é usual o ajuste de modelos de semivariograma. Para contornar essa deficiência Balaguer *et al.*

(2010) propuseram um conjunto de 14 atributos extraídos de outros parâmetros do semivariograma como o primeiro máximo, primeiro mínimo e segundo máximo, por serem capazes de resumir informações mais relevantes do ponto de vista da descrição da variabilidade dos valores dos *pixels*.

A classificação da cobertura do solo pode envolver, em alguns casos, um grande número de atributos, o que não indica, necessariamente, uma maior capacidade descritiva das classes em questão. Em alguns casos apenas um pequeno conjunto de atributos contém as informações mais relevantes, enquanto o restante adiciona pouco ou nada à classificação. De acordo com Wu *et al.* (2015) alguns atributos não desempenham papéis importantes na classificação, podendo até mesmo ser dependentes entre si acarretando em redundância, ou ainda prejudicar a identificação dos objetos.

O processo que envolve a decisão de quais variáveis são mais ou menos importantes em um processo de classificação é uma questão delicada. Geralmente a tomada de decisão é feita com base na intuição do pesquisador respaldada pela literatura, ou seja, baseada em critérios que podem ser considerados mais subjetivos do que objetivos. Cada caso e área de estudo têm suas peculiaridades que devem ser estudadas individualmente para que possam ser extraídas informações realmente relevantes para solucionar um problema específico.

Com base no exposto, este trabalho tem como objetivo investigar, por meio de uma análise exploratória, a contribuição de um conjunto de atributos espectrais e espaciais na identificação de quatro classes em uma área com ocorrência de plantios de Seringueira no Mato Grosso do Sul.

2.2. Material e métodos

2.2.1. Área de Estudo

Uma imagem proveniente do sensor OLI, à bordo do satélite *Landsat 8*, cobre a área de estudo, com órbita/ponto: 223/73. A área de estudo engloba parte da mesorregião Leste do estado, e nela estão contidas duas microrregiões: Cassilândia e Paranaíba, compreendendo parte dos municípios de Cassilândia, Chapadão do Sul, Inocência e Paranaíba (Figura 7). Como a cena em questão cobre também o estado de Goiás, foi feito um recorte na cena utilizando o limite do estado de Mato Grosso do Sul.

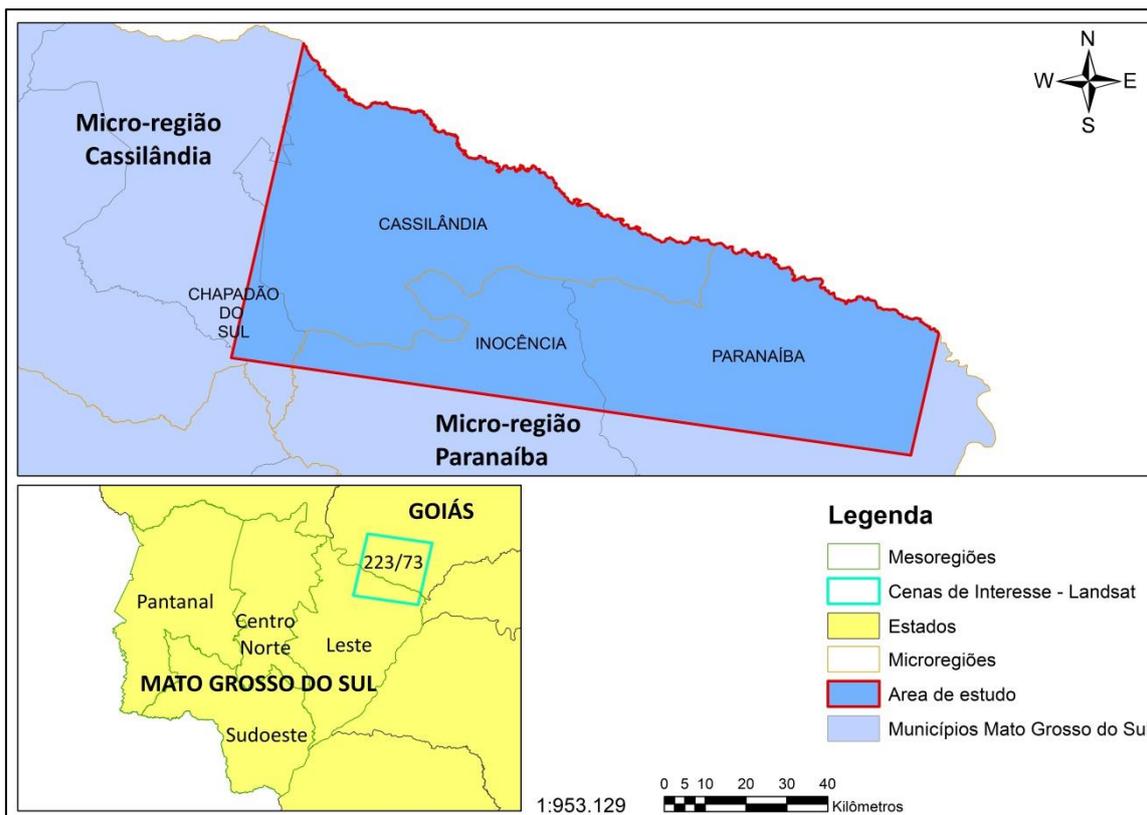


Figura 7. Cena utilizada para cobrir a área de estudo, localizada no Mato Grosso do Sul.

O clima da região é, segundo a classificação de Koppen, Aw, caracterizado como clima tropical quente e úmido. A temperatura média é de 26° C, com estação chuvosa no verão e seca no inverno. A precipitação pluvial total anual varia entre 900 e 1.400 mm; e a altitude média, entre 350 e 470 m.

A vegetação do Mato Grosso do Sul é composta por fitofisionomias do Cerrado, Pantanal e Mata Atlântica, sendo que 60,14% da área do estudo se encontram ocupada pelo Cerrado (SILVA *et al.*, 2011b). A área de estudo, não compreende o Pantanal, ocorrendo apenas Mata Atlântica e Cerrado.

2.2.2. Banco de dados (Imagens e dados)

As imagens foram adquiridas do órgão de Pesquisa Geológica dos Estados Unidos para Observação da Terra e Ciência (USGS/ EROS – do inglês *United States Geological Survey for Earth Observation and Science*) através do site *Earth Explorer* no nível 2 de processamento (*on-demand*), então além de correções geométricas e

radiométricas contam também com correção atmosférica, sendo então compostas por valores de reflectância de superfície.

A cena foi adquirida em duas datas, 07/10/2016 e 07/08/2017, possibilitando a inserção de uma componente temporal às análises. Essa componente foi responsável por agregar a característica caducifólia da Seringueira, trazendo a informação dos seringais completamente enfolhados na primeira data e completamente desfolhados na segunda data. Isso significa que para o mesmo objeto foram extraídos dois atributos espectrais representando a mesma fonte de informação, um para a primeira data e um para a segunda, e assim para todos os outros atributos.

Para a escolha das datas de estudo, este trabalho assumiu duas restrições. A primeira foi imposta pela cobertura de nuvens. A presença de nuvens não permite a identificação correta dos objetos e nem mesmo a correta segmentação, já que as sombras das nuvens e as próprias nuvens seriam consideradas objetos. A segunda foi imposta pela cultura, e se trata da desfolha natural da Seringueira, que geralmente acontece no início da estação seca, sendo a precipitação o fator de maior influência.

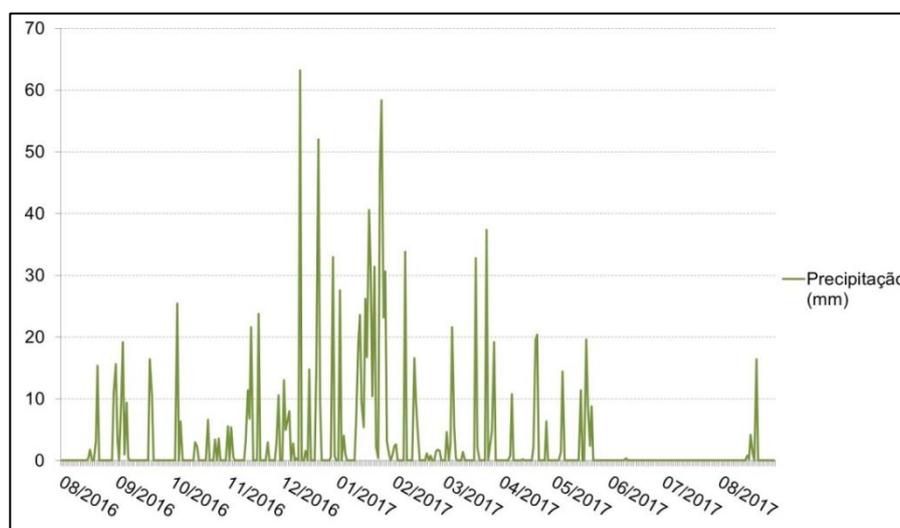


Figura 8. Dados de precipitação (mm) diária da estação meteorológica de Paranaíba – MS, fornecidos pelo INMET para o período de 01/08/2016 a 31/08/2017.

Os dados de precipitação disponibilizados pelo Instituto Nacional de Meteorologia – INMET (2018) para o município de Paranaíba foram utilizados na escolha das datas (Figura 8). O mês da primeira data (07/10/2016 - imagem com plantios enfolhados) foi precedido por chuvas bem distribuídas a partir do final do mês de agosto e, o mês da segunda data (07/08/2017 - imagem com plantios desfolhados) foi

precedido de um período de seca severa, que se iniciou no final do mês de maio. Portanto, as imagens selecionadas apresentaram cenas sem nuvens e plantios de Seringueira completamente enfolhados na primeira data e os mesmos plantios totalmente desfolhados na segunda.

2.2.3. Metodologia

O trabalho fundamentou-se na análise exploratória dos dados para avaliar a sensibilidade do classificador às Unidades Amostrais (UAs) coletadas. Essa análise exploratória foi feita através da geração e da avaliação de testes, apresentados por meio de comparações que deram origem às conclusões iniciais, utilizadas para nortear as classificações tratadas no Capítulo 3. Cada teste foi proposto para particularmente atender a uma necessidade específica de informação para ser usada no teste posterior, e assim sucessivamente.

A metodologia utilizada está resumida no fluxograma apresentado na Figura 9. A primeira etapa foi a segmentação. Em seguida, foram coletadas as UAs e realizada a extração das características espectrais e espaciais de cada um dos objetos gerados na segmentação. Os atributos que apresentaram valor nulo para a maioria ou para todos os objetos foram excluídos. Por último os atributos deram entrada no algoritmo *Random Forest* (RF) para a classificação.

Nos testes apresentados neste capítulo, selecionou-se aleatoriamente 66% das UAs para treinar o classificador, e os outros 34% para avaliar a acurácia. Isso significa que apenas as UAs foram utilizadas tanto para treinar quanto para avaliar o classificador, sendo a avaliação da acurácia da classificação feita utilizando as matrizes de confusão e as acurácias geral (AG), do produtor (AP) e do usuário (AU), geradas pelo próprio classificador.

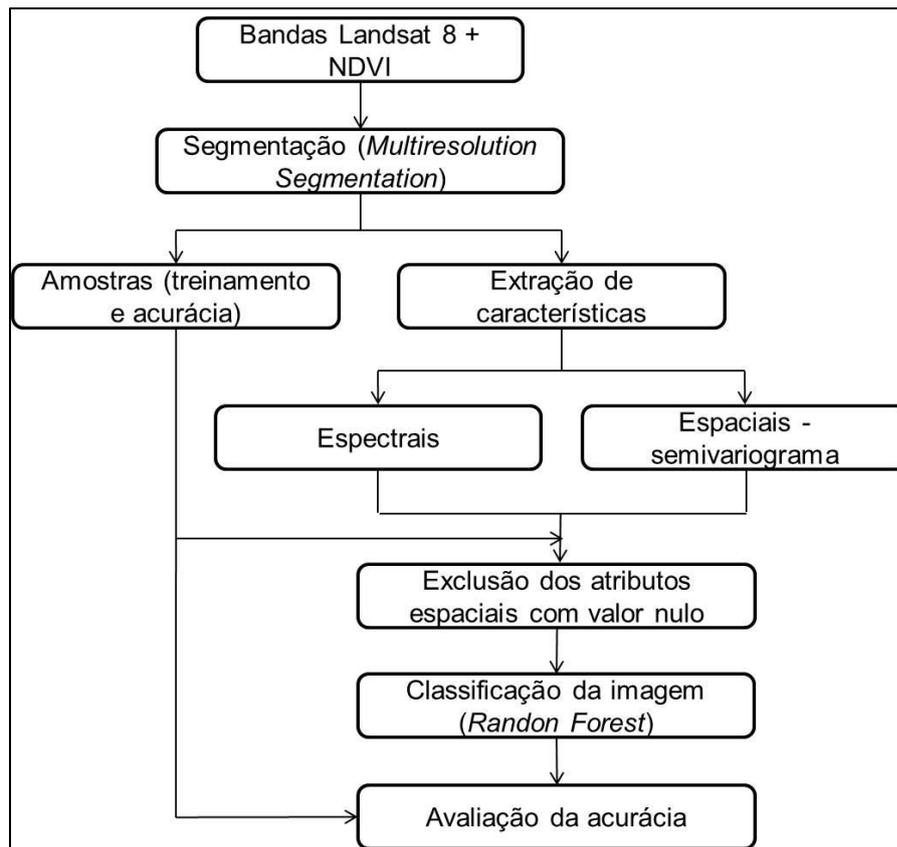


Figura 9. Fluxograma da metodologia utilizada no Capítulo 2.

2.2.3.1. Segmentação

A segmentação divide a imagem em objetos e é o primeiro passo para que seja possível a extração dos atributos por objeto. No presente estudo, as imagens foram segmentadas de acordo com suas características espectrais, utilizando o algoritmo de segmentação *Multiresolution Segmentation* do software *Ecognition*, conforme tratado em Silveira *et al.* (2017). Segundo os autores o algoritmo cria objetos homogêneos, analisando de maneira geral as resoluções nos diferentes tipos de dados.

Para a segmentação de imagens é necessário configurar três parâmetros: Escala, Compacidade e Forma, que para com Mui *et al.* (2015) controlam respectivamente o tamanho, a forma e a variação espectral dos objetos. Dentre os três, o mais crítico é o de Escala, por que segundo Wu *et al.* (2015) se relaciona com a resolução da imagem, influenciando diretamente no tamanho dos objetos que serão gerados. O tamanho mínimo dos objetos tem que levar em consideração o número mínimo de *pixels* dentro de cada objeto necessário para gerar o semivariograma, isso

por que segundo Silveira *et al.* (2017), a escala de segmentação influencia no tamanho dos objetos, que está conectado por sua vez a distância *lag* do semivariograma.

Nesse estudo, os parâmetros apropriados foram encontrados por tentativa e erro, conforme encontrado em Silveira *et al.* (2017), com atenção especial aos polígonos de Seringueira, por que para essa classe objetos grandes não são interessantes do ponto de vista de descrição dos limites devido a grande ocorrência de pequenos plantios na região. À medida que é priorizado a geração de objetos maiores, as classes de uso do solo que possuem menores objetos são conseqüentemente englobadas pelas classes de uso do solo que apresentam frequentemente polígonos maiores.

2.2.3.2. Amostragem

Foram consideradas quatro classes de cobertura do solo: Seringueira, Eucalipto, Nativa e Outros Usos. A classe Outros Usos é composta por todas as outras classes que não são nem Seringueira, nem Eucalipto e nem Nativa, fazendo com que essa classe tenha uma grande variação de tipos de objetos, como por exemplo, pasto, água, área urbana, culturas temporárias e etc.

O conjunto de UAs foi selecionado manualmente por interpretação visual a partir de conhecimento adquirido em campo a respeito da área. Devido à deciduidade apresentada pela Seringueira as UAs foram repassadas e analisadas uma a uma quanto a possibilidade de despertar dúvidas no classificador devido à cobertura foliar (enfolhada e desfolhada). Esse processo atuou como um filtro, onde UAs que não representavam a realidade fisiológica da classe no momento do imageamento (como um polígono de Seringueira, por exemplo, que por algum motivo estava desfolhado quando deveria estar enfolhado, ou vice e versa) foram descartadas.

A coleta de UAs também foi orientada a não causar superamostragem de uma classe com relação às outras. Quanto maior o número de UAs para avaliação da acurácia de uma classe menor o impacto de uma única UA nas acurácias do produtor e do usuário. Mas, a superamostragem de uma classe em relação às outras é prejudicial ao processo de amostragem e de classificação, e deve-se cuidar para que isso não ocorra. Esse cuidado envolve uma análise criteriosa a respeito da cobertura de uma determinada classe na área de estudo. Isso significa que a classe Outros Usos, por exemplo, é capaz de nos fornecer maior número de UAs sem que ocorra superamostragem da mesma, já

que é a classe de maior ocorrência na área de estudo, já as classes Seringueira e Eucalipto são capazes de fornecer um número bem menor de UAs sem que haja superamostragem, por que em termos de cobertura são classes bem menos numerosas na área de estudo.

Foram coletadas 109 UAs, 19 (17,4%) para a classe Seringueira, 14 (12,8%) para a classe Eucalipto, 41 (37,6%) para a classe Nativa e 35 (32,2%) para a classe Outros Usos. Dessa amostra, 66% foram destinadas ao grupo de amostras de treinamento e 34% ao grupo de amostras de acurácia.

2.2.3.3. Extração de atributos

Os atributos espaciais foram extraídos pelo software FETEX 2.0, como proposto por Ruiz *et al.* (2011), que é uma ferramenta para extração de informações e análise de imagens a partir de uma abordagem orientada a objeto, e foi projetado para trabalhar com uso e cobertura do solo, auxiliando na classificação dos objetos.

Nesse trabalho foram extraídos atributos espectrais a partir de valores das seguintes fontes de informação: das seis bandas espectrais captadas pelo sensor OLI à bordo do satélite Landsat 8 – Azul (B), Verde (G), Vermelho (R), Infravermelho próximo (NIR – do inglês *Near Infra-Red*), Infravermelho médio 1 (SWIR 1 – do inglês *Short Wave Infra-red Region*) e Infravermelho médio 2 (SWIR 2) - e do índice de vegetação NDVI (do inglês *The Normalized Difference Vegetation Index*, que é traduzido como Índice de Vegetação da Diferença Normalizada). Já os atributos espaciais foram extraídos de semivariogramas gerados a partir de valores das fontes de informação do Vermelho, do NIR e do NDVI.

2.2.3.3.1. Cálculo dos semivariogramas e extração de atributos espaciais por objeto

A distância *lag* é um dos fatores mais importantes a se atentar quando o semivariograma é gerado individualmente para cada objeto. De acordo com Miranda *et al.* (1998) a seleção da distância *lag* depende das classes de estudo devido a grande diversidade no padrão de variabilidade espacial que pode ser encontrada na imagem. Para a geração dos semivariogramas a distância *lag* utilizada foi de um *pixel*, como

forma de obter o máximo detalhamento da organização espacial interna dos objetos das classes de vegetação estudadas. Um *pixel* de uma imagem do sensor OLI/Landsat 8 tem 30 metros, e o número de *lags* foi fixado em 20 para garantir a geração dos semivariogramas, o que resultou em uma distância total de 600 metros.

Inicialmente foram utilizados como atributos espaciais todos os propostos por Balaguer *et al.* (2010) (Tabela 2). Esses atributos são divididos pelos autores de acordo com a posição dos *lags* utilizados em sua definição em: Próximos a origem, Até o primeiro máximo e Entre o primeiro e segundo máximos. Eles foram utilizados por agregarem mais informações relevantes e representativas, devido a capacidade de identificar pontos importantes e realçar a informação contida nos primeiros *lags*, onde a correlação espacial é maior. Mas, alguns não foram utilizados por não auxiliarem na separabilidade das classes, e esse assunto é tratado no tópico 2.1.1.1.

Tabela 2. Atributos espaciais descritos em Balaguer *et al.* (2010) divididas em: (A) Atributos extraídos próximos à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos

Grupo	Descrição	Atributos
A	Razão entre a variância total e a semivariância na primeira distância	$RVF = \frac{\text{variância}}{\gamma_i}$
A	Razão entre os valores de semivariância na primeira e na segunda distância	$RSF = \frac{\gamma_1}{\gamma_2}$
A	Derivada próxima da origem	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
A	Segunda derivada do semivariograma na terceira distância	$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$
B	Primeiro valor máximo de distância	$FML = h_{\max_1}$
B	Média dos valores de semivariância até o primeiro máximo	$MFM = \frac{1}{\max_1} \sum \gamma_i$
B	Variância dos valores de semivariância	$VFM = \frac{1}{\max_1} (\gamma_i - \gamma)^2$
B	Razão entre a semivariância no primeiro máximo e a média do semivariograma até esse máximo	$RMM = \frac{\gamma_{\max_1}}{\gamma_{\max_1}^{\text{média}}}$

CONTINUA...

...CONTINUAÇÃO

B	Diferença entre a média dos valores do semivariograma até esse máximo	$DMF = MFM - \gamma_i$
B	Diferença de segunda ordem entre o primeiro intervalo e o primeiro máximo	$SDF = \gamma_{\max_1} - 2\gamma \frac{\max_1}{2} + \gamma_1$
B	Área entre o valor do semivariograma na primeira distância e a função do semivariograma até o primeiro máximo	$AMF = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) + \gamma_{\max_1} \right) - (\gamma_1 (h_{\max_1} - h_1))$
C	Distância entre a localização do primeiro máximo local e o segundo máximo local	$DMS = h_{\max_2} - h_{\max_1}$
C	Distância entre o primeiro máximo e o primeiro mínimo	$DMM = h_{\min_1} - h_{\max_1}$
C	Área da abertura	$HA = \frac{h}{2} (\gamma_{\max_1} + 2(\sum_{i=\max_1+1}^{\max_2-1} \gamma_i) + \gamma_{\max_2}) - \frac{1}{2} (h_{\max_2} - h_{\max_1})(\gamma_{\max_2} - \gamma_{\max_1})$

Fonte: Adaptado Balaguer *et al.* (2010).

2.2.3.3.2. Extração de atributos espectrais

Tratando dos atributos espectrais, a variabilidade espectral das classes foi representada utilizando as fontes de informação (R, G, B, NIR, SWIR 1, SWIR 2 e NDVI) das duas datas (enfolhada e desfolhada), totalizando 14 fontes de informação disponíveis para a extração dos atributos espectrais. De cada fonte de informação foi extraída um atributo espectral por objeto, representado pela média dos valores de reflectância dos *pixels*.

À medida que os testes foram progredindo, as fontes de informação foram utilizadas ou não de acordo com a sua necessidade para a identificação espectral. Dessa forma foram testados no total 14 atributos espectrais, sendo sete para cada data de análise.

2.2.3.4. Exclusão de atributos espaciais

Os atributos que apresentaram valor 0 para a maioria ou para todos os objetos foram excluídos para que esses valores não influenciassem na interpretação e designação das classes do classificador. Valor 0 significa que, por algum motivo, o semivariograma não pôde ser gerado para o objeto e por isso o atributo apresentou esse valor. Manter esses atributos iria causar confusão da classificação e consequente queda nas acurácias. Os atributos espaciais gerados foram calculados para cada objeto e só depois os atributos foram excluídos.

2.2.3.5. Classificação

Com os objetos gerados e as características espaciais e espectrais extraídas, foi iniciado o processo de classificação. O classificador foi treinado utilizando como entrada os atributos selecionados para cada teste e a amostra de treinamento. Foi utilizado como classificador o RF, conforme em De Vries *et al.* (2016), que foi implementado no *software* de fonte aberta Weka 3.8.

Nesse *software*, segundo Belgiu e Dragu (2016), para a geração das árvores é necessário a configuração de dois parâmetros: o número de árvores de decisão que serão geradas (*Ntree*) e o número de variáveis à serem selecionadas e testadas para a divisão em cada nó, quando as árvores estão sendo geradas (*Mtry*). O *Mtry* foi deixado no valor padrão, que é o *log* do número de parâmetros + 1, conforme utilizado em Silveira *et al.* (2017), e o *Ntree* foi de 500 para cada teste por que segundo Lawrence (2006) a maioria dos estudos utiliza 500 árvores por que os erros tendem a estabilizar após esse número.

2.2.3.6. Avaliação da classificação

Como em conjunto com os atributos espaciais foram utilizados também atributos espectrais, os testes serão analisados de acordo com três categorias:

- (1) Teste puramente espectral: Atributos originados da média dos valores dos *pixels* extraídos das fontes de informação;
- (2) Teste puramente espacial: Atributos derivados de semivariogramas extraídos de valores das fontes de informação;

- (3) Teste espectral-espacial: Atributos das duas categorias de atributos utilizados em conjunto.

A avaliação da classificação foi feita usando a matriz de confusão e suas acurácias, conforme em Silveira *et al.* (2017).

- (1) Acurácia do Produtor: Indica a probabilidade de um polígono de referência, ou seja, de uma UA ser classificada corretamente, ou seja, porcentagem de polígonos que foram classificados na classe 'x' e que eram efetivamente pertencentes à classe 'x' no campo. Relaciona-se ao classificador, sendo uma medida da classificação. O inverso da AP é o Erro de Omissão (EO), ou seja, porcentagem de polígonos que foram classificados na classe 'x', mas que não pertenciam à classe 'x' no campo.
- (2) Acurácia do Usuário (AU): Indica a probabilidade de um polígono classificado na imagem realmente pertencer a essa classe na verdade de campo. É a porcentagem de polígonos que pertencem à classe 'x', de acordo com a verdade de campo, e que foram classificados como tal. Relaciona-se a amostragem, sendo uma medida da amostra fornecida pelo analista. O inverso da AU é o Erro de Comissão (EC) ou inclusão, que é a porcentagem de polígonos que tinha-se certeza que pertenciam à classe 'x', mas foram classificados como outra classe.
- (3) Acurácia Global (AG): Indica a precisão geral de uma classificação, que é calculada dividindo o número total de UAs classificadas corretamente pelo número total de UAs.

2.2.4. Análise exploratória dos dados

A seguir todos os testes serão detalhados, e após o último teste as informações encontram-se resumidas na Tabela 3.

TESTES PURAMENTE ESPECTRAIS

Foram gerados seis testes puramente espectrais para entender o impacto na classificação da inclusão de atributos espectrais extraídos de valores das fontes de informação estudadas (R, G, B, NIR, SWIR 1, SWIR 2 e NDVI) e da inclusão da

componente temporal (utilização da segunda data de análise) para extração de atributos. A descrição detalhada de cada teste se encontra a seguir:

Teste 1: Teste feito utilizando as fontes de informação R, G, B, NIR, SWIR 1 e SWIR 2 do sensor OLI/Landsat 8 e o NDVI, para extrair informações espectrais. Foi utilizada nesse teste apenas a primeira imagem (07/10/2016 – plantios de Seringueira enfolhados), portanto esse teste contou com a utilização de sete atributos puramente espectrais.

Teste 2: Como a Seringueira é uma espécie que apresenta deciduidade, para trabalhar com a diferença na estrutura do dossel foi necessário utilizar imagens em duas datas, conforme discutido no início da descrição da metodologia. O teste 2 foi gerado para entender os ganhos nas acurácias com a inclusão da componente espacial, ou seja, com a inclusão de atributos espectrais gerados a partir de valores extraídos de fontes de informação pertencentes à segunda data (07/08/2017 – plantios de Seringueira desfolhados).

A identificação das classes de estudo nesse teste foi feita utilizando as mesmas fontes de informações para extração dos atributos espectrais utilizados no teste anterior (R, G, B, NIR, SWIR 1, SWIR 2 e NDVI), mas nesse teste cada um desses atributos foram extraídos das imagens das duas datas concomitantemente. Portanto esse teste contou com a utilização de 14 atributos puramente espectrais, o dobro dos atributos espectrais utilizados no teste anterior.

A utilização de um grande número de atributos em classificações do uso do solo, ao contrário do que pensamos de imediato pode atrapalhar a classificação ao invés de ajudar. Essa questão suscita alguns questionamentos como, por exemplo: ‘quanto mais atributos, melhor a AG da classificação?’ ‘Uma interação negativa entre dois atributos pode fazer com que a AG da classificação diminua?’ ou ainda ‘quais os melhores atributos a serem utilizados para trabalhar com as classes de uso do solo selecionadas?’. Objetivando entender a interação das fontes de informação e amparar a seleção dos melhores atributos os testes 3, 4, 5 e 6 foram realizados.

Teste 3: Para entender a necessidade e o efeito na separabilidade das classes de atributos espectrais extraídos das fontes de informação do G, B, SWIR 1 e SWIR 2, no teste 3 essas fontes de informação foram excluídas, e por isso esse teste contou com a utilização de 6 atributos puramente espectrais.

O índice de vegetação NDVI é gerado a partir de valores das fontes de informação R e NIR. Para compreender qual a melhor forma de se trabalhar e/ou manipular os valores de reflectância dos *pixels* dessas três importantes fontes de informação na separabilidade das quatro classes de uso do solo (Seringueira, Eucalipto, Nativa e Outros Usos), essas fontes foram testadas duas a duas, conforme segue.

Teste 4: Esse teste foi feito utilizando atributos espectrais extraídos de valores das fontes de informação R e NIR, portanto esse teste contou com a utilização de quatro atributos espectrais.

Teste 5: Esse teste foi feito utilizando atributos espectrais extraídos de valores das fontes de informação R e NDVI, portanto esse teste contou com a utilização de quatro atributos espectrais.

Teste 6: Esse teste foi feito utilizando atributos espectrais extraídos de valores das fontes de informação NIR e NDVI, portanto esse teste contou com a utilização de quatro atributos espectrais.

TESTES PURAMENTE ESPACIAIS

Para ser possível a compreensão do efeito dos atributos espaciais nas acurácias e o poder de descrição das classes desses atributos, foi necessária a realização de alguns testes utilizando apenas atributos espaciais. Foram gerados quatro testes puramente espaciais, e os atributos espaciais utilizados foram gerados a partir de valores extraídos de três fontes de informação R, NIR e NDVI.

Os atributos espaciais foram extraídos por objeto, mas uma fonte de informação é capaz de fornecer até 14 atributos espaciais, se forem utilizados todos os atributos propostos por Balaguer *et al.* (2010). Conforme discutido na subseção 2.1.1.1 foi necessário a exclusão de alguns atributos, por isso para as fontes de informação R e NIR nesse Capítulo foram utilizados 11 atributos espaciais e para a fonte de informação NDVI foram utilizados apenas 3 atributos espaciais.

Teste 7: Esse teste foi feito utilizando atributos espaciais extraídos de valores das fontes de informação NDVI, R e NIR, portanto esse teste contou com a utilização de 25 atributos espaciais (11 do R, 11 do NIR e três do NDVI). Os atributos espaciais utilizados nesse teste estão especificados na Tabela 3.

Para entender como os atributos gerados de cada uma das três fontes de informação se relacionam com as classes de uso do solo e são capazes de identificá-las, foram gerados testes utilizando os atributos espaciais extraídos de cada uma das três fontes de informação separadamente, conforme segue.

Teste 8: Esse teste foi feito utilizando atributos espaciais extraídos de valores das fontes de informação R, portanto esse teste contou com a utilização de 11 atributos espaciais. Os atributos espaciais utilizados nesse teste estão especificados na Tabela 3.

Teste 9: Esse teste foi feito utilizando atributos espaciais extraídos de valores das fontes de informação NIR, portanto esse teste contou com a utilização de 11 atributos espaciais. Os atributos espaciais utilizados nesse teste estão especificados na Tabela 3.

Teste 10: Esse teste foi feito utilizando atributos espaciais extraídos de valores das fontes de informação NDVI, portanto esse teste contou com a utilização de três atributos espaciais. Os atributos espaciais utilizados nesse teste estão especificados na Tabela 3.

TESTES EPECTRAIS-ESPACIAIS

A partir dos resultados obtidos nos dois grupos de testes anteriores, foram gerados quatro testes espectrais-espaciais, onde atributos espectrais e espaciais foram utilizados em conjunto. Esses testes foram realizados para esclarecer como os atributos espaciais e espectrais se relacionam e influenciam a separabilidade das classes de estudo, e conseqüentemente as acurácias das classificações.

Nesse grupo de testes foram utilizadas as fontes de informação R, NIR e NDVI. Os atributos espectrais foram mantidos constantes nos quatro testes, sendo utilizadas as fontes de informação NIR e NDVI para extrair esses atributos. Já os atributos espaciais variaram por que foi identificado no grupo de testes puramente espaciais que classificações dessa natureza não originam AG aceitáveis. Dessa forma, na busca por maior e melhor compreensão do processo e conseqüente melhora na capacidade descritiva das classes, foi necessário alternar duas as duas a utilização das fontes de informação para extração de atributos espaciais. Nesse processo foram utilizadas as fontes de informação R, NIR e NDVI para extração de valores e geração de semivariogramas.

Da mesma forma que nos testes puramente espectrais, de cada fonte de informação foi extraído um atributo espectral por objeto. Os atributos espaciais também foram extraídos por objeto, conforme os testes puramente espaciais.

Teste 11: Esse teste é caracterizado por ser uma fusão dos testes 6 (puramente espectral) e 7 (puramente espacial). Dessa forma, a classificação foi feita utilizando atributos espectrais extraídos de valores das fontes de informação NIR e NDVI, em conjunto com atributos espaciais gerados a partir de semivariogramas extraídos de valores das fontes de informação R, NIR e NDVI. Portanto esse teste contou com a utilização de quatro atributos espectrais e 25 espaciais, totalizando 29 atributos. Os atributos espaciais são especificados na Tabela 3.

Teste 12: Teste realizado utilizando atributos espectrais extraídos de valores das fontes de informação NIR e NDVI, em conjunto com atributos espaciais gerados a partir de semivariogramas extraídos de valores das fontes de informação R e NIR. Portanto esse teste contou com a utilização de quatro atributos espectrais e 22 espaciais, totalizando 26 atributos. Os atributos espaciais são especificados na Tabela 3.

Teste 13: Teste realizado utilizando atributos espectrais extraídos de valores das fontes de informação NIR e NDVI, em conjunto com atributos espaciais gerados a partir de semivariogramas extraídos de valores das fontes de informação R e NDVI. Portanto esse teste contou com a utilização de quatro atributos espectrais e 14 espaciais, totalizando 18 atributos. Os atributos espaciais são especificados na Tabela 3.

Teste 14: Teste realizado utilizando atributos espectrais extraídos de valores das fontes de informação NIR e NDVI, em conjunto com atributos espaciais gerados a partir de semivariogramas extraídos de valores das fontes de informação NIR e NDVI. Portanto esse teste contou com a utilização de quatro atributos espectrais e 14 espaciais, totalizando 18 atributos. Os atributos espaciais são especificados na Tabela 3.

Tabela 3. Detalhamento dos atributos utilizados nos testes e resumo dos testes realizados, com seus respectivos objetivos principais e tipos de atributos utilizados. Linha tracejada divide os três grupos de testes (puramente espectrais, puramente espaciais e espectrais-espaciais).

Teste	Atributos					Nº de atributos	Objetivo principal
	Espectrais		Espaciais				
	Data 1	Data 2	R	NIR	NDVI		
1	R, G, B, NIR, SWIR 1, SWIR 2 e NDVI	-	-	-	-	7	Iniciar testes espectrais
2	R, G, B, NIR, SWIR 1, SWIR 2 e NDVI	R, G, B, NIR, SWIR 1, SWIR 2 e NDVI	-	-	-	14	Inclusão componente temporal
3	R, NIR e NDVI	R, NIR e NDVI	-	-	-	6	Diminuir atributos
4	R e NIR	R e NIR	-	-	-	4	Compreender fontes
5	R e NDVI	R e NDVI	-	-	-	4	Compreender fontes
6	NIR e NDVI	NIR e NDVI	-	-	-	4	Compreender fontes
7	-	-	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF e RMM	25	Iniciar testes espaciais

CONTINUA...

...CONTINUAÇÃO

8	-	-	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	-	-	11	Compreender fontes
9	-	-	-	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	-	11	Compreender fontes
10	-	-	-	-	RVF, RSF e RMM	3	Compreender fontes
11	NIR e NDVI	NIR e NDVI	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF e RMM	29	Unir melhores resultados

CONTINUA...

...CONTINUAÇÃO

12	NIR e NDVI	NIR e NDVI	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	-	26	Compreender interação fontes de informação - atributos espaciais
13	NIR e NDVI	NIR e NDVI	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	-	RVF, RSF e RMM	18	Compreender interação fontes de informação - atributos espaciais
14	NIR e NDVI	NIR e NDVI	-	RVF, RSF, FDO, SDT, FML, MFM, VFM, DMF, RMM, SDF e AFM	RVF, RSF e RMM	18	Compreender interação fontes de informação - atributos espaciais

Fonte: Este trabalho.

2.1. Resultados e discussão

2.1.1. Segmentação

Notou-se que a medida que o valor do parâmetro de escala aumentava, aumentava-se também o tamanho dos objetos, por que esse parâmetro segundo Batista (2006) permite maior número de iterações e agrupamentos, o que resulta em objetos maiores. Devido aos plantios de Seringueira e Eucalipto da área de estudo serem pequenos, no presente estudo trabalhou-se em um limiar desse parâmetro entre 100 e

200, para que fosse possível a identificação dos limites desses plantios e consequente formação dos polígonos coerentes com as classes de uso do solo abordadas.

Já para o parâmetro de forma, com o aumento do valor o tamanho dos objetos diminuiu e a confusão dos limites desses objetos aumentou, originando polígonos menos fiéis a realidade. Segundo Batista (2006) quanto menor o parâmetro de forma, mais suave o contorno, e por isso optou-se trabalhar com um limiar do parâmetro de forma entre 0,2 e 0,5.

O parâmetro de compacidade descreve a concentração de *pixels* sobre um objeto conforme o preenchimento desses *pixels* e pode ser entendido como a compactação do objeto.

Após uma série de testes, chegou-se no valor de 120 para o parâmetro de escala e de 0,2 para o de forma. O parâmetro compacidade foi mantido no *default*, que é de 0,5. Esses valores descreveram os limites dos objetos no contexto do objetivo do estudo. A Figura 10 ilustra a segmentação gerada a partir dos parâmetros citados.

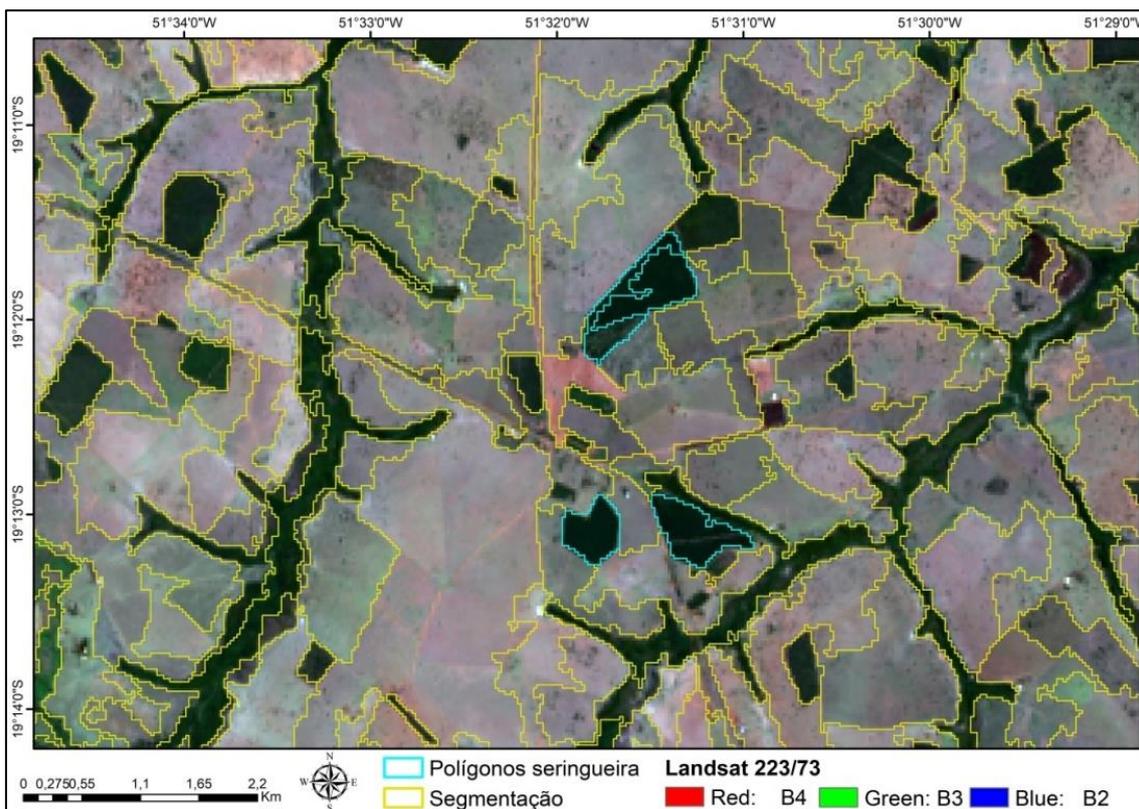


Figura 10. Segmentação gerada a partir da configuração – Escala: 120, Forma: 0,2 e Compacidade: 0,5 - do algoritmo *Multiresolution Segmentation* implementada no software *Ecognition*.

2.1.2. Semivariogramas para as classes em estudo

Foram gerados semivariogramas para cada classe de estudo para que fosse possível observar o comportamento dos semivariogramas de cada classe, que estão ilustrados nas Figuras de 11 a 13 para as fontes de informação R, NIR e NDVI, que foi de onde as informações espaciais foram extraídas no presente Capítulo.

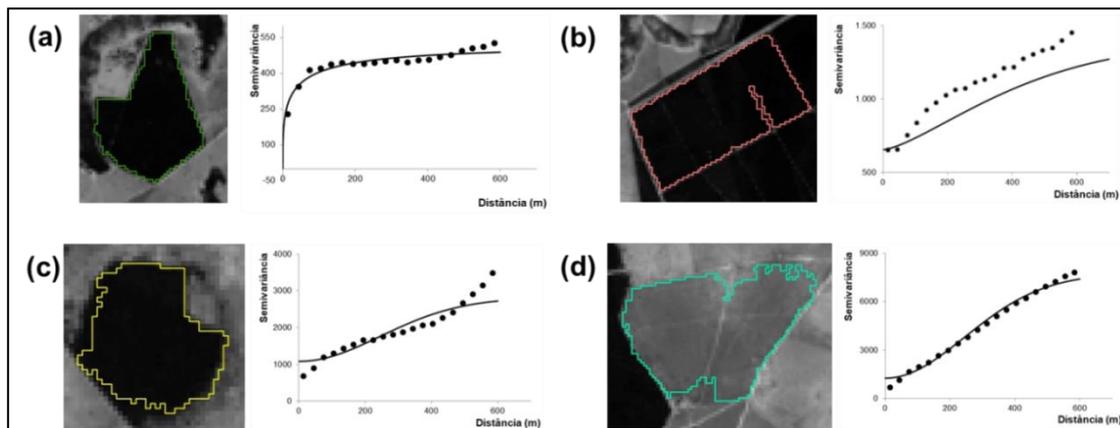


Figura 11. Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação Vermelho.

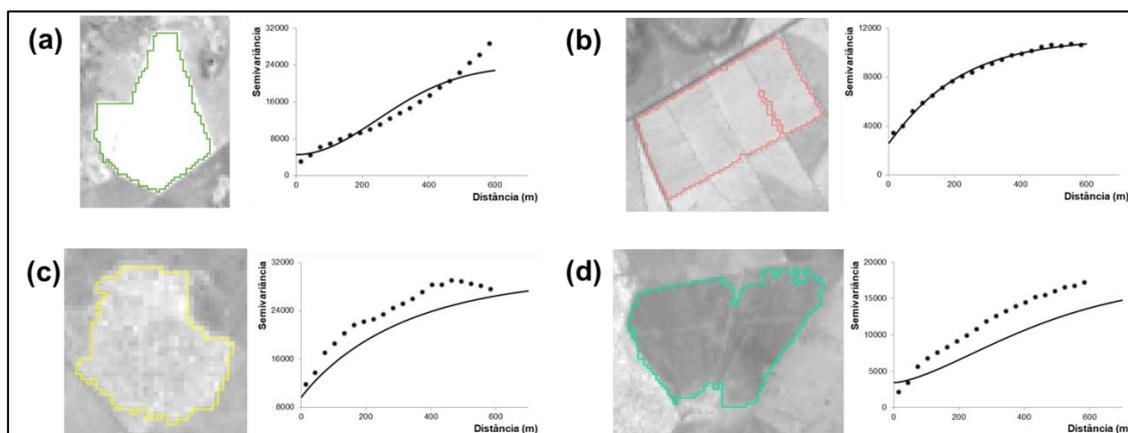


Figura 12. Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação NIR.

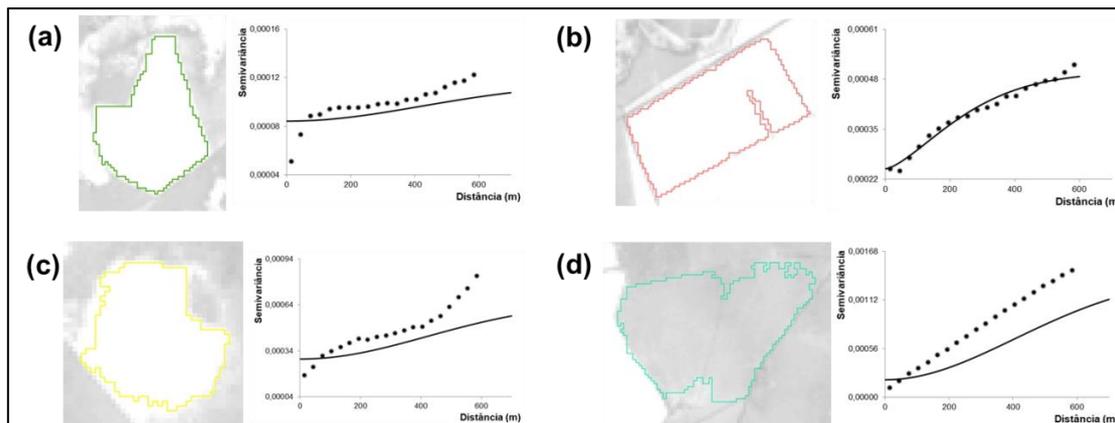


Figura 13. Semivariogramas das classes (a) Seringueira, (b) Eucalipto, (c) Nativa e (d) Outros Usos para a fonte de informação NDVI.

A classe Outros Usos é uma classe composta por vários tipos de objetos, pois abrange desde pastagens até corpos d'água e área urbana. O objeto selecionado para representar a classe Outros Usos nas Figuras de 11 a 13 é um polígono de pastagem, que é o que compõe a grande maioria dos polígonos representantes dessa classe.

Através da análise dos semivariogramas é possível observar que cada classe de estudo apresenta um comportamento bem específico para as três fontes de informação, evidenciando o potencial dos atributos extraídos do semivariogramas na discriminação de diferentes classes de cobertura do solo.

O semivariograma ilustrado na Figura 11, para fonte de informação R, para a classe Seringueira apresentou comportamento interessante, apresentando comportamento exponencial até atingir o patamar, se estabilizando a 100 metros.

Outras curvas que apresentaram comportamento exponencial foram as ilustradas para a fonte de informação NIR na Figura 12, para as classes Eucalipto e Nativa.

Todas as outras curvas ilustradas nas Figuras de 11 a 13 apresentaram comportamento Gaussiano.

2.1.3. Testes

Os testes forneceram informações a respeito das UAs, da área de estudo e das classes. Aqui se encontram apresentados os resultados dos testes discutidos na subseção 2.1.2 e os atributos selecionados para cada um deles. Um resumo dos atributos utilizados em cada teste pode ser encontrado na Tabela 3.

Quanto à utilização dos atributos espectrais, nos testes de 1 a 6 as fontes de informação (R, G, B, NIR, SWIR 1, SWIR 2 e NDVI) foram utilizadas para extração dos atributos espectrais de acordo com as necessidades, e à medida que os testes foram progredindo as fontes foram selecionadas. Como resultado da seleção encontrou-se no teste 6 melhores resultados utilizando apenas as fontes de informação NIR e NDVI. A partir deste teste, todos os outros (com exceção dos testes de 7 a 10, que são puramente espaciais) foram feitos utilizando os atributos espectrais extraídos dessas fontes de informação.

Com relação aos atributos espaciais, inicialmente no teste 7 (primeiro a utilizar atributos com natureza espacial) foram utilizados atributos espaciais gerados de valores extraídos de três fontes de informação, R, NIR e NDVI, após seleção dos atributos espaciais que não forneceram valores 0 para a maioria ou para todos os polígonos. Assim, no total foram utilizados nos testes 11 atributos espaciais extraídos da fonte de informação R, 11 extraídos da fonte de informação NIR e três extraídos da fonte de informação NDVI. Com relação a posição dos *lags* (Tabela 2) os atributos utilizados estão divididos conforme Tabela 4:

Tabela 4. Atributos utilizados nos testes com suas respectivas divisões: (A) Atributos extraídos próximas à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos (S=Sim, a fonte de informação foi utilizada e N=Não, a fonte de informação não foi utilizada).

Grupo	Atributo	Fonte de Informação		
		R	NIR	NDVI
A	RVF	S	S	S
	RSF	S	S	S
	FDO	S	S	N
	SDT	S	S	N
B	FML	S	S	N
	MFM	S	S	N
	VFM	S	S	N
	RMM	S	S	S
	DMF	S	S	N
	SDF	S	S	N
	AMF	S	S	N
C	DMS	N	N	N
	DMM	N	N	N
	HA	N	N	N

Os outros três testes puramente espaciais foram realizados para esclarecer dúvidas relacionadas à interação das fontes de informação R, NIR e NDVI (utilizadas para a extração dos atributos espaciais) com os atributos propostos por Balaguer *et al.* (2010). Entender essa interação foi necessário para esclarecer como as fontes de informação se comportaram ao serem utilizadas como base para a extração dos atributos espaciais, ou em outras palavras, para entendermos quão relevante à classificação as fontes se mostraram ao serem utilizadas para extração de valores para serem utilizados na geração de semivariogramas e posterior formação de atributos espaciais.

O teste 11 (primeiro teste espectral-espacial) foi uma junção dos testes 6 e 7, o que totalizou 29 atributos (4 espectrais + 25 espaciais), se tornando o teste com maior número de atributos. Nos testes de 12 a 14 as fontes de informação para extração de atributos espaciais foram alternadas, já os atributos espectrais se mantiveram constantes em todos os testes do terceiro grupo. A utilização alternada das fontes de informação para extração de atributos espaciais foi necessária para a compreensão de como as fontes de informação interagiram entre si ao serem utilizadas duas a duas para extrair atributos espaciais. Esses testes foram feitos a partir de resultados prévios do estudo, onde foi identificada diferente sensibilidade das classes de estudo a atributos espaciais gerados a partir das fontes de informação utilizadas para extração de atributos espaciais.

2.1.4. Comparações entre os testes

Em cada subseção referente às comparações os resultados das matrizes de confusão são apresentados e discutidos de forma separada para cada teste e após a apresentação e discussão os mesmos são comparados.

2.1.4.1. Comparação 1

A comparação apresentada nesta subseção foi realizada entre testes puramente espectrais (1 e 2) com o objetivo de estudar a necessidade de incorporação de uma componente temporal, e tem as informações apresentadas na Tabela 5. As matrizes de confusão dos testes aqui comparados se encontram na Tabela 6.

Tabela 5. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados nessa comparação 1.

COMPARAÇÃO 1		Teste	
		1	2
Nº de atributos utilizados		7	14
Polígonos classificados	Correto	25	35
	Incorreto	8	2
Seringueira	AP	91,7	100
	AU	91,7	75
Nativa	AP	60	94,1
	AU	50	100
Eucalipto	AP	60	83,3
	AU	85,7	83,3
Outros Usos	AP	83,3	100
	AU	62,5	100
Acurácia Global		75,8	94,6

Tabela 6. Matrizes de confusão dos testes 1 e 2 (acurácias do produtor nas colunas e acurácias do usuário nas linhas).

1	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	11	0	0	1
Nativa	0	3	3	0
Eucalipto	0	1	6	0
Outros Usos	1	1	1	5

2	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	3	0	1	0
Nativa	0	16	0	0
Eucalipto	0	1	5	0
Outros Usos	0	0	0	11

Teste 1: Fontes de informação provenientes do sensor OLI/Landsat 8 (RGB, SWIR e NIR) e o NDVI sem adição da componente temporal (apenas imagem da primeira data) – total de 7 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 1, para a classe Seringueira conclui-se que foi confundida apenas com a classe Outros Usos. A omissão (que representa os Erros de Omissão - EO) da Seringueira na classe Outros Usos foi de 8,3%. A inclusão (que representa os Erros de Comissão - EC) da classe Seringueira na classe Outros Usos também apresentou o valor de 8,3%.

A classe Nativa foi confundida com as classes Eucalipto e Outros Usos. O EO da classe Nativa foi de 20%, nas classes Eucalipto e Outros Usos, e o EC da classe Nativa aconteceu na classe Outros Usos, mas apresentou valor maior ainda, de 50%.

A classe Eucalipto foi confundida com as classes Nativa e Outros Usos. O EO da classe Eucalipto foi de 40%, nas classes Nativa (30%) e Outros Usos (10%). Já o EC da classe Eucalipto aconteceu apenas na classe Outros Usos, na ordem de 14,3%,

A classe Outros Usos se confundiu com as classes Seringueira, Nativa e Eucalipto. O EO aconteceu apenas na classe Seringueira, na ordem de 16,7%. O EC da classe Outros Usos foi de 37,5%, sendo no valor de 12,5% nas classes Seringueira, Nativa e Eucalipto.

Dessa forma, no teste 1 a AG apresentada foi de 75,8%. A classe Seringueira foi a melhor discriminada neste teste, com as maiores AP e AU, na ordem de 91,7%.

Teste 2: Fontes de informação provenientes do sensor OLI/Landsat 8 (RGB, SWIR e NIR) e o NDVI com adição da componente temporal (utilizando imagens das duas datas) – total de 14 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 2, a classe Seringueira, da mesma forma que no teste anterior, só foi confundida com a classe Outros Usos. O EO da Seringueira na classe Outros Usos diminuiu, passando de 8,3% para 0%. Já o EC da classe Seringueira na classe Outros Usos aumentou bastante, passando de 8,3% para 25%.

A classe Nativa foi confundida apenas com a classe Eucalipto, diferente do teste 1, onde ela foi confundida com as classes Eucalipto e Outros Usos. A confusão dessa classe diminuiu muito, apresentando as maiores acurácias dos dois testes para esta classe. O EO da Nativa apresentou uma queda significativa, passando de 40% para 5,9%, o que também aconteceu com o EC da classe Nativa, que apresentou queda em maior escala ainda, passando de 50% para 0%.

A classe Eucalipto continuou sendo confundida com a classe Nativa, mas nesse teste a confusão deixou de acontecer com a classe Outros Usos e passou a acontecer com a classe Seringueira. O EO da classe Eucalipto apareceu apenas na classe Seringueira e diminuiu bastante, passando de 40% para 16,7%. Já o EC apareceu apenas na classe Nativa, onde demonstrou pequena aumento, passando de 14,3% para 16,7%.

A classe Outros Usos apresentou nesse teste os melhores valores de AP e a AU para as quatro classes. Esta classe deixou de ser confundida com todas as classes, diferente do teste 1, onde ela se confundiu com todas elas. Dessa forma, a classe Outros Usos não apresentou EO e nem EC, resultado de acurácias máximas.

Dessa forma, no teste 2 a AG aumentou de 75,8% para 94,6%, mas para a classe Seringueira, que é o alvo do presente estudo, houve queda da AU. Mas, para todas as outras acurácias, exceto para a AU das classes Eucalipto e Seringueira que diminuíram, esse teste apresentou acurácias iguais ou superiores. Mesmo a classe objetivo do presente estudo sendo a Seringueira, não é apenas baseado na identificação desta classe que deve ser feita a escolha dos melhores resultados. À medida que a discriminação de todas as classes de estudo melhora, melhora também classificação da área como um todo, por que significa que o classificador foi capaz de interpretar melhor as entradas oferecidas para todas as classes.

No teste 2 a classe Seringueira deixou de se confundir com a classe Outros Usos, e esse detalhe deve ser levado em consideração. A confusão destas duas classes era de se esperar, já que espectralmente essas duas classes são parecidas na época de desfolha da Seringueira, e foi exatamente por esse motivo que a inclusão da componente temporal foi necessária. Esses resultados indicam que a componente temporal cumpriu seu objetivo, auxiliando na identificação da classe Seringueira e separabilidade desta com a classe Outros Usos, devido à deciduidade. Então, mesmo a classe Seringueira apresentando no teste 2 AP de 75%, proveniente de uma maior confusão com a classe Eucalipto, o teste 2 foi o que apresentou resultados globais mais satisfatórios.

2.1.4.2. Comparação 2

A comparação 2 foi realizada também entre testes puramente espectrais (2, 3, 4, 5 e 6) e foi feita para entender a capacidade descritiva de cada uma das fontes de informação, e a interação entre os atributos espectrais extraídos. O entendimento dessas questões respondeu importantes perguntas relacionadas, principalmente, a qualidade e quantidade dos atributos espectrais utilizados. As informações dos testes estão apresentadas na Tabela 7 e as matrizes de confusão na Tabela 8.

Tabela 7. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados nessa comparação 2.

COMPARAÇÃO 2		Teste				
		2	3	4	5	6
Nº de atributos utilizados		14	6	4	4	4
Polígonos classificados	Correto	35	34	32	29	35
	Incorreto	2	3	5	8	2
Seringueira	AP	100	100	100	100	80
	AU	75	75	75	50	100
Nativa	AP	94,1	93,8	87,5	77,8	94,1
	AU	100	93,8	87,5	87,5	100
Eucalipto	AP	83,3	71,4	57,1	50	100
	AU	83,3	83,3	66,7	50	66,7
Outros Usos	AP	100	100	100	90,9	100
	AU	100	100	100	90,9	100
Acurácia Global		94,6	91,9	86,5	78,4	94,6

Tabela 8. Matrizes de confusão dos testes 2, 3, 4, 5 e 6 (acurácias do produtor nas colunas e acurácias do usuário nas linhas).

2	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	3	0	1	0
Nativa	0	16	0	0
Eucalipto	0	1	5	0
Outros Usos	0	0	0	11

3	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	3	0	1	0
Nativa	0	15	1	0
Eucalipto	0	1	5	0
Outros Usos	0	0	0	11

4	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	3	0	1	0
Nativa	0	14	2	0
Eucalipto	0	2	4	0
Outros Usos	0	0	0	11

CONTINUA...

...CONTINUAÇÃO

5	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	2	0	1	1
Nativa	0	14	2	0
Eucalipto	0	3	3	0
Outros Usos	0	1	0	10

6	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	4	0	0	0
Nativa	0	16	0	0
Eucalipto	1	1	4	0
Outros Usos	0	0	0	11

Teste 3: Fontes de informação R, NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) – total de 6 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 3, para a classe Seringueira percebe-se que da mesma forma que no teste 2, esta classe só foi confundida com a classe Eucalipto. Mantendo o mesmo padrão apresentado no teste 2, a Seringueira não apresentou EO, e o EC se manteve estável em 25%, o que significa que para estes testes a Seringueira apresentou $\frac{1}{4}$ de inclusão na classe Eucalipto. Essa inclusão era de se esperar, visto que a hipótese inicial para a escolha das classes de estudo é que essas duas classes são espectralmente parecidas, e podem se confundir.

Para a classe Nativa percebe-se que, como no teste 2, essa classe continuou sendo confundida apenas com a classe Eucalipto. A diferença entre os testes 2 e 3 é que no teste 2 houve apenas EO, já no testes 3 houve EO e EC. O EO aumentou um pouco, passando de 5,9% para 6,2% de omissão da classe Nativa na classe Eucalipto. O EC da classe Nativa aumentou, passando de 0% para 6,2% na classe Eucalipto, mas apesar dessa queda na AU, houve equilíbrio maior entre as acurácias dessa classe.

Para a classe Eucalipto percebe-se que, como no teste 2, essa classe continuou sendo confundida com as classes Nativa e Seringueira, mas no teste 3 essa confusão aumentou. O EO demonstrou considerável aumento, passando de 16,7% para 28,6%. Isso aconteceu por que no teste 2 a classe Eucalipto não demonstrou omissão na classe Nativa, apenas inclusão, já no teste 3 apresentou tanto omissão quanto inclusão na

classe Nativa. A omissão foi da ordem de 14,3% tanto para a classe Seringueira quanto para a classe Nativa. Já o EC permaneceu estável em 16,7% apenas na classe Nativa.

Para a classe Outros Usos percebe-se que, como no teste 2, a classe Outros Usos não se confundiu com nenhuma classe, mantendo suas acurácias em 100%, que significa que esta classe não sofreu omissão e nem inclusão em nenhuma outra classe. Esse resultado pode indicar que essa classe não se confundiu com as outras três classes mesmo quando as bandas do B, do G e as do SWIR não foram utilizadas.

Dessa forma, em comparação ao teste 2, neste teste houve diminuição da AG, que passou de 94,6% para 91,9%. Mesmo que o teste 3 tenha apresentado uma AG ligeiramente menor, as acurácias dos testes 2 e 3 se mostraram equilibradas como um todo. Mas, a classe Eucalipto apresentou resultados mais delicados, e no teste 2 esta classe apresentou melhores acurácias. Isso pode sugerir que a classe Eucalipto é mais sensível a retirada dos atributos espectrais extraídos das fontes de informação G, B, SWIR 1 e SWIR 2, e por esse motivo tenha apresentado AP tão baixa no teste 3. Da mesma forma, podemos sugerir que todos os outros resultados foram causados pela exclusão dos atributos espectrais extraídos de valores dessas fontes de informação. Como já foi consagrado na literatura o poder das fontes de informação R, NIR e NDVI na identificação da vegetação, essas foram utilizadas para serem alternadas nos teste 4, 5 e 6.

Teste 4: Fontes de informação R e NIR com adição da componente temporal (utilizando imagens das duas datas) – total de 6 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 4, para a classe Seringueira percebe-se que da mesma forma que nos testes 2 e 3, esta classe só foi confundida com a classe Eucalipto e manteve os mesmo resultados, com APs e AUs idênticas às dos outros dois testes.

Da mesma forma que nos testes 2 e 3, a classe Nativa continuou sendo confundida apenas com a classe Eucalipto, mas no teste 4 apresentou acurácias inferiores. No teste 2 a classe Nativa apresentou apenas EO, já no teste 3 a classe apresentou EO e EC. No teste 4 o padrão ocorrido no teste 3 foi mantido, mas os erros aconteceram o dobro de vezes do teste 3. O EO da classe Nativa na classe Eucalipto aconteceu na ordem de 12,5%, assim como o EC.

Para a classe Eucalipto, como nos testes 2 e 3, ela continuou sendo confundida com as classes Nativa e Seringueira, mas a confusão, que já havia apresentado aumento do teste 2 para o teste 3, continuou aumentando no teste 4, que apresentou as menores acurácias. O EO subiu de 28,6% no teste 3 para 42,9% no teste 4. Na classe Seringueira a classe Eucalipto foi omitida 14,3%, enquanto que na classe Nativa a omissão foi de 28,6%. O EC da classe Eucalipto aconteceu apenas na classe Nativa na ordem de 33,3%, duas vezes mais do ocorrido nos testes 2 e 3, onde o EC foi de 16,7%.

A classe Outros Usos manteve o mesmo padrão identificado nos testes 2 e 3, onde não se confundiu com nenhuma outra classe, apresentando acurácias máximas.

Dessa forma, em comparação com os testes 2 e 3, no teste 4 houve queda da AG, que passou de 94,6% para 91,9% e depois para 86,5%. No teste 4 as acurácias se encontraram mais desbalanceadas do que no teste 3, mas as acurácias das classes Seringueira e Outros Usos permaneceram estáveis. Isso pode indicar que as fontes de informação R e NIR utilizadas em conjunto tem o mesmo poder de identificação das classes Seringueira e Outros Usos, que as fontes B, G, R, NIR, SWIR 1, SWIR 2 e NDVI utilizadas em conjunto (teste 2), ou ainda que as fontes R, NIR e NDVI utilizadas em conjunto (teste 3). Resumindo: esses resultados podem indicar que não é necessário utilizar as sete fontes de informação para identificar as classes Seringueira e Outros Usos, já que isso pode ser feito utilizando apenas duas, R e NIR, por que pelo que parece essas fontes descrevem melhor essas classes.

Mas por outro lado, para as classes Nativa e Eucalipto, a medida que fontes de informação foram retiradas das análises as acurácias apresentaram decréscimo e se mostraram mais desbalanceadas. No teste 2 foram utilizados sete fontes de informação atributos (14 atributos) e a AG foi a maior entre os testes 2, 3 e 4. No teste 3 foram utilizados três fontes de informação (seis atributos) e as acurácias começaram a diminuir e a se mostrar desbalanceadas. Já no teste 4 o número de fontes de informação diminuiu ainda mais passando para duas (quatro atributos), e o resultado foi acurácias menores e mais desbalanceadas para as classes Nativa e Eucalipto. Esses resultados podem indicar que para identificar as classes Nativa e Eucalipto, apenas as fontes de informação R e NIR não são suficientes, principalmente para a classe Eucalipto que apresentou os resultados mais críticos.

Teste 5: Fontes de informação R e NDVI com adição da componente temporal (utilizando imagens das duas datas) – total de 6 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 5, para a classe Seringueira percebe-se que diferente dos testes anteriores (2, 3 e 4) onde esta classe se confundiu apenas com a classe Eucalipto, no teste 5 esta classe se confundiu também com a classe Outros Usos. A AP se mostrou estável em 100%, não apresentando EO, mas a AU apresentou queda, passando de 100% para 80%. O EC aconteceu nas classes Eucalipto e Outros Usos, na ordem de 10% em ambos os casos. A classe Seringueira só havia sido confundida com a classe Outros Usos no teste 1.

Os resultados para a classe Nativa mostram que, diferente do ocorrido nos testes 2, 3 e 4 onde esta classe se confundiu apenas com a classe Eucalipto, no teste 5 a classe Nativa também se confundiu com a classe Outros Usos, apresentando os piores resultados dentre os testes 2, 3, 4 e 5. A AP caiu de 87,5% para 77,8%, já a AU se manteve estável em 87,5%. O EO aconteceu na classe Eucalipto na ordem de 16,6% e na classe Outros Usos de 5,6%. O EC aconteceu apenas na classe Eucalipto na ordem de 12,5%. A classe Nativa não havia sido confundida com a classe Outros Usos em nenhum outro teste.

Para a classe Eucalipto, como nos testes 3 e 4, ela continuou sendo confundida com as classes Nativa e Seringueira, mas a confusão, que já havia apresentado aumento do teste 3 para o teste 4, continuou aumentando no teste 5 que apresentou as menores acurácias até então. O EO subiu de 42,9% no teste 4 para 50% no teste 5. Na classe Seringueira a classe Eucalipto foi omitida 16,7%, enquanto que na classe Nativa a omissão foi de 33,3%. O EC da classe Eucalipto aconteceu apenas na classe Nativa passando de 33,3% no teste 4 para 50% no teste 5.

A classe Outros Usos, diferente do evidenciado nos testes 2, 3 e 4 onde esta classe não se confundiu com nenhuma outra classe, no teste 5 passou a se confundir com as classes Nativa e Seringueira. O EO aconteceu na classe Seringueira na ordem de 9,1%, e o EC aconteceu na classe Nativa também na ordem de 9,1%. A classe Outros Usos só havia sido confundida com a classe Outros Usos no teste 1.

Dessa forma, em comparação com os testes 2, 3 e 4 a AG foi a menor até então, apresentando o valor de 78,4%. Nenhuma acurácia demonstrou aumento e apenas a AP da classe Seringueira e AU da classe Nativa se mantiveram iguais.

No teste 5 a falta de balanço entre as acurácias ficou ainda mais evidente e em maior escala. As classes Seringueira e Outros Usos, que haviam demonstrado boa resposta a diminuição do número de atributos e das fontes de informação e apresentaram bons resultados utilizando as fontes de informação R e NIR apenas (teste 4), neste teste apresentaram queda das acurácias. Para as classes Nativa e Eucalipto, que se mostraram sensíveis e apresentaram resposta negativa a diminuição do número de atributos e fontes de informação (teste 4), as acurácias continuaram demonstrando queda e falta de balanço.

Ao utilizar a fonte de informação do NDVI (teste 5) no lugar do NIR (teste 4), em conjunto com o R para extração de atributos espectrais, os resultados foram piores. Isso pode indicar que, mesmo que os resultados para as classes Nativa e Eucalipto não tenham sido bons no teste 4, em conjunto com o R o NIR é melhor para identificar as quatro classes do que o NDVI.

Uma observação importante a ser levada em consideração é que no teste 5 as classes de vegetação de porte arbóreo (Seringueira, Nativa e Eucalipto) passaram a se confundir com a classe de vegetação rasteira (Outros Usos), assim como a classe de vegetação rasteira passou a se confundir com duas das três classes de vegetação de porte arbóreo (Seringueira e Nativa). Isso pode indicar confusão estrutural devido a incapacidade das fontes de informação R e NDVI de descrever as classes abordadas.

Até então a melhor AG foi encontrada no teste 2, ou seja, as melhores fontes de informação para identificar as quatro classes de uso do solo são G, B, R, NIR, SWIR 1, SWIR 2 e NDVI, quando comparados os testes 2, 3, 4 e 5.

Teste 6: Fontes de informação NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) – total de 6 atributos espectrais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 6, para a classe Seringueira percebe-se que o padrão apresentado nos testes 2, 3 e 4 voltou a ser exibido, se confundindo apenas com a classe Eucalipto. Mas, diferente dos testes 2, 3, 4, 5 e 6 deixou de apresentar EC e passou a apresentar EO. O EO da classe Seringueira na Classe Eucalipto passou de 0% no teste 5 para 20% no teste 6. Em contrapartida, o EC passou de 50 % no teste 5 para 0% no teste 6.

A classe Nativa se confundiu apenas com a classe Eucalipto e os resultados mostram que, diferente do ocorrido nos testes 3, 4 e 5 onde a classe apresentou tanto EO

quanto EC, no teste 6 ela apresentou apenas EO, assim como no teste 2. Os resultados para a classe Nativa foram idênticos aos do teste 2, sendo os melhores resultados dentre todos os testes para esta classe. O EO aconteceu na classe Eucalipto na ordem de 5,9%.

Para a classe Eucalipto, como nos testes 2, 3, 4 e 5, ela continuou sendo confundida com as classes Nativa e Seringueira, mas a confusão diminuiu e a classe passou a exibir apenas EC, fato esse que não havia acontecido em nenhum dos testes anteriores, fazendo com que o teste 6 apresentasse os melhores resultados para a classe Eucalipto. O EC aconteceu na ordem de 33,3%, apresentando inclusão de 16,7% tanto para a classe Seringueira quanto para a classe Nativa.

A classe Outros Usos voltou a exibir o padrão identificado nos testes 2, 3 e 4, onde não se confundiu com nenhuma das outras três classes.

Dessa forma, em comparação com os testes 2, 3, 4 e 5 a AG apresentou o maior valor, que foi o mesmo exibido pelo teste 2, na ordem de 94,6%.

No teste 6 a falta de balanço entre as acurácias foi recuperada e as classes Seringueira e Outros Usos, que haviam demonstrado boa resposta a diminuição do número de atributos e das fontes de informação no teste 4 (apenas as fontes R e NIR), mas queda nas acurácias no teste 5 (apenas as fontes de informação R e NDVI), voltaram a apresentar boas acurácias. As classes Nativa e Eucalipto se mostraram sensíveis e apresentaram resposta negativa a diminuição do número de atributos e das fontes de informação nos testes 4 (apenas as fontes R e NIR) e 5 (apenas as fontes R e NDVI), passaram a apresentar boa acurácias, com exceção da AU da classe Eucalipto.

Ao utilizar as fontes de informação NIR e NDVI (teste 6) os resultados foram parecidos com os do teste 2. Dessa forma, ao compararmos esses testes, que foram os melhores até então, percebemos que mesmo que a classe Eucalipto tenha apresentado AU não satisfatória, o teste 6 apresentou melhores acurácias para a classe Seringueira.

Os resultados do teste 5 deram a entender que as fontes de informação R e NIR utilizadas em conjunto eram melhores para identificar as quatro classes do que as fontes de informação R e NDVI (teste 5), mas que a melhor alternativa para identificar as quatro classes de uso do solo ainda era a apresentada no teste 2 (utilizar todas as fontes de informação propostas - G, B, R, NIR, SWIR 1, SWIR 2 e NDVI). O teste 6 evidenciou uma outra face dos resultados, onde a utilização isolada das fontes de informação NIR e NDVI apresentou os mesmos resultados da utilização de todas as fontes de informação propostas para extração de atributos espectrais (teste 2).

Assim podemos começar a tentar responder as três perguntas feitas anteriormente na subseção 2.1.1. ‘Quanto mais atributos, melhor a AG da classificação?’, ‘Uma interação negativa entre dois atributos pode fazer com que a AG da classificação diminua?’ e ‘Quais os melhores atributos a serem utilizados para trabalhar com as classes de uso do solo selecionadas?’.

Por meio dos resultados obtidos até aqui percebemos que não é a quantidade de atributos e nem de fontes de informação que faz uma classificação ser melhor ou pior, e sim a “qualidade” desses atributos e fontes de informação, onde qualidade se refere a capacidade das fontes de informação e seus atributos de identificar as classes. O teste 2 contou com 14 atributos (7 fontes de informação) e forneceu resultados similares ao teste 6, que contou com 4 atributos (2 fontes de informação).

Mas pode-se perceber também que há interação negativa entre atributos gerados de fontes de informação diferentes. Os testes 4 e 5 também utilizaram apenas 4 atributos, e nem por isso apresentaram bons resultados. Percebemos que em ambos os casos foi utilizada a fonte de informação R, mas ao uni-la a fonte de informação NDVI (teste 4) o resultado é inferior do que ao uni-la a fonte de informação NIR (teste 5). Dessa forma pode-se pensar que, então, o problema se encontra na fonte de informação NDVI, pois foi a única modificação feita entre os testes 4 e 5. Mas, ao se utilizar a fonte de informação NDVI com a fonte de informação NIR (teste 6), obtivemos os melhores resultados. Isso pode indicar que quando a fonte NDVI é utilizada com a fonte R há uma interação negativa entre os atributos, que não acontece quando a fonte NDVI é utilizada com a fonte NIR.

Esses resultados nos levam a acreditar que os melhores atributos para serem utilizados para trabalhar com as classes de uso do solo selecionadas (Seringueira, Nativa, Eucalipto e Outros Usos) são os atributos espectrais extraídos das fontes de informação NDVI e NIR, por que apresentaram resultados similares de quando os atributos foram extraídos das sete fontes de informação (teste 2), mas com a vantagem de utilizar um menor número de atributos, que permite maior compreensão do processo e da capacidade de identificação dos atributos utilizados. Por esse motivo, no grupo de testes espectrais-espaciais (subseção 2.1.4.4) foram utilizadas apenas essas duas fontes de informação para extração de atributos espectrais.

2.1.4.3. Comparação 3

A comparação apresentada nesta subseção foi realizada entre testes puramente espaciais (7, 8, 9 e 10) com o objetivo de estudar o poder de identificação das classes pelos atributos espaciais propostos, e as fontes de informações utilizadas para extraí-los. As informações dos testes estão apresentadas na Tabela 9 e as matrizes de confusão na Tabela 10.

Tabela 9. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados na Comparação 3.

COMPARAÇÃO 3		Teste			
		7	8	9	10
Nº de atributos utilizados		25	11	11	3
Polígonos classificados	Correto	23	12	19	14
	Incorreto	14	25	18	23
Seringueira	AP	33	16,7	16,7	25
	AU	75	25	50	50
Nativa	AP	71,4	42,1	71,4	40,9
	AU	62,5	50	62,5	56,3
Eucalipto	AP	0	0	0	50
	AU	0	0	0	16,7
Outros Usos	AP	71,4	27,3	70	40
	AU	90,9	27,3	63,6	18,2
Acurácia Global		62,2	32,4	51,4	37,8

Tabela 10. Matrizes de confusão dos testes 7, 8, 9 e 10 (acurácias do produtor nas colunas e acurácias do usuário nas linhas).

7	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	3	0	0	1
Nativa	4	10	0	2
Eucalipto	2	3	0	1
Outros Usos	0	1	0	10

8	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	1	0	1	2
Nativa	3	8	0	5
Eucalipto	0	5	0	1
Outros Usos	2	6	0	3

CONTINUA...

...CONTINUAÇÃO

9	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	2	0	0	2
Nativa	5	10	1	0
Eucalipto	3	2	0	1
Outros Usos	2	2	0	7

10	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	2	1	0	1
Nativa	4	9	1	2
Eucalipto	0	5	1	0
Outros Usos	2	7	0	2

Teste 7: Fontes de informação R, NIR e NDVI – total de 25 atributos espaciais (11 do R, 11 do NIR e três do NDVI).

Examinando os valores fora da diagonal principal da matriz de confusão do teste 7, para Seringueira percebe-se que ela foi confundida com as outras três classes. Esse resultado deve ser analisado com bastante cautela, já que a confusão com todas as classes não havia acontecido em nenhum outro teste. O EO da classe Seringueira aconteceu na ordem de 67%, nas classes Nativa (44,7%) e Eucalipto (22,3%). A omissão nas classes Nativa e Eucalipto faz todo sentido em análises espaciais, já que as três classes são compostas por árvores, que por sua vez possuem maior similaridade nas características espaciais do que com a classe Outros Usos. O EC desta classe foi de 25%, mas a inclusão da classe Seringueira aconteceu na classe Outros Usos. Isso provavelmente ocorreu por que a Seringueira na fase de desfolha tem maior semelhança espectral com a classe Outros Usos, por predominar a resposta espectral do solo, do que com as classes Nativa e Eucalipto, já que essas classes não apresentam deciduidade. Assim, esses resultados apontam que, provavelmente, apenas os atributos espaciais não são suficientes para diferenciar a classe Seringueira das demais classes arbóreas, assim como a classe Seringueira da classe Outros Usos na fase de desfolha.

A classe Nativa, assim como a classe Seringueira, também se confundiu com todas as classes, e essa confusão só havia acontecido no teste 1. Embora os resultados para essa classe não sejam bons, as acurácias foram maiores do que as da classe Seringueira. O EO aconteceu na ordem de 28,6%, nas classes Eucalipto (21,5%) e Outros Usos (7,1%). O EC foi maior, na ordem de 37,5% com as classes Seringueira

(25%) e Outros Usos (12,5%). Percebe-se tanto para o EO quanto para o EC, os maiores erros aconteceram nas classes de vegetação arbórea, quando comparada a classe de vegetação rasteira. Esse resultado respalda a especulação formulada anteriormente para a classe Seringueira de que os atributos espaciais sozinhos não são capazes de separar classes arbóreas.

A classe Eucalipto foi completamente excluída pelo classificador. Então para esse teste os piores resultados foram para esta classe, onde o classificador não conseguiu classificar nenhum polígono como tal. Os resultados da classe Eucalipto colaboram com a conclusão de que uma classificação baseada apenas nos atributos espaciais extraídos de valores das fontes de informação R, NIR e NDVI não é viável, já que a classe Eucalipto foi completamente excluída. Além disso, todos os valores de acurácia se mostraram ou iguais ou muito inferiores aos resultados da classificação gerada apenas com atributos espectrais extraídos de valores das mesmas fontes de informação (teste 3).

A classe Outros Usos apresentou os melhores resultados para este teste, mas se confundiu com todas as classes, o que havia acontecido apenas nos testes 1 e 5. O EO aconteceu na ordem de 28,6% com as três classes, sendo de 7,1% com as classes Seringueira e Eucalipto e de 14,3% com a classe Nativa. Já o EC aconteceu apenas com a classe Nativa no valor de 9,1%.

Dessa forma, a AG deste teste foi de 62,2%, menor AG identificada até então.

Teste 8: Fonte de informação R – total de 11 atributos espaciais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 8, para Seringueira percebe-se que ela continuou sendo confundida com as outras três classes, assim como o identificado no teste 7. O EO da classe Seringueira aumentou ainda mais quando comparado com o teste 7, passando de 67% para 83,3%. A omissão aconteceu nas classes Nativa (50%) e Outros Usos (33,3%). O EC da classe Seringueira triplicou, passando de 25% para 75%. A inclusão aconteceu nas classes Eucalipto (25%) e Outros Usos (50%). Os resultados continuam apontando para a incapacidade dos atributos espaciais em diferenciar a classe Seringueira das demais classes arbóreas, assim como a classe Seringueira da classe Outros Usos na fase de desfolha.

A classe Nativa também continuou se confundindo com todas as classes e apresentou acurácias menores do que as identificadas no teste 7, mas ainda maiores do que as da classe Seringueira. O EO duplicou em comparação com o teste 7, passando de

28,6% para 57,9%, e continuou acontecendo nas classes Eucalipto (26,3%) e Outros Usos (31,6%). O EC também aumentou, mas em menor escala, passando de 37,5% para 50%, continuando nas classes Seringueira (18,7%) e Outros Usos (31,3%). Ao contrário do identificado no teste 7, os maiores erros aconteceram na classe Outros Usos (vegetação rasteira). A inconsistência desses resultados indica a dificuldade dos atributos espaciais de separar classes de vegetação, principalmente quando os atributos são gerados de semivariogramas extraídos apenas de valores da fonte de informação R.

A classe Eucalipto continuou sendo completamente excluída pelo classificador, mantendo o mesmo padrão identificado no teste 7. Assim como no teste 7, os resultados para a classe Eucalipto indicam que uma classificação baseada apenas em atributos espaciais, em especial em atributos espaciais extraídos apenas da fonte de informação R, não é viável.

A classe Outros Usos continuou se confundindo com todas as classes, assim como o identificado no teste 7. O EO aumentou muito, passando de 28,6% para 72,7%. Continuou acontecendo com as três classes, sendo na ordem de 18,2% com a classe Seringueira, 9,1% com a classe Eucalipto e 45,4% com a classe Nativa. O EC também apresentou grande queda, passando de 9,1% para 72,7%. A inclusão aconteceu com as classes Seringueira (18,2%) e Nativa (54,5%).

Dessa forma, em comparação com o teste 7 a AG desse teste apresentou um valor menor ainda, quase duas vezes menor. A AG passou de 62,2% para 32,4%, e para a classe Seringueira, que é o alvo do estudo, os resultados foram piores ainda com a AP duas vezes menor e a AU três vezes menor.

Tanto nos testes 7 e 8 a classe Eucalipto foi excluída pelo classificador, então independente da AG apresentada, esse resultado por si só já é suficiente para desqualificar a utilização isolada de atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 7), ou apenas do R (teste 8).

Teste 9: Fonte de informação NIR – total de 11 atributos espaciais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 9, para Seringueira percebe-se que ela continuou sendo confundida com as outras três classes, assim como o identificado nos testes 7 e 8. O EO da classe Seringueira continuou apresentando o mesmo valor do teste 7 de 83,3%. Mas diferindo do teste 7, além das classes Nativa (41,6%) e Outros Usos (16,7%), a omissão aconteceu também

com a classe Eucalipto (25%). O EC da classe Seringueira diminuiu, passando de 75% para 50%. A inclusão aconteceu apenas na classe Outros Usos, diferente do teste 7, onde aconteceu também na classe Eucalipto. Os resultados continuam corroborando para a conclusão de que atributos espaciais isolados isoladamente são incapazes de diferenciar a classe Seringueira das demais classes arbóreas, assim como a classe Seringueira da classe Outros Usos na fase de desfolha.

A classe Nativa continuou se confundindo com todas as classes, da mesma forma que nos testes 7 e 8. No teste 9 a classe Nativa apresentou os mesmos valores do teste 7, com as melhores acurácias entre as quatro classes. O EO diminuiu em comparação com o teste 8, passando de 57,9% para 28,6%, e continuou acontecendo nas classes Eucalipto (14,3%) e Outros Usos (14,3%). O EC também diminuiu, mas em menor escala, passando de 50% para 37,5%, nas classes Seringueira (31,2%) e Eucalipto (6,3%).

A classe Eucalipto continuou sendo completamente excluída pelo classificador, mantendo o mesmo padrão identificado nos testes 7 e 8, reforçando a conclusão de que uma classificação baseada apenas em atributos espaciais não é viável, mas não apenas pelos baixos valores das acurácias, e também pela exclusão de uma classe.

A classe Outros Usos continuou se confundindo com as outras três classes, assim como nos testes 7 e 8. Em comparação com o teste 8 o EO diminuiu muito, passando de 72,7% para 30%. A omissão aconteceu com as classes Seringueira (20%) e Eucalipto (10%). O EC também apresentou grande queda, passando de 72,7% para 36,4%. A inclusão aconteceu com as classes Seringueira (18,2%) e Nativa (18,2%).

Dessa forma, em comparação com o teste 8 a AG desse teste apresentou considerável aumento. A AG passou de 32,4% para 51,4%. As classes Nativa e Outros Usos apresentaram resultados interessantes quando comparados ao teste 8. Nessas classes identificaram-se melhores resultados e, em uma análise isolada das classes, de certa forma aceitável. Mas, o Eucalipto continuou sendo excluído pelo classificador, então mantendo a decisão do teste 7, independente da AG apresentada, esses resultados desqualificam a utilização isolada de atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 7), apenas do R (teste 8) ou apenas do NIR (teste 9).

Outro ponto interessante revelado por esses resultados é a similaridade entre as acurácias da classe Nativa nos testes 7 e 9. No teste 7 foram utilizadas as fontes de informação R, NIR e NDVI para extrair os atributos espaciais, já no teste 9 foi utilizada

apenas a fonte de informação R. Esse resultado pode indicar que para a classe Nativa, a fonte de informação com maior capacidade de fornecer atributos espaciais mais representativos e importantes à classificação é a fonte de informação R.

Teste 10: Fonte de informação NDVI – total de 3 atributos espaciais.

Examinando os valores fora da diagonal principal da matriz de confusão do teste 10, para Seringueira percebe-se que ela continuou sendo confundida com as classes Nativa e Outros Usos, mas diferente dos testes 7, 8 e 9, nesse teste ela deixou de se confundir com a classe Eucalipto.

Comparando-se ao teste 9, o EO da classe Seringueira apresentou queda, passando de 83,3% para 75%. A omissão aconteceu nas classes Nativa (50%) e Outros Usos (25%). O EC da classe Seringueira continuou da mesma ordem do identificado no teste 9, no valor de 50%. A inclusão também aconteceu nas classes Nativa (25%) e Outros Usos (25%).

A classe Nativa continuou se confundindo com todas as classes, da mesma forma que nos testes 7, 8 e 9. No teste 10 a classe Nativa apresentou acurácias menores, quando comparado ao teste 9. O EO duplicou, passando de 28,6% para 59,1%, e aconteceu nas três classes, diferente dos testes 7, 8 e 9 onde a omissão aconteceu apenas nas classes Eucalipto e Outros Usos. Na classe Seringueira a omissão foi da ordem de 4,6%, na classe Eucalipto de 22,7% e na classe Outros Usos de 31,8%. O EC também aumentou, mas em menor escala, passando de 37,5% para 43,7%, nas classes Seringueira (24,9%), Eucalipto (6,3%) e Outros Usos (12,5%).

Diferente do identificado em todos os outros testes puramente espaciais (testes 7, 8 e 9) no teste 10 a classe Eucalipto não foi excluída pelo classificador. Do ponto de vista da exclusão de classes, esse resultado contradiz o identificado anteriormente de que uma classificação baseada apenas em atributos espaciais não é viável por que exclui alguma classe. Mas do ponto de vista dos valores apresentados para as acurácias, o teste 10 continua colaborando com a teoria especulada de que classificações feitas apenas com atributos espaciais não são viáveis.

A classe Outros Usos continuou se confundindo com as outras três classes, assim como nos testes 7, 8 e 9. Em comparação com o teste 9 o EO duplicou, passando de 30% para 60%. A omissão aconteceu com as classes Seringueira (20%) e Nativa (40%). O EC apresentou aumento em maior escala ainda, passando de 36,4% para 81,8%. A

inclusão aconteceu com as classes Seringueira (18,2%) e Nativa (63,6%), assim como no teste 9. Dentre os testes puramente espaciais a classe Outros Usos apresentou nesse teste, junto com o testes 8, os piores valores.

Dessa forma, em comparação com o teste 9 a AG desse teste apresentou considerável queda, passando de 51,4% para 37,8%.

No teste 10 o Eucalipto deixou de ser excluído pelo classificador, Mas por outro lado apresentou a segunda pior AG dentre os quatro testes puramente espaciais, perdendo apenas para o teste 8, onde foi utilizada apenas a fonte de informação R para extrair os atributos espaciais. Então mantendo o formulado em todos os quatro testes puramente espaciais, independente da AG apresentada, esses resultados desqualificam a utilização isolada de atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 7), apenas do R (teste 8), apenas do NIR (teste 9) ou apenas do NDVI (teste 10).

Mas deve-se ressaltar a quantidade de atributos utilizados no teste 10. Enquanto o teste 7 utilizou 25 atributos e os testes 8 e 9 utilizaram 11 atributos cada, o teste 10 utilizou apenas três atributos. Esse resultado traz novamente uma das questões levantadas na subseção 2.1.2 e discutidas na subseção anterior (2.1.4.2). ‘Quanto mais atributos, melhor a AG da classificação?’. O teste 8 e 10 apresentaram as piores AGs, no entanto, o teste 8 utilizou oito atributos a mais que o teste 10, e ainda assim o teste 10 apresentou (37,8%) AG superior ao teste 8 (32,4%). Esses resultados corroboram com os encontrados no final da subseção dedicada a Comparação 2 (2.1.4.2), de que não é a quantidade de atributos que importa, e sim a qualidade, a capacidade da fonte de informação de identificar determinada classe. No caso da classificação 10 essa questão fica bem clara, já que quando a fonte de informação NDVI foi utilizada isoladamente para extrair atributos espaciais a classe Eucalipto não foi excluída pelo classificador, mas quando ela foi utilizada em conjunto com as fontes de informação R e NIR (teste 7) essa classe foi excluída.

Um outro ponto onde deve-se fixar atenção é o fato de a única classe excluída ter sido a classe Eucalipto. Pode-se discutir duas possibilidades para explicar esse fato. A primeira é que a classe Eucalipto seja sensível a atributos espaciais, e sua identificação utilizando apenas tais atributos é difícil de ser alcançada. A segunda é que a amostragem pode não ter sido representativa pra esta classe, ou que a amostragem pode

não ter captado bem a variação dessa classe na área de estudo e isso tenha confundido o classificador.

De acordo com os resultados discutidos na presente subseção é impossível eleger o melhor teste, tendo em vista que nenhum dos testes simulados foi capaz de fornecer acurácias aceitáveis e/ou identificação das quatro classes de uso do solo abordadas. Por esse motivo na subseção seguinte, destinada aos testes espectrais-espaciais, as fontes de informação destinadas a extração de atributos espaciais foram utilizadas, e alternadas, duas a duas, juntamente com atributos espectrais extraídos das fontes de informação NIR e NDVI, que foram as fontes que proporcionaram melhor identificação das classes dentre os seis testes puramente espectrais simulados.

2.1.4.4. Comparação 4

A comparação apresentada nesta subseção foi realizada entre testes espectrais-espaciais (11, 12, 13 e 14), que foram gerados a partir dos resultados obtidos nas comparações anteriores para esclarecer dúvidas a respeito do relacionamento entre atributos espaciais e espectrais e como eles influenciam a separabilidade das classes de estudo, e conseqüentemente as acurácias. As informações dos testes estão apresentadas na Tabela 11 e as matrizes de confusão na Tabela 12.

Tabela 11. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global dos testes confrontados na Comparação 4.

COMPARAÇÃO 4		Teste			
		11	12	13	14
Nº de atributos utilizados		29	26	18	18
Polígonos classificados	Correto	35	35	36	35
	Incorreto	2	2	1	2
Seringueira	AP	80	80	100	80
	AU	100	100	100	100
Nativa	AP	94,1	94,1	94,1	94,1
	AU	100	100	100	100
Eucalipto	AP	100	100	100	100
	AU	66,7	66,7	83,3	66,7
Outros Usos	AP	100	100	100	100
	AU	100	100	100	100
Acurácia Global		94,6	94,6	97,3	94,6

Tabela 12. Matrizes de confusão dos testes 11, 12, 13 e 14 (acurácias do produtor nas colunas e acurácias do usuário nas linhas).

11	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	4	0	0	0
Nativa	0	16	0	0
Eucalipto	1	1	4	0
Outros Usos	0	0	0	11

12	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	4	0	0	0
Nativa	0	16	0	0
Eucalipto	1	1	4	0
Outros Usos	0	0	0	11

13	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	4	0	0	0
Nativa	0	16	0	0
Eucalipto	0	1	5	0
Outros Usos	0	0	0	11

14	Seringueira	Nativa	Eucalipto	Outros Usos
Seringueira	4	0	0	0
Nativa	0	16	0	0
Eucalipto	1	1	4	0
Outros Usos	0	0	0	11

Teste 11: Fontes de informação de atributos espectrais NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) e fontes de informação de atributos espaciais R, NIR e NDVI – total de 29 atributos, 4 espectrais e 25 espaciais (11 do R, 11 do NIR e três do NDVI).

Examinando os valores fora da diagonal principal da matriz de confusão do teste 11, para a classe Seringueira percebe-se que quando foram utilizados atributos espectrais (extraídos das fontes de informação NIR e NDVI) e espaciais (extraídos das fontes de informação R, NIR e NDVI) em conjunto, ela se confundiu apenas com a classe Eucalipto, assim como nos testes 2, 3, 4 e 6. É interessante atentar ao fato de que a confusão dessa classe no teste 11 aconteceu de forma similar a identificada no teste 6, onde foram utilizados os mesmos atributos espectrais. Dessa forma, tanto o EO quanto o EC foram idênticos aos do teste 6. O EO foi da ordem de 20% na classe Eucalipto. Não

houve EC por que a classe Seringueira não foi incluída em nenhuma outra classe. Os resultados deste teste para esta classe, juntamente com os identificados no teste 6 que foram similares, foram os melhores dentre todos os testes discutidos até então.

A classe Nativa também foi confundida apenas com a classe Eucalipto, o que tinha acontecido apenas no teste 2, apresentando AP de 94,1 %, maior ainda que a da classe Seringueira (80%). Assim como na classe Seringueira, na classe Nativa não houve EC, resultado de Uma AU máxima. O EO aconteceu na ordem de 5,9%.

A classe Eucalipto apresentou resultados similares aos encontrados para o teste 6, sendo confundida com as classes Seringueira e Nativa, assim como nos testes 3, 4, 5 e 6. A classe não apresentou EO, não sendo omitida em nenhuma outra classe. Com relação ao EC, aconteceu na ordem de 33,3%, sendo 16,6% tanto para a classe Seringueira quanto para a classe Nativa. Até então os melhores resultados para esta classe haviam sido encontrados nos testes 2 e 6.

Para a classe Outros Usos, o padrão encontrado nos testes 2, 3, 4 e 6 voltou a acontecer, com a classe apresentando acurácias máximas, não sendo confundida com nenhuma outra classe.

Dessa forma, a AG deste teste foi de 94,6%, apresentando, junto com os testes 2 e 6, a maior AG encontrada até então.

O valor de todas as APs e AUs foi idêntico aos valores encontrados no teste 6, que utilizou os mesmos atributos espectrais utilizados no teste 11. A Diferença entre os testes 6 e 11 é que no teste 11 utilizou-se também atributos espaciais extraídos das fontes R, NIR e NDVI. Dessa forma, esse resultado pode sugerir que os atributos espaciais aqui utilizados não fizeram diferença nenhuma.

Caso a especulação feita anteriormente seja confirmada, a melhor estratégia seria utilizar apenas os atributos espectrais (teste 6), já que foram adicionados 25 atributos espaciais e não houve impacto nenhum nas acurácias. Concluiu-se nas Comparações 2 e 3 que não é a quantidade de atributos que importa e sim a qualidade, logo se os atributos adicionados não surtiram efeito na AG devem ser retirados das análises.

Teste 12: Fontes de informação de atributos espectrais NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) e fontes de informação de atributos espaciais R e NIR – total de 26 atributos, 4 espectrais e 22 espaciais (11 do R e 11 do NIR).

Examinando os valores fora da diagonal principal da matriz de confusão do teste 12, percebe-se que todos os resultados foram idênticos aos identificados nos testes 6 e 11. Esse resultado suscita um raciocínio interessante. Ao adicionarem-se atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 11) ou apenas as fontes de informação R e NIR (teste 12) os resultados são os mesmos obtidos quando não foram utilizados atributos espaciais.

A ideia formulada no final do teste 11 começa a ganhar aqui um corpo mais conciso, pois mesmo ao se retirar a fonte do NDVI das análises os resultados permaneceram intactos. Isso pode significar que além de a inclusão de atributos espaciais não ter surtido efeito positivo, eles também não atrapalharam. Esse resultado pode indicar que atributos espectrais são menos volúveis do que atributos espaciais, pois na comparação destinada aos testes puramente espaciais (Comparação³ – subseção 2.1.4.3) percebemos que atributos dessa natureza tendem ser muito delicados e qualquer manipulação nos atributos utilizados causou grande variação nas acurácias.

Assim, como o formulado ao fim dos resultados do teste 11, a melhor estratégia neste caso parece ser a exclusão dos atributos espaciais, pois fica cada vez mais claro que a utilização de atributos espaciais não surtiu efeito nenhum nas acurácias quando são utilizados em conjunto com atributos espectrais. A diferença básica entre este teste e o teste 11 é que no teste 11 foram adicionados 25 atributos espaciais, já neste foram adicionados 22, já que a fonte de informação NDVI foi retirada das análises.

Teste 13: Fontes de informação de atributos espectrais NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) e fontes de informação de atributos espaciais R e NDVI – total de 18 atributos, 4 espectrais e 14 espaciais (11 do R e três do NDVI).

Examinando os valores fora da diagonal principal da matriz de confusão do teste 13, para a classe Seringueira percebe-se que quando foram utilizados atributos espectrais (extraídos das fontes de informação NIR e NDVI) e espaciais (extraídos das fontes de informação R e NDVI) em conjunto, ela não se confundiu com nenhuma outra

classe, não apresentando, portanto, nem EO e nem EC. Esse resultado não havia sido identificado em nenhum outro teste, sendo o melhor resultado para a classe Seringueira dentre todos os outros testes.

A classe Nativa também foi confundida apenas com a classe Eucalipto, assim como nos testes 2, 11 e 12, apresentando AP de 94,1%. A classe Nativa não apresentou EC, resultado de uma AU máxima. O EO aconteceu na ordem de 5,9%. Os resultados para a classe Nativa se mantiveram constantes nos testes 6, 11, 12 e 13. Isso pode indicar que essa classe não é sensível a inclusão e/ou exclusão dos atributos espaciais extraídos das fontes de informação R, NIR e NDVI.

A classe Eucalipto apresentou no teste 13 os melhores resultados dentre todos os testes simulados até então. A classe foi confundida apenas com a classe Nativa. Assim como nos testes 6, 11 e 12, a classe Eucalipto no teste 13 não apresentou EO, e o EC foi o menor identificado para esta classe, na ordem de 16,7%.

A classe Outros Usos continuou apresentando o padrão encontrado nos testes 2, 3, 4, 6, 11 e 12, com acurácias máximas, sem ser confundida com nenhuma outra classe.

Dessa forma, a AG deste teste passou de 94,6% para 97,3%, apresentando a maior AG identificada.

Ao serem adicionados atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 11) ou apenas as fontes de informação R e NIR (teste 12) os resultados foram os mesmos obtidos quando não foram utilizados atributos espaciais (teste 6). Mas, ao serem adicionados atributos espaciais extraídos das fontes de informação R e NDVI (teste 13) obtiveram-se acurácias melhores e mais balanceadas para todas as classes.

Esses resultados retornam aos resultados obtidos na subseção 2.1.4.2 (Comparação 2), onde foi identificado que existem fontes de informação que são capazes fornecer melhores atributos, do ponto de vista da identificação das classes de uso do solo, do que outras. A Comparação 2 foi feita entre testes puramente espectrais, e foi identificado que os melhores atributos espectrais para serem utilizados para trabalhar com as classes de uso do solo selecionadas (Seringueira, Nativa, Eucalipto e Outros Usos) foram os atributos espectrais extraídos das fontes de informação NDVI e NIR (teste 6), por que apresentaram resultados similares de quando os atributos foram extraídos das sete fontes de informação (teste 2), mas com a vantagem de utilizar um menor número de atributos.

Mesmo que os resultados encontrados na Comparação 2 tenham sido formulados para testes puramente espectrais podem ser aplicados na presente subseção, por que quando foram utilizados atributos espectrais e espaciais em conjunto houve diferença nas acurácias, conforme as fontes de informação utilizadas para extrair atributos espaciais, resultado diferente do identificado nos testes 11 e 12.

No final dos resultados dos testes 11 e 12 iniciou-se a discussão a respeito do efeito da inclusão de atributos espaciais às análises. Nesses testes foi identificado que a inclusão de atributos espaciais não surtiu efeito positivo e nem negativo, sugerindo que a melhor estratégia fosse a exclusão de atributos espaciais das análises, utilizando apenas atributos espectrais nas classificações. Mas ao alternarmos mais umas vezes as fontes de informação para extração de atributos espaciais, utilizando as fontes R e NDVI, percebeu-se que houve uma interação positiva entre os atributos extraídos das duas fontes de informação, o que causou aumento da AG.

Provavelmente não foi identificada interação negativa entre os atributos extraídos das fontes de informação (R, NIR e NDVI) ao se alternar as fontes de informação utilizadas, por que conforme a ideia formulada no final dos resultados do teste 12, os atributos espectrais parecem ser menos volúveis do que atributos espaciais, sendo os atributos de natureza espacial muito delicados, onde qualquer manipulação nos atributos utilizados pode causar grande variação nas acurácias. Em outras palavras, é necessária interação negativa entre atributos espaciais muito mais forte para que as acurácias de classificações espectrais-espaciais apresente queda, e isso se deve, provavelmente, à consistência que atributos espectrais conferem a classificação.

Devido aos resultados apresentados pelo teste 13, podemos afirmar que a melhor estratégia, até então, é utilizar atributos espectrais extraídos das fontes de informação NIR e NDVI em conjunto com atributos espaciais extraídos das fontes de informação R e NDVI.

Teste 14: Fontes de informação de atributos espectrais NIR e NDVI com adição da componente temporal (utilizando imagens das duas datas) e fontes de informação de atributos espaciais NIR e NDVI – total de 18 atributos, 4 espectrais e 14 espaciais (11 do NIR e três do NDVI).

Examinando os valores fora da diagonal principal da matriz de confusão do teste 14, percebe-se que, todos os resultados foram idênticos aos identificados nos testes

6, 11 e 12. Ao adicionarem-se atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 11), ou apenas atributos extraídos das fontes de informação R e NIR (teste 12), ou ainda apenas atributos extraídos das fontes de informação NIR e NDVI (teste 14) os resultados são os mesmos obtidos quando não foram utilizados atributos espaciais.

A classe Seringueira apresentou ótimos resultados na Comparação 4, se mostrando pouco sensível do ponto de vista de manipulação das fontes de informação para extração de atributos espaciais. Mas mesmo se mostrando pouco sensível, a melhor abordagem para identificação desta classe é utilizar em conjunto com atributos espectrais extraídos das fontes de informação NIR e NDVI, atributos espaciais extraídos das fontes de informação R e NDVI (teste 13).

A classe Eucalipto foi a mais sensível a manipulação das fontes de informação para extração de atributos espaciais dentre as quatro classes. Os resultados evidenciados para esta classe desconstrói a conclusão formulada de que a classe Eucalipto é muito sensível a exclusão dos atributos espaciais gerados a partir de valores das fontes de informação R, NIR e NDVI, levando a acreditar que na verdade essa classe não é sensível da mesma maneira as fontes de informação utilizadas para extração de valores para geração dos atributos espaciais. Esses resultados sugerem que a melhor abordagem para se trabalhar com a classe Eucalipto é utilizar em conjunto com atributos espectrais extraídos das fontes de informação NIR e NDVI, atributos espaciais extraídos das fontes de informação R e NDVI (teste 13), mesma conclusão alcançada para a classe Seringueira.

As classes Nativa e Outros Usos não se mostraram sensíveis a manipulação e/ou alteração das fontes de informação utilizadas para extração dos atributos espaciais. Essas classes não demonstraram oscilação nas acurácias nos testes 6, 11, 12, 13 e 14.

A ideia formulada no final dos testes 11 e 12 é reforçada aqui, mas após os resultados identificados para o teste 13 ganha uma nova interpretação. Nos testes 11 e 12 sugeriu-se que a inclusão de atributos espaciais não surtiu efeito nenhum nas classificações. Esse resultado foi evidenciado também no teste 14, que obteve os mesmos resultados dos testes citados apenas utilizando as fontes de informação do NIR e do NDVI para extrair atributos espaciais. Mas no teste 13 identificou-se aumento da AG e APs e AUs mais balanceadas para todas as classes.

Isso significa que as fontes de informação R e NDVI (teste 13) em conjunto fornecem atributos espaciais mais robustos para identificação das classes Seringueira, Nativa, Eucalipto e Outros Usos, do que utilizar apenas atributos espectrais (teste 6), utilizar em conjunto com os atributos espectrais atributos espaciais extraídos das fontes de informação R, NIR e NDVI (teste 11), utilizar em conjunto com os atributos espectrais atributos espaciais extraídos das fontes de informação R e NIR (teste 12), ou ainda que utilizar em conjunto com os atributos espectrais atributos espaciais extraídos das fontes de informação NIR e NDVI (teste 14), colaborando com o sugerido ao final dos resultados do teste 13.

2.2. Conclusões

A inclusão da componente temporal aos testes puramente espectrais diminuiu a confusão da classe Seringueira com a classe Outros Usos, indicando que esta componente cumpriu seu dever no auxílio à identificação da classe Seringueira e separabilidade desta classe da classe Outros Usos, devido à deciduidade.

A exclusão das fontes de informação G, B, SWIR 1 e SWIR 2 resultou em acurácias mais equilibradas.

As fontes de informação NDVI e NIR fornecem os melhores atributos espectrais para descrição das classes em classificações puramente espectrais.

Foi identificada interação negativa entre atributos espectrais gerados de diferentes fontes de informação, onde destaca-se a interação negativa entre as fontes NDVI e R, e interação positiva entre as fontes NDVI e NIR, em classificações puramente espectrais.

As classes demonstraram sensibilidade diferente à retirada e/ou alternância das fontes de informação utilizadas para geração de atributos espaciais, sendo a classe Eucalipto a mais sensível à manipulação das fontes de informação para extração de atributos espaciais.

Não houve classificação puramente espacial simulada capaz de fornecer acurácias aceitáveis e/ou identificação das quatro classes de uso do solo abordadas, devido, principalmente, à incapacidade de separar classes arbóreas entre si, e a classe Outros Usos da classe Seringueira.

A qualidade dos atributos e das fontes de informação é mais importante do que a quantidade dos atributos, onde qualidade refere-se à capacidade das fontes de informação e seus atributos de identificar determinada classe.

Os melhores resultados foram encontrados em classificação espectral-espacial que utilizou as fontes de informação NDVI e NIR para extrair atributos espectrais em conjunto com as fontes de informação NDVI e R para extrair atributos espaciais.

REFERÊNCIAS

- BALAGUER, A.; RUIZ, L. A.; HERMOSILLA, T.; RECIO, J. A. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. **Computers & Geoscience**, Houston-TX, v. 36, p. 231-240, 2010.
- BATISTA, M. H. **Classificação hierárquica orientada a objeto em imagens de alta resolução especial empregando atributos espaciais e espectrais**. 2006. 199 p. Dissertação (Mestrado em Sensoriamento Remoto) – Universidade Federal do Rio Grande do Sul – UFRGS, Porto Alegre-RS, 2006.
- BELGIU, M.; DRAGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 114, p. 24–31, 2016.
- CHEN, B.; CAO, J.; WANG, J.; WU, Z.; TAO, Z.; CHEN, J.; YANG, C.; XIE, G. Estimation of rubber stand age in typhoon and chilling injury afflicted area with Landsat TM data: A case study in Hainan Island, China. **Forest Ecology and Management**, Flagstaff-AZ, v. 274, p. 222–230, 2012.
- CURRAN, P. The semivariogram in remote sensing: an introduction. **Remote Sensing of Environment**. Toronto-ON, v. 24, p. 493–507, 1988.
- DEVRIES, B.; PRATIHAST, A. K.; VERBESSELT J.; KOOISTRA L.; HEROLD, M. Characterizing forest change using community-based monitoring data and Landsat time series. **Plos One**, College Park-MD, v. 11, n. 3, 2016.
- DONG, J.; XIAO, X.; CHEN, B.; TORBICK, N.; JIN, C.; ZHANG, G.; BIRADAR, C. Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. **Remote Sensing of Environment**, New York-NY, v. 134, p. 392–402, 2013.
- EKADINATA, A.; WIDAYATI, A.; VINCENT, G. Rubber agroforest identification using object-based classification in Bungo District, Jambi, Indonesia. In: ASIAN CONFERENCE ON REMOTE SENSING, 25., 2004, Chiang Mai-Thailand. **Anais...** Chiang Mai-Thailand: ACRS, 2004. p. 551-556
- ESPERANTE, D. **O que esperar dos preços da borracha natural para 2016?** São Paulo-SP: Seringueira, 2015. Disponível em: <http://seringueira.com.br/noticia_seringueira_140116.php> Acesso em: 18 fev. 2016.
- GHOSH, A.; JOSHI, P. K. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. **International Journal of Applied Earth Observation Geoinformation**, Enschede-Overijssel, v. 26, p. 298–311, 2014

INMET – INSTITUTO NACIONAL DE METEOROLOGIA. **Série histórica** – Dados diários, 2018. Disponível em: <<http://www.inmet.gov.br/portal/>>. Acesso em: 01 jul. 2018.

JUSOFF, K.; YUSOFF, M. H. M. New approaches in estimating rubberwood standing volume using airborne hyperspectral sensing. **Modern Applied Science**, North York-ON, v. 3, n. 4, 62-70, 2009.

LAWRENCE, R. L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). **Remote Sensing**, v. 100, p. 356–362, 2006.

LEAL, S. T. **A heveicultura na mesorregião leste do Estado de Mato Grosso do Sul: Aspectos técnicos e econômicos**. 2017. 108 p. Tese (Mestrado em Agronomia) - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, Ilha Solteira, 2017.

LI, Z.; FOX, J. M. Integrating Mahalanobis typicalities with a neural network for rubber distribution mapping. **Remote Sensing Letters**, Nottingham: UK, v. 2, p. 157-166, 2011.

MIRANDA, F. P., FONSECA, L. E. N., CARR, J. R. Semivariogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest. **IEEE International Journal of Remote Sensing**, London-UK, v. 19, n. 3, p. 549–556, 1998.

MUI, A.; HE, Y.; WENG, Q. An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. **International Journal of Geographical Information Science**, London-UK, v. 109, p. 30–46, 2015.

PENNACCHIO, H. L. **Borracha natural**. Brasília-DF: Companhia Nacional de Abastecimento - CONAB, 2014. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_07_16_13_52_53_borrach_ajunho2014.pdf> Acesso em: 20 jun. 2018.

RUIZ, L. A.; RECIO, J. A.; HERMOSILLA, T. A feature extraction software tool for agricultural object-based image analysis. **Computers and Electronics in Agriculture**, Amsterdã-Holanda do Norte, v. 76, p. 284-296, 2011.

SILVA, A. F. **Geoestatística e sensoriamento remoto na classificação de imagens em áreas cultivadas com citros**. 2011. 96 p. Dissertação (Mestrado em Agronomia) – Universidade Estadual Paulista “Júlio de Mesquita” – UNESP, Botucatu, 2011a.

SILVA, J. S. V.; SPERANZA, E. A.; VENDRUSCULO, L. G.; ESQUERDO, J. C. D. M.; MAURO, R. A.; BIANCHINI, S. L.; FLORENCE, R. O. **Projeto Geo MS: melhorando o Sistema de Licenciamento Ambiental do Estado do Mato Grosso do Sul**. Campinas/SP: Embrapa Informática Agropecuária, 2011b. 64 p.

SILVEIRA, E. M. O.; MENEZES, M. M.; ACERBI JÚNIOR, F. W.; TERRA, M. C. N. S.; MELLO, J. M. Assessment of geostatistical features for object-based image

classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Orlando-FL, v. 11, n. 3, 2017.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing I: scene models and simulated images. **Remote Sensing of Environment**. Toronto-ON, v. 25, p. 323–348, 1988a.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing II: real digital images. **Remote Sensing of Environment**. Toronto-ON, v. 25, p. 323–34349–3798, 1988b.

WU, X.; PENG, J.; SHAN, J.; CUI, W. Evaluation of semivariogram features for object based image classification. **Geo-spatial Information Science**, Londres-UK, v. 18, n. 4, p. 159-170, 2015.

CAPÍTULO 3. AVALIAÇÃO DE GRUPOS DE ATRIBUTOS ESPACIAIS INSERIDOS NA CLASSIFICAÇÃO DIGITAL DE BASE ESPECTRAL DE PLANTIOS DE SERINGUEIRA NO MATO GROSSO DO SUL

RESUMO: A heveicultura vem crescendo de forma expressiva no estado do Mato Grosso do Sul, e para auxiliar no monitoramento dessas áreas imagens de satélite podem ser utilizadas. O mapeamento de seringais é um desafio, principalmente devido à deciduidade apresentada pela espécie na época seca, que altera a resposta dos alvos à radiação eletromagnética, devido à ausência e presença de folhas ao longo do ano. Entretanto, para tentar contornar o problema de ambiguidade espectral dos dados, a mesma característica pode auxiliar na identificação dos plantios de Seringueira ao serem incluídos nas análises dados auxiliares como, por exemplo, componentes temporais por meio de diferentes datas de análise e atributos espaciais extraídos de valores de semivariogramas gerados a partir da variação dos valores de reflectância dos *pixels* por objeto. Assim, o objetivo do presente trabalho é ampliar o estudo da incorporação de atributos espaciais no processo de classificação de base espectral, buscando aumentar a separabilidade de quatro principais classes de estudo (Seringueira, Nativa, Eucalipto e Outros Usos) e estudar a utilização do *ranking* de importância dos atributos fornecido pelo algoritmo de classificação. Para isso foi utilizada uma abordagem orientada à objeto e imagens do sensor OLI/Landsat 8. Os atributos espectrais foram mantidos constantes, sendo extraídos de valores de reflectância de superfície presentes nas bandas espectrais do vermelho e infravermelho próximo. Os atributos espaciais variaram, e para isso foram extraídos de valores de reflectância de superfície presentes nas bandas espectrais do vermelho e do infravermelho próximo, e do índice de vegetação NDVI. A classificação foi realizada através do algoritmo *Random Forest*. Como resultados principais foi observado que, de maneira geral, os atributos espaciais mais interessantes a serem empregados em classificações espectrais-espaciais são os atributos do primeiro grupo, e que o grupo a qual os atributos espaciais utilizados pertencem é mais importante do que a quantidade de atributos espaciais e do que a fonte de informação utilizada.

Palavras-chave: sensoriamento remoto, random forest, *hevea brasiliensis*, landsat 8, semivariograma, heveicultura.

CHAPTER 3. EVALUATION OF SPACE ATTRIBUTES GROUPS INCORPORATED IN THE DIGITAL SPECTRAL BASE CLASSIFICATION OF RUBBER TREE CULTIVATION IN MATO GROSSO DO SUL

ABSTRACT: Rubber tree cultivation has been expanding in the state of Mato Grosso do Sul and is mostly monitored remotely via satellite imaging. Defoliation in the dry season however alters response to electromagnetic radiation due to the absence and presence of leaves throughout the year. Meanwhile, to avoid spectral ambiguity of the data, the same characteristic can aid the identification of Rubber Tree by including auxiliary data such as temporal components with different dates of analysis and spatial attributes extracted from semivariogram values generated from the variation of the reflection values of pixels per object. Present work intends to expand the study of incorporation of spatial attributes study in spectral-based classification, improving separability of the four study classes (Rubber Tree, Eucalyptus, Native and Others Uses) and also prospecting the use of importance ranking attributes provided by the classification algorithm. For this, an object-oriented approach and OLI/Landsat 8 sensor images were used. The spectral attributes were kept constant, being extracted from surface reflectance values present in spectral bands of red and near infrared. The spatial attributes varied and were extracted from surface reflectance values present in the red and near infrared spectral bands and from the NDVI vegetation index. The classification was performed through the Random Forest algorithm. As the main results it was observed that in general the most interesting spatial attributes to be employed in spectral-spatial classifications are the attributes of the first group, and also that the group to which the spatial attributes used belong is more important than the number of spatial attributes and the source used.

Key words: remote sensing, random forest, *Hevea brasiliensis*, landsat 8, semivariograma, heveculture.

3.1. Introdução

Segundo Leal (2017), a borracha é considerada uma *commodity* estratégica para os mercados globais, sendo a formação dos preços mundiais explicada por uma complexa interação de fatores, e embora o Brasil seja o centro de origem da Seringueira, é considerado um país sem expressão no mercado mundial.

A produção mundial mais significativa é observada nos países asiáticos, que participam em mais de 80% da produção mundial de borracha natural. De acordo com dados do IBGE (2017), a produção brasileira atende a apenas 41,44% da demanda nacional, sendo necessária, segundo Scaloppi *et al.* (2017), uma importação média anual de mais de 500 milhões de dólares.

Com a crescente demanda global por produtos que utilizam borracha natural em suas composições houve em resposta um aumento no número de seringais, que ocorreu em muitas regiões onde originalmente o plantio era considerado inadequado. Assim, a produção brasileira se deslocou da Região Amazônica para outras regiões. No entanto, informações precisas a respeito dos plantios não estão disponíveis, o que limita substancialmente nossa compreensão da expansão da cultura no país.

Segundo Chen *et al.* (2012a) existe uma grande necessidade, tanto do ponto de vista do gerenciamento da produção quanto da modelagem de ecossistemas, de atualização de dados como área, idade, volume e distribuição espacial de plantios de espécies, como por exemplo a Seringueira, cultura que tem demonstrado considerável expressão na balança comercial nacional.

Para Rodriguez-Galiano *et al.* (2012) o monitoramento da cobertura do solo através do sensoriamento remoto (SR) requer métodos de classificação robustos, que executem o mapeamento de paisagens complexas e suas categorias. Nesse contexto, passaram a fazer parte das análises de dados de SR as máquinas de aprendizagem, que adquirem conhecimento de forma automática a fim de formular conclusões sobre um conjunto de dados.

Dentre os inúmeros algoritmos de classificação com base em máquinas de aprendizagem destaca-se o *Random Forest* (RF), que de acordo com Oshiro (2013) é um algoritmo que constrói muitas árvores de decisão que são utilizadas para classificar Unidades Amostrais (UAs) de acordo com a maioria de votos das árvores geradas. O RF é um algoritmo de classificação robusto, não paramétrico, com a habilidade de utilizar

muitas variáveis de predição, que apresenta eficiente implementação em grandes conjuntos de dados e possui uma estrutura que permite o uso de árvores geradas em posteriores aplicações (DEVRIES, 2016). Nas últimas duas décadas a utilização do RF tem ganhado particular atenção por produzir classificações excelentes e apresentar tempos de processamento interessantes (BREIMAN, 2001). Belgiu e Dragu (2016) acreditam que esse classificador lida com sucesso com a alta dimensionalidade e multicolinearidade dos dados de entrada, de uma maneira rápida e sensível ao superajuste.

É considerado o estado da arte no campo da classificação do uso do solo, mas, como os processos de reconhecimento de padrões tradicionais, utilizam essencialmente atributos espectrais para formação de padrões no processo de classificação da paisagem, ignorando a dependência espacial entre os segmentos vizinhos em suas análises.

Para tentar melhorar a identificação das classes de uso do solo essa dependência espacial entre os *pixels* pode ser fornecida ao sistema através da utilização de atributos conhecidos no presente trabalho como espaciais. Para Blaschke (2010) informações espaciais tem sido frequentemente utilizadas para melhorar a performance de classificações orientadas a objeto. Assim, essas novas análises e atributos espaciais foram incorporados aos Sistemas de Informação Geográfica (SIGs), objetivando melhorar os procedimentos tradicionais. Balaguer *et al.* (2010), Wu *et al.* (2015) e Silveira *et al.* (2017) conseguiram melhores performances ao combinar atributos espectrais e atributos espaciais extraídos de semivariogramas na classificação da cobertura do solo.

Um exemplo de atributos espaciais são os 14 atributos propostos por Balaguer *et al.* (2010), que, ao contrário de modelagens convencionais de semivariogramas onde são utilizados os parâmetros básicos de um semivariograma: Efeito Pepita, Patamar e Alcance para extração de informações descritivas, utilizam em seu cálculo parâmetros como o primeiro máximo, primeiro mínimo e segundo máximo. Os atributos foram divididos pelos autores em três grupos, de acordo com a posição na curva do semivariograma onde se encontravam os valores utilizados para a geração de cada atributo. Esses grupos são: Atributos próximos à origem, atributos extraídos de valores até o primeiro máximo e Atributos extraídos de valores entre o primeiro e segundo máximos.

Segundo Balaguer *et al.* (2010) esses atributos resultam em acurácias maiores e mais balanceadas, quando comparadas a utilização de outros métodos de incorporação da dependência espacial entre os valores de reflectância dos *pixels*, por que os parâmetros utilizados agregam mais informações relevantes e representativas para esta aplicação, pois são capazes de identificar pontos importantes e realçar informações contidas em menores distâncias, onde há maior correlação espacial.

Para variáveis contínuas como a reflectância, o semivariograma experimental é definido a partir da variância espacial interna dos objetos, e para trabalhar com essa variância e ser possível gerar semivariogramas é necessário fixar uma distância mínima entre dois pontos, conhecida como distância *lag*. Essa distância corresponde ao eixo x do semivariograma e em análises de imagens de satélite representa a distância entre dois *pixels*, já que o *pixel* é a menor unidade de análise. A distância *lag* não deve ser maior do que a extensão espacial do objeto e nem demasiadamente pequena. Para Woodcock *et al.* (1988a e b), é necessário que o tamanho do objeto seja maior que o alcance do semivariograma para que a primeira parte do mesmo seja caracterizado, e grande o suficiente para revelar periodicidade, que é identificada quando há ocorrência de padrões repetitivos sobre superfícies espacialmente variáveis, como superfícies modificadas pelo homem (áreas urbanas e agrícolas) ou ainda algumas formações florestais que apresentem deciduidade, que é o caso da Seringueira.

Assim, este trabalho teve como objetivo estudar dois grupos de atributos espaciais e avaliar a contribuição de cada grupo na identificação de quatro classes distintas em uma área com ocorrência de plantios de Seringueira no Mato Grosso do Sul, através das diferenças espectrais apresentadas por essas classes, e estudar a utilização do *Ranking* de importância dos atributos, fornecido pelo algoritmo de classificação.

3.2. Material e métodos

3.2.1. Área de Estudo

No Capítulo 2 a fonte de informação NDVI possibilitou a extração de apenas três atributos espaciais, e por esse motivo foi necessário a ampliação da área de trabalho do Capítulo 3. Assim, a área de estudo é composta por duas imagens provenientes do

sensor OLI, à bordo do satélite *Landsat 8*, com órbitas/pontos: 224/73 e 223/73, conforme Figura 14.

Essas cenas abrangem as mesoregiões Leste e Centro-Norte do estado do Mato Grosso do Sul e se localizam entre os paralelos 17°57' e 19°36' de latitude sul e os meridianos 51°02' e 54°08' de longitude oeste. Nela estão contidas três microregiões do estado: Alto-Taquari, Cassilândia e Paranaíba, compreendendo na totalidade os municípios de Costa Rica e Cassilândia, e parcialmente os municípios de Alcinópolis, Figueirão, Camapuã, Paraíso das Águas, Chapadão do Sul, Inocência e Paranaíba.

Como as duas cenas em questão cobrem também o estado e Goiás, foi feito um recorte nas cenas utilizando o limite do estado de Mato Grosso do Sul como limite da área de estudo.

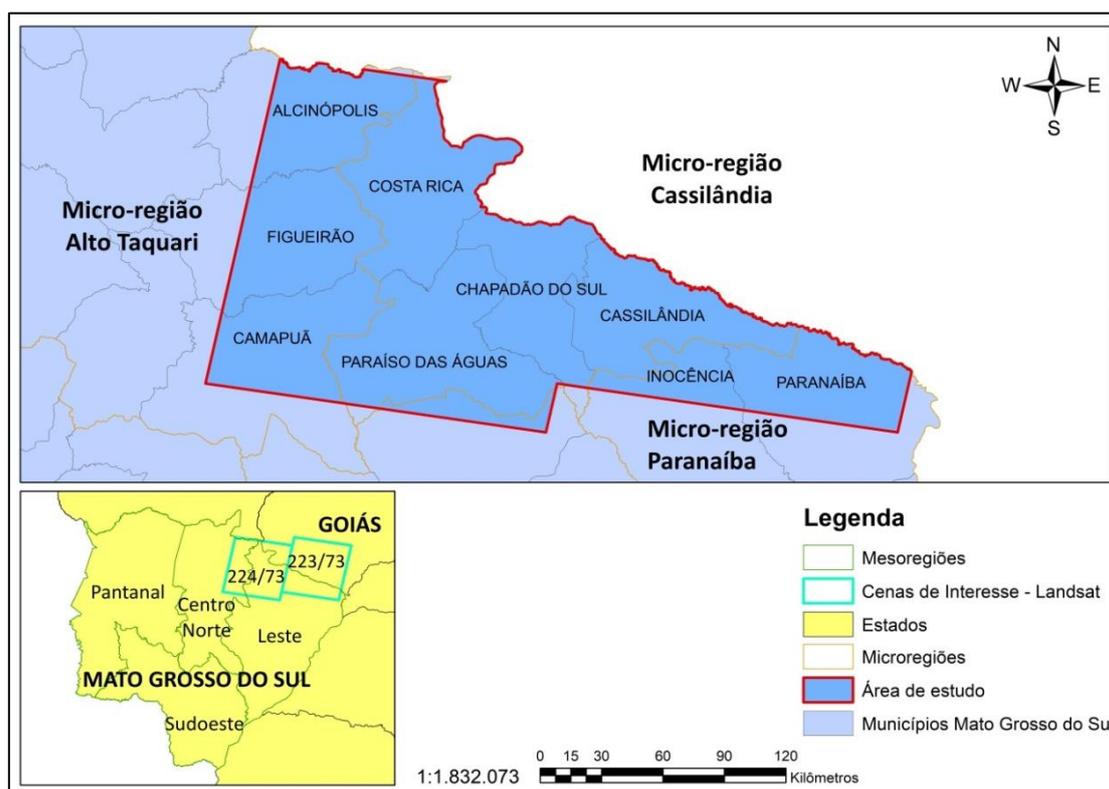


Figura 14. Cenas utilizadas para cobrir área de estudo, localizadas no Mato Grosso do Sul.

O clima da região é, segundo a classificação de Köppen, Aw, caracterizado como clima tropical quente e úmido. A temperatura média é de 26° C, com estação chuvosa no verão e seca no inverno. A precipitação pluvial total anual varia entre 900 e 1.400 mm; e a altitude média, entre 350 e 470 m.

A vegetação do Mato Grosso do Sul é composta por fitofisionomias do Cerrado, Pantanal e Mata Atlântica, sendo que 60,14% da área do estado é ocupada pelo Cerrado (SILVA *et al.*, 2011). A área de estudo, não compreende o Pantanal, ocorrendo apenas Mata Atlântica e Cerrado.

3.2.2. Banco de dados (Imagens e dados)

As imagens foram adquiridas do órgão de Pesquisa Geológica dos Estados Unidos para Observação da Terra e Ciência (USGS/ EROS – do inglês *United States Geological Survey for Earth Observation and Science*) através do site *Earth Explorer* no nível 2 de processamento (*on-demand*), então além de correções geométrica e radiométrica contam também com correção atmosférica, sendo então compostas por valores de reflectância de superfície.

A cena foi adquirida em duas datas, 07/10/2016 e 07/08/2017, possibilitando a inserção de uma componente temporal às análises, responsável por agregar a característica caducifólia da Seringueira, trazendo a informação dos seringueiros completamente enfolhados na primeira data e completamente desfolhados na segunda data. Isso significa que para o mesmo objeto foram extraídos dois atributos espectrais representando a mesma fonte de informação, um para a primeira data e um para a segunda.

Para a escolha das datas de estudo, foram assumidas duas restrições. A primeira foi imposta pela cobertura de nuvens, já que a presença de nuvens não permite a identificação correta dos objetos e nem mesmo a correta segmentação, já que as sombras das nuvens e as próprias nuvens seriam consideradas objetos. A segunda foi imposta pela cultura, e se trata da desfolha natural da Seringueira, que geralmente acontece no início da estação seca, sendo a precipitação o fator de maior influência.

Os dados de precipitação disponibilizados pelo Instituto Nacional de Meteorologia – INMET (2018) para o município de Paranaíba foram utilizados na escolha das datas. De acordo com o gráfico de precipitação observado na Figura 15, o mês da primeira data (07/10/2016 - imagem com plantios enfolhados) foi precedido por chuvas bem distribuídas a partir do final do mês de agosto e, o mês da segunda data (07/08/2017 - imagem com plantios desfolhados) foi precedido de um período de seca severa, que se iniciou no final do mês de maio.

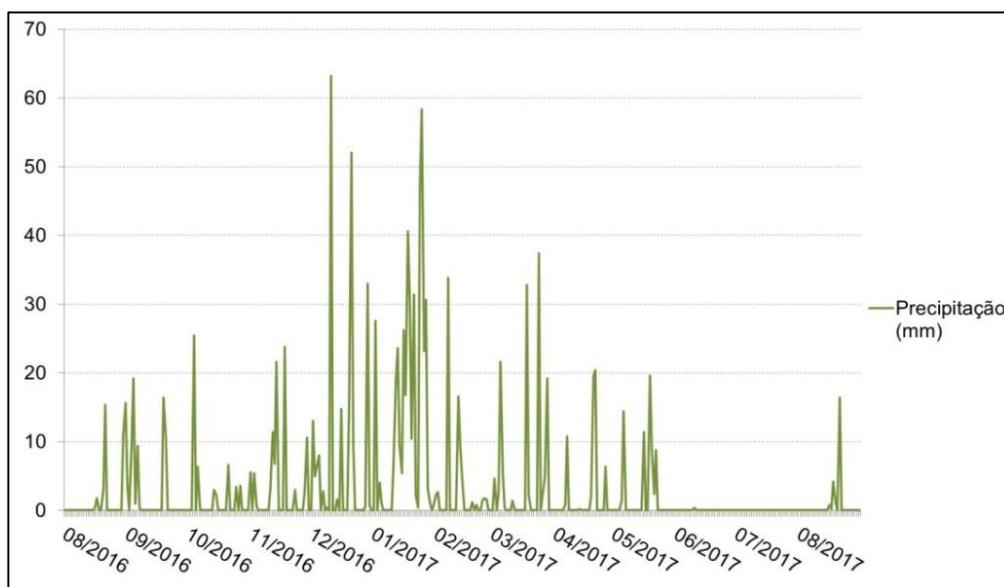


Figura 15. Precipitação (mm) diária da estação meteorológica de Paranaíba – MS, para o período de 01/08/2016 a 31/08/2017.

Portanto, as imagens selecionadas apresentaram cenas sem nuvens, e plantios de Seringueira completamente enfolhados na primeira data e os mesmos plantios totalmente desfolhados na segunda.

3.2.3. Metodologia

A metodologia utilizada foi a mesma utilizada no Capítulo 2 e está resumida no fluxograma apresentado na Figura 16. A primeira etapa foi a segmentação das imagens. Em seguida, foram coletadas as UAs e realizada a extração das características espectrais e espaciais de cada um dos objetos gerados na segmentação. Os atributos deram entrada no algoritmo *Random Forest* (RF) para a classificação. As UAs foram divididas em dois grupos, um grupo de treinamento do classificador e um grupo de avaliação da acurácia das classificações. Dessa forma, nas classificações discutidas neste capítulo 66% das UAs foram utilizadas para treinar o classificador, e os outros 34% para avaliar a acurácia das classificações, que é o último passo da metodologia. Para isso foi utilizada a acurácia geral (AG) das classificações, geradas a partir das acurácias do produtor (AP) e do usuário (AU).

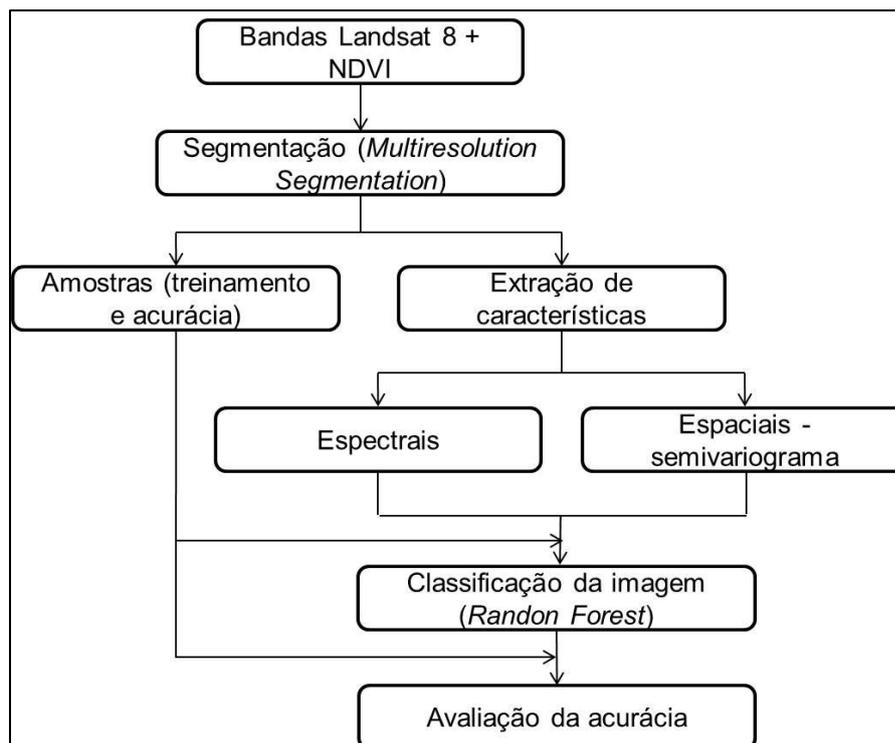


Figura 16. Fluxograma da metodologia utilizada no Capítulo 3.

3.2.3.1. Segmentação

A segmentação divide a imagem em objetos e é o primeiro passo para que seja possível a extração dos atributos por objeto. No presente estudo, as imagens foram segmentadas de acordo com suas características espectrais e espaciais, utilizando o algoritmo de segmentação *Multiresolution Segmentation* do software *Ecognition*, conforme tratado em Silveira *et al.* (2017) e no Capítulo 2 da presente dissertação.

Para a segmentação de imagens é necessário configurar três parâmetros: Forma, Compacidade e Escala, que para com Mui *et al.* (2015), controlam respectivamente o tamanho, a forma e a variação espectral dos objetos criados. O tamanho mínimo dos objetos tem que ser configurado levando em consideração o número mínimo de *pixels* dentro de cada objeto necessário para gerar o semivariograma, já que segundo Silveira *et al.* (2017), a geração de objetos tem que estar conectada a distância *lag*, que é a distância mínima necessária entre os pares de pontos para o cálculo do semivariograma.

Nesse estudo, a segmentação foi feita da mesma forma que no Capítulo, 2 por se tratar da mesma região e devido à descrição satisfatória dos limites dos objetos

alcançada, em especial para os objetos de Seringueira, que são encontrados na região em pequenos objetos. Logo, os parâmetros utilizados foram: 120 para o parâmetro de escala, 0,2 para o de forma, e 0,5 para o de compacidade.

3.2.3.2. Amostragem

Foram consideradas quatro classes de cobertura do solo: Seringueira, Eucalipto, Nativa e Outros Usos. A classe Outros Usos é composta por todas as outras classes que não são nem Seringueira, nem Eucalipto e nem Nativa, fazendo com que esta classe tenha uma grande variação de tipos de objetos, como por exemplo, água, área urbana, culturas temporárias e etc.

O conjunto de UAs foi selecionado manualmente por interpretação visual a partir de conhecimento adquirido em campo a respeito da área. Devido à deciduidade apresentada pela Seringueira as UAs foram repassadas e analisadas uma a uma quanto a possibilidade de despertar dúvidas no classificador devido à cobertura foliar (enfolhada e desfolhada). Esse processo atuou como um filtro, onde UAs que não representavam a realidade fisiológica da classe no momento do imageamento (como um polígono de Seringueira, por exemplo, que por algum motivo estava desfolhado quando deveria estar enfolhado, ou vice e versa) foram descartadas.

Assim os polígonos candidatos às UAs foram criteriosamente analisados e as que, por interpretação do analista, pudessem despertar dualidade na interpretação do classificador não foram incluídas na amostra. Foram coletadas 284 UAs, 34 (11,9%) para a classe Seringueira, 72 (25,4%) para a classe Eucalipto, 82 (28,9%) para a classe Nativa e 96 (33,8%) para a classe Outros Usos.

3.2.3.3. Extração de atributos

Os atributos espectrais e espaciais foram extraídos pelo software FETEX 2.0, como proposto por Ruiz *et al.* (2011), que é uma ferramenta para extração de informações e análise de imagens a partir de uma abordagem orientada a objeto, que foi projetado para trabalhar com uso e cobertura do solo auxiliando na classificação dos objetos, portanto a extração dos atributos foi feita por objeto.

As conclusões preliminares alcançadas no Capítulo 2 dessa dissertação foram utilizadas como base para as classificações do presente capítulo. Por esse motivo os atributos espaciais foram extraídos de semivariogramas a partir de valores das seguintes fontes de informação: duas bandas espectrais captadas pelo sensor OLI à bordo do satélite Landsat 8 –Vermelho (R) e NIR (do inglês *Near Infra Red*, que é traduzido como Infravermelho Próximo) - e do índice de vegetação NDVI (do inglês *The Normalized Difference Vegetation Index*, que é traduzido como Índice de Vegetação da Diferença Normalizada). Já os atributos espectrais foram extraídos a partir de valores médios de cada objeto das fontes de informação do R e do NIR.

3.2.3.3.1. Cálculo dos semivariogramas e extração de atributos espaciais por objeto

A distância *lag* é um dos fatores mais importantes a se atentar quando o semivariograma é gerado individualmente para cada objeto. De acordo com Miranda *et al.* (1998) a seleção da distância *lag* depende das classes de estudo devido a grande diversidade no padrão de variabilidade espacial que pode ser encontrada na imagem. Para a geração dos semivariogramas a distância *lag* utilizada foi de um *pixel*, que em uma imagem do sensor OLI/Landsat 8 equivale a 30 metros. O número de *lags* foi fixado em 20 para garantir a geração dos semivariogramas, o que resultou em uma distância total de 600 metros.

A partir dos resultados obtidos no Capítulo 2, no presente capítulo foram utilizados como atributos espaciais apenas os atributos do primeiro e do segundo grupos propostos por Balaguer *et al.* (2010) (Tabela 13). Esses atributos são divididos pelos autores de acordo com a posição dos *lags* utilizados em sua definição em: Próximos a origem e Até o primeiro máximo.

Tabela 13. Atributos espaciais descritos em Balaguer *et al.* (2010) divididas em: (A) Atributos extraídos próximas à origem, (B) Atributos extraídos de valores até o primeiro máximo e (C) Atributos extraídos de valores entre o primeiro e segundo máximos

Grupo	Descrição	Atributos
A	Razão entre a variância total e a semivariância na primeira distância	$RVF = \frac{\text{variância}}{\gamma_i}$
A	Razão entre os valores de semivariância na primeira e na segunda distância	$RSF = \frac{\gamma_1}{\gamma_2}$
A	Derivada próxima da origem	$FDO = \frac{\gamma_2 - \gamma_1}{h}$
A	Segunda derivada do semivariograma na terceira distância	$SDT = \frac{\gamma_4 - 2\gamma_3 + \gamma_2}{h^2}$
B	Primeiro valor máximo de distância	$FML = h_{\max_1}$
B	Média dos valores de semivariância até o primeiro máximo	$MFM = \frac{1}{\max_1} \sum \gamma_i$
B	Variância dos valores de semivariância	$VFM = \frac{1}{\max_1} (\gamma_i - \gamma)^2$
B	Razão entre a semivariância no primeiro máximo e a média do semivariograma até esse máximo	$RMM = \frac{\gamma_{\max_1}}{\gamma_{\max_1}^{\text{média}}}$
B	Diferença entre a média dos valores do semivariograma até esse máximo	$DMF = MFM - \gamma_i$
B	Diferença de segunda ordem entre o primeiro intervalo e o primeiro máximo	$SDF = \gamma_{\max_1} - 2\gamma \frac{\max_1}{2} + \gamma_1$
B	Área entre o valor do semivariograma na primeira distância e a função do semivariograma até o primeiro máximo	$AMF = \frac{h}{2} \left(\gamma_1 + 2 \left(\sum_{i=2}^{\max_1-1} \gamma_i \right) + \gamma_{\max_1} \right) - (\gamma_1 (h_{\max_1} - h_1))$

Fonte: Adaptado Balaguer *et al.* (2010).

3.2.3.3.2. Extração de atributos espectrais

Tratando dos atributos espectrais, os mesmos foram mantidos constantes em todas as classificações, variando apenas a utilização dos atributos espaciais. Os atributos

espectrais utilizados foram extraídos das fontes de informação R e NIR, das duas datas (enfolhada e desfolhada), totalizando quatro fontes de informação disponíveis para a extração dos atributos espectrais, sendo duas para cada data de análise. De cada fonte de informação foi extraída um atributo espectral por objeto, representado pela média dos valores de reflectância dos *pixels*.

2.2.4.1. Contribuição dos atributos

É essencial que seja feito um estudo dos atributos e suas contribuições à separabilidade das classes, conforme observado em Balaguer *et al.* (2010), Wu *et al.* (2015) e Silveira *et al.* (2017). No presente estudo utilizou-se também o *ranking* de atributos mais importantes de cada uma das classificações fornecido pelo próprio algoritmo de classificação, o *Random Forest* (RF), que é baseado na diminuição média das impurezas.

Em uma árvore de decisão cada nó contém um teste para algum atributo, cada ramo descendente corresponde a um possível valor deste atributo. O conjunto de ramos é distinto e cada folha está associada a uma classe, sendo cada percurso da árvore, da raiz à folha, correspondente a uma regra de classificação (GAMA, 2004). O critério utilizado para realizar partições é o da utilização do atributo para a classificação. Aplica-se por este critério um determinado ganho de informação a cada atributo, e o atributo selecionado como teste para o corrente nó é aquele que possui maior ganho de informação, onde a partir deste inicia-se um novo processo de partição (SILVA, 2005).

Segundo Onoda (2001), quando a árvore é utilizada para classificação os critérios de partição mais conhecidos são baseados na entropia e índice de Gini, mas o RF utiliza a entropia, que segundo Mitchell (1997) é o cálculo do ganho de informação baseado em uma medida utilizada na teoria da informação. A entropia caracteriza a impureza dos dados e em um conjunto de dados é uma medida da falta da homogeneidade dos dados de entrada em relação a sua classificação.

3.2.4. Classificação

Com os objetos gerados e as características espaciais e espectrais extraídas, foi iniciado o processo de classificação. O classificador foi treinado utilizando como

entrada os atributos selecionados para cada teste e as UAs de treinamento. Foi utilizado como classificador o RF, conforme em Devries *et al.* (2016), que foi implementado no *software* de fonte aberta Weka 3.8.

Nesse *software*, segundo Belgiu e Dragu (2016), para a geração das árvores é necessário a configuração de dois parâmetros: o número de árvores de decisão que serão geradas (*Ntree*) e o número de variáveis à serem selecionadas e testadas para a divisão em cada nó, quando as árvores estão sendo geradas (*Mtry*). O *Mtry* foi deixado no valor padrão, que é o *log* do número de parâmetros + 1, conforme utilizado em Silveira *et al.* (2017), e o *Ntree* foi de 500 para cada teste por que segundo Lawrence (2006) a maioria dos estudos utiliza 500 árvores por que os erros tendem a estabilizar após esse número.

Foram geradas 9 classificações com o objetivo de estudar o ganho nas AGs das classificações ao utilizar o primeiro ou o segundo grupo dos atributos espaciais, e também de identificar a existência ou não de correlação negativa entre os atributos dos dois grupos. Nas avaliações foram utilizados o *ranking* de importância dos atributos fornecido pelo Weka e as acurácias.

Todas as classificações foram feitas utilizando todos os objetos gerados na área de estudo, e não apenas os objetos contidos na amostra, como aconteceu no Capítulo 2. A seguir estão listadas as classificações produzidas nesse artigo, e na Tabela 14 essas informações estão resumidas.

Classificação 1: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais do primeiro e do segundo grupo, extraídos da fonte de informação R. Portanto essa classificação contou com 15 atributos, 4 espectrais e 11 espaciais.

Classificação 2: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do primeiro grupo, extraídos da fonte de informação R. Portanto essa classificação contou com 8 atributos, 4 espectrais e 4 espaciais.

Classificação 3: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do

segundo grupo, extraídos da fonte de informação R. Portanto essa classificação contou com 11 atributos, 4 espectrais e 7 espaciais.

Classificação 4: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais do primeiro e do segundo grupo, extraídos da fonte de informação NIR. Portanto essa classificação contou com 15 atributos, 4 espectrais e 11 espaciais.

Classificação 5: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do primeiro grupo, extraídos da fonte de informação NIR. Portanto essa classificação contou com 8 atributos, 4 espectrais e 4 espaciais.

Classificação 6: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do segundo grupo, extraídos da fonte de informação NIR. Portanto essa classificação contou com 11 atributos, 4 espectrais e 7 espaciais.

Classificação 7: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais do primeiro e do segundo grupo, extraídos da fonte de informação NDVI. Portanto essa classificação contou com 15 atributos, 4 espectrais e 11 espaciais.

Classificação 8: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do primeiro grupo, extraídos da fonte de informação NDVI. Portanto essa classificação contou com 8 atributos, 4 espectrais e 4 espaciais.

Classificação 9: Classificação espectral-espacial. Os atributos espectrais foram originados de valores extraídos das fontes de informação R e NIR das duas datas (utilizando a componente temporal). Foram utilizados os atributos espaciais apenas do segundo grupo, extraídos da fonte de informação NDVI. Portanto essa classificação contou com 11 atributos, 4 espectrais e 7 espaciais.

Tabela 14. Resumo dos atributos utilizados nas classificações realizadas e o número de atributos contidos em cada classificação (S= Sim, foi utilizada a fonte de informação marcada na classificação).

Classificação	Atributos								Atributos
	Espectrais				Espaciais				
	R	NIR	R - 1º grupo	R - 2º grupo	NIR - 1º grupo	NIR - 2º grupo	NDVI - 1º grupo	NDVI - 2º grupo	
1	S	S	S	S	-	-	-	-	15
2	S	S	S	-	-	-	-	-	8
3	S	S	-	S	-	-	-	-	11
4	S	S	-	-	S	S	-	-	15
5	S	S	-	-	S	-	-	-	8
6	S	S	-	-	-	S	-	-	11
7	S	S	-	-	-	-	S	S	15
8	S	S	-	-	-	-	S	-	8
9	S	S	-	-	-	-	-	S	11

3.2.5. Avaliação da classificação

A avaliação da classificação foi feita utilizando as acurácias das classificações, conforme segue:

- (1) Acurácia do Produtor: Indica a probabilidade de um polígono de referência, ou seja, de uma UA ser classificada corretamente, ou seja, porcentagem de polígonos que foram classificados na classe 'x' e que eram efetivamente pertencentes à classe 'x' no campo. Relaciona-se ao classificador, sendo uma medida da classificação. O inverso da AP é o Erro de Omissão (EO), ou seja, porcentagem de polígonos que foram classificados na classe 'x', mas que não pertenciam à classe 'x' no campo.
- (2) Acurácia do Usuário (AU): Indica a probabilidade de um polígono classificado na imagem realmente pertencer a essa classe na verdade de campo. É a porcentagem de polígonos que pertencem à classe 'x', de acordo com a verdade de campo, e que foram classificados como tal. Relaciona-se a amostragem, sendo uma medida da amostra fornecida pelo analista. O inverso da AU é o Erro de Comissão (EC) ou inclusão, que é a porcentagem de polígonos que tinha-se certeza que pertenciam à classe 'x', mas foram classificados como outra classe.

- (3) Acurácia Global (AG): Indica a precisão geral de uma classificação, que é calculada dividindo o número total de UAs classificadas corretamente pelo número total de UAs.

3.3. Resultados e discussão

3.3.1. Classificações

Todas as classificações geradas utilizaram atributos espectrais e espaciais, os atributos espectrais foram mantidos constantes e os atributos espaciais variaram. Como os melhores resultados no Capítulo 2 foram obtidos com a utilização dos atributos espectrais extraídos das fontes de informação NIR e NDVI, essas fontes foram utilizadas para extração dos atributos espectrais. Os atributos espaciais foram extraídos das fontes R, NIR e NDVI.

Diferente do Capítulo 2, o objetivo é entender como cada um dos grupos de atributos propostos por Balaguer *et al.* (2010) contribuem para a acurácia das classificações. É importante lembrar que o terceiro grupo, designado pelos autores como “Atributos extraídos de valores entre o primeiro e segundo máximos”, não foi utilizado na presente dissertação. O resumo dos atributos utilizados em cada classificação se encontra na Tabela 14.

Na subseção do Capítulo 2 dedicada às classificações espectrais-espaciais (2.1.4.4) as fontes de informação para extração de atributos espaciais foram utilizadas duas a duas, por que foi identificado na seção dedicada aos testes puramente espaciais (2.1.4.3) que quando os atributos espaciais foram utilizados isoladamente não se obtiveram acurácias aceitáveis. Mas, para podermos tratar dos grupos de atributos espaciais propostos por Balaguer *et al.* (2010) é necessário que sejam simuladas classificações espectrais-espaciais utilizando as fontes de informação para extração de atributos espaciais isoladamente. Mesmo que a fonte de informação NIR tenha demonstrado resultados inferiores às fontes de informação R e NDVI, conforme identificado na subseção 2.1.4.4, ela foi estudada no presente capítulo.

Portanto, nesta subseção serão discutidas nove classificações, três de cada uma das fontes de informação, conforme elucidado na subseção 3.2.4. Os resultados se encontram nas Tabelas 15, 16 e 17, e para desenvolver melhor o raciocínio vamos tratar

das classificações três a três, utilizando como critério a fonte de informação utilizada. As três tabelas seguem a mesma lógica, a primeira classificação de cada tabela (1, 4 e 7) é referente a utilização dos atributos dos dois grupos (11 atributos) da fonte de informação em questão. A segunda classificação de cada tabela (2, 5 e 8) é referente a utilização dos quatro atributos do primeiro grupo (Atributos próximos a origem) da fonte de informação em questão. E a terceira classificação de cada tabela (3, 6 e 9) é referente a utilização dos sete atributos do segundo grupo (Atributos até o primeiro máximo) da fonte de informação em questão.

3.3.1.1. Atributos espaciais extraídos da fonte de informação R

Os resultados obtidos para as três classificações que utilizaram a fonte de informação R para extração de atributos espaciais se encontram resumidos na Tabela 15. A Classificação 1 utilizou 15 atributos no total, sendo quatro espectrais e 11 espaciais. A Classificação 2 utilizou 8 atributos no total, sendo quatro espectrais e quatro espaciais. E a Classificação 3 utilizou 11 atributos no total, sendo quatro espectrais e sete espaciais.

Tabela 15. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação R.

Classificações		1	2	3
Nº de atributos utilizados		15	8	11
Polígonos classificados	Correto	88	87	90
	Incorreto	9	10	7
Seringueira	AP	73,3	73,3	78,6
	AU	91,7	91,7	91,7
Nativa	AP	96,3	96,2	96,4
	AU	89,7	86,2	93,1
Eucalipto	AP	92	88,5	92,3
	AU	85,2	85,2	88,9
Outros Usos	AP	93,3	93,3	96,6
	AU	96,6	96,6	96,6
Acurácia Global (AG)		90,7	89,7	92,8

Por meio dos resultados obtidos para a fonte de informação R, percebe-se que a maior AG foi encontrada quando foram utilizados apenas os sete atributos espaciais do

segundo grupo (Classificação 3), no valor de 92,8%. A AG intermediária apresentou o valor 90,7% e foi obtida quando foram utilizados os 11 atributos espaciais (Classificação 1). Já a AG mais baixa foi obtida quando foram utilizados os quatro atributos espaciais do primeiro grupo (Classificação 2) e apresentou o valor de 89,7%.

Esses resultados indicam que houve interação negativa entre os atributos do primeiro e segundo grupos, que refletiu no valor da AG. Esse resultado é sustentado pelo aumento evidenciado na AG da Classificação 3, que foi obtida com a retirada dos atributos do primeiro grupo da análise. Outra forte evidência é o fato da Classificação 2, que foi obtida quando foram utilizados apenas os atributos espaciais do primeiro grupo, ter apresentado a menor AG das três classificações.

Os resultados obtidos para a variação dos atributos extraídos da fonte de informação R estão de acordo com a conclusão obtida no Capítulo 2 de que não é a quantidade de atributos que importa e sim a qualidade. Se a quantidade de atributos fosse mais importante, a maior AG seria da Classificação 1 que utilizou 15 atributos, e não da Classificação 3 que utilizou 11.

Percebe-se também que as diferenças entre as acurácias das três classificações não foram grandes, com as classificações com maior e menor AG apresentando diferença de apenas 3,1, e também considerando que o pior resultado em questão não é um valor baixo, sendo aceito como suficiente, que é o caso da AG de 89,7% para a identificação de quatro classes em estudo com particular dificuldade.

Nas Tabelas 16, 17 e 18 se encontram os *rankings* de importância dos atributos para cada uma das três classificações, seguido da discussão dos resultados.

Tabela 16. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 1 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (2)	0.63	383
2°	NDVI (1)	0.61	400
3°	NIR (2)	0.58	339
4°	NIR (1)	0.52	387
5°	RVF (1° grupo)	0.47	142
6°	FDO (1° grupo)	0.46	97

CONTINUA...

			...CONTINUAÇÃO
7°	SDT (1° grupo)	0.40	92
8°	MFM (2° grupo)	0.40	92
9°	RSF (1° grupo)	0.39	145
10°	FML (2° grupo)	0.39	84
11°	RMM (2° grupo)	0.38	94
12°	DMF (2° grupo)	0.37	63
13°	VFM (2° grupo)	0.33	82
14°	SDF (2° grupo)	0.30	100
15°	AFM (2° grupo)	0.26	63

Através do *ranking* de importância dos atributos percebe-se que os quatro primeiros atributos, ou seja, os atributos mais importantes são os atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. O único atributo espacial que se encontra entre os cinco atributos mais importantes é o RVF, que pertence ao primeiro grupo de atributos. De modo geral os atributos espaciais do primeiro grupo estão melhores colocados no *ranking* do que os do segundo grupo, sendo que apenas o atributo MFM do segundo grupo se encontra entre os atributos do primeiro grupo.

Mesmo com os atributos do primeiro grupo ocupando os primeiros lugares do *ranking* da Classificação 1, foi observado que a maior AG foi obtida ao serem utilizados apenas os atributos do segundo grupo (Classificação 3).

Tabela 17. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 2 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.61	334
2°	NDVI (2)	0.55	303
3°	NIR (1)	0.54	418
4°	NIR (2)	0.53	328
5°	RVF	0.53	99
6°	SDT	0.37	71
7°	FDO	0.31	87
8°	RSF	0.29	117

Através do *ranking* de importância dos atributos percebe-se que, assim como na Classificação 1, os quatro atributos mais importantes são os atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. De forma semelhante a Classificação 1 o único atributo espacial do primeiro grupo que se encontra entre os cinco atributos mais importantes é o RVF.

Tabela 18. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 2 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.62	365
2°	NDVI (2)	0.58	366
3°	NIR (1)	0.55	400
4°	NIR (2)	0.54	348
5°	DMF	0.45	57
6°	MFM	0.42	112
7°	FML	0.39	135
8°	VFM	0.38	94
9°	RMM	0.36	89
10°	AFM	0.31	68
11°	SDF	0.30	82

Através do *ranking* de importância dos atributos percebe-se que, assim como nas Classificações 1 e 2, os quatro atributos mais importantes são os atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. O único atributo espacial do segundo grupo que se encontra entre os cinco atributos mais importantes é o DMF.

3.3.1.2. Atributos espaciais extraídos da fonte de informação NIR

Os resultados obtidos para as três classificações que utilizaram a fonte de informação NIR para extração de atributos espaciais se encontram resumidos na Tabela 19. A Classificação 4 utilizou 15 atributos no total, sendo quatro espectrais e 11 espaciais. A Classificação 5 utilizou 8 atributos no total, sendo quatro espectrais e

quatro espaciais. E a Classificação 6 utilizou 11 atributos no total, sendo quatro espectrais e sete espaciais.

Tabela 19. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação NIR.

Classificações		4	5	6
Nº de atributos utilizados		15	8	11
Polígonos classificados	Correto	90	90	88
	Incorreto	7	7	9
Seringueira	AP	73,3	73,3	73,3
	AU	91,7	91,7	91,7
Nativa	AP	96,6	96,6	96,3
	AU	96,6	96,6	89,7
Eucalipto	AP	95,8	95,8	92
	AU	85,2	85,2	85,2
Outros Usos	AP	96,6	96,6	93,3
	AU	96,6	96,6	96,6
Acurácia Global (AG)		92,8	92,8	90,7

Por meio dos resultados obtidos para a fonte de informação NIR, percebe-se que a maior AG foi encontrada quando foram utilizados os 11 atributos (Classificação 4) ou apenas os quatro atributos espaciais do primeiro grupo (Classificação 5), no valor de 92,8%. Já a AG mais baixa foi obtida quando foram utilizados os sete atributos espaciais do segundo grupo (Classificação 6) e apresentou o valor de 90,7%.

Esses resultados indicam que não houve interação negativa entre os atributos do primeiro e segundo grupos. Esse resultado se sustenta pelo fato de que quando foram utilizados todos os atributos (Classificação 4) a classificação apresentou o mesmo valor de quando foram utilizados apenas os quatro atributos do primeiro grupo de análise (Classificação 5). Isso significa que se houvesse interação negativa entre os atributos dos dois grupos distintos o valor da AG da classificação que utilizou os 11 atributos (Classificação 4) deveria ser diferente da AG das classificações que utilizaram apenas os atributos do primeiro (Classificação 5) e do segundo grupos (Classificação 6). O que pode ter ocorrido é que apenas os atributos do segundo grupo não foram capazes de descrever as classes tão bem quanto os atributos do primeiro grupo.

Os resultados obtidos para a variação dos atributos extraídos da fonte de informação NIR também estão de acordo com a conclusão obtida no Capítulo 2 e com

os resultados obtidos na subseção anterior de que não é a quantidade de atributos que importa e sim a qualidade. A Classificação 5 utilizou apenas 8 atributos e obteve o mesmo resultado da Classificação 4, que utilizou 15. Já a Classificação 6 utilizou 11 atributos e apresentou a AG mais baixa dentre as três classificações que utilizaram a fonte de informação NIR para extrair atributos espaciais.

Uma observação interessante com relação aos resultados das classificações que utilizaram a fonte de informação NIR para extrair atributos espaciais e com a questão da quantidade de atributos é que as Classificações 4 e 5, que apresentaram a mesma AG, apresentaram também APs e AUs idênticas para as quatro classes de estudo, mesmo utilizando quantidade diferente de atributos. Isso reforça a superioridade do fator qualidade de descrição dos atributos com relação ao fator quantidade de atributos utilizados.

As diferenças entre as acurácias das três classificações foi menor ainda que a identificada na subseção anterior, com as classificações com maior e menor AG apresentando diferença de apenas 2,1, com o pior resultado em questão apresentando valor adequado, que é o caso da AG de 90,7% para a identificação de quatro classes em estudo com particular dificuldade.

Nas Tabelas 20, 21 e 22 se encontram os *rankings* de importância dos atributos para cada uma das três classificações, seguido da discussão dos resultados.

Tabela 20. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 4 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.67	382
2°	NDVI (2)	0.61	403
3°	NIR (2)	0.58	342
4°	NIR (1)	0.56	366
5°	RVF (1° grupo)	0.48	153
6°	SDT (1° grupo)	0.43	98
7°	FML (2° grupo)	0.43	65
8°	RSF (1° grupo)	0.43	125
9°	FDO (1° grupo)	0.37	108
10°	RMM (2° grupo)	0.37	115

CONTINUA...

...CONTINUAÇÃO

11°	SDF (2° grupo)	0.36	80
12°	VFM (2° grupo)	0.35	88
13°	MFM (2° grupo)	0.35	97
14°	DMF (2° grupo)	0.35	62
15°	AFM (2° grupo)	0.33	66

Através do *ranking* de importância dos atributos percebe-se que o padrão das três classificações da subseção anterior foi mantido, com os quatro primeiros atributos sendo ocupados por atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. Assim como na Classificação 1 (que utilizou os 11 atributos) o único atributo espacial que se encontra entre os cinco atributos mais importantes é o RVF, que pertence ao primeiro grupo de atributos. De modo geral os atributos espaciais do primeiro grupo estão melhores colocados no *ranking* do que os do segundo grupo, sendo que apenas o atributo FML do segundo grupo se encontra entre os atributos do primeiro grupo.

Os atributos do primeiro grupo ocuparam os primeiros lugares do *ranking* da Classificação 4, e ao contrário do identificado na subseção anterior, foi observado que a maior AG foi obtida ao serem utilizados os primeiros atributos do *ranking*, que no caso foram os atributos do primeiro grupo (Classificação 5).

Tabela 21. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 5 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.6	336
2°	NIR (2)	0.54	320
3°	NDVI (2)	0.52	337
4°	NIR (1)	0.52	435
5°	RSF	0.44	69
6°	RVF	0.39	137
7°	FDO	0.29	90
8°	SDT	0.24	73

Através do *ranking* de importância dos atributos percebe-se que os quatro atributos mais importantes são os atributos espectrais, mas ao contrário do ocorrido em todas as classificações analisadas até então neste capítulo, os atributos se alternaram entre extraídos da fonte de informação NDVI e NIR. Diferente do identificado nas Classificações 1, 2 e 4, o único atributo espacial do primeiro grupo que se encontra entre os cinco atributos mais importantes é o RSF.

Tabela 22. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 6 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.6	360
2°	NDVI (2)	0.57	353
3°	NIR (2)	0.55	324
4°	NIR (1)	0.54	438
5°	FML	0.53	91
6°	VFM	0.42	80
7°	MFM	0.37	114
8°	RMM	0.37	129
9°	DMF	0.34	62
10°	SDF	0.3	90
11°	AFM	0.29	65

Através do *ranking* de importância dos atributos percebe-se que, assim como nas Classificações 1, 2, 3 e 4, os quatro atributos mais importantes são os atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. O único atributo espacial do segundo grupo que se encontra entre os cinco atributos mais importantes é o FML.

3.3.1.3. Atributos espaciais extraídos da fonte de informação NDVI

Os resultados obtidos para as três classificações que utilizaram a fonte de informação NDVI para extração de atributos espaciais se encontram resumidos na Tabela 23. A Classificação 7 utilizou 15 atributos no total, sendo quatro espectrais e 11 espaciais. A Classificação 8 utilizou 8 atributos no total, sendo quatro espectrais e

quatro espaciais. E a Classificação 9 utilizou 11 atributos no total, sendo quatro espectrais e sete espaciais.

Tabela 23. Resultados das Acurácias do Produtor (AP), do Usuário (AU) e Global, das classificações que utilizaram a fonte de informação NDVI.

Classificações		7	8	9
Nº de atributos utilizados		15	8	11
Polígonos classificados	Correto	86	88	87
	Incorreto	11	9	10
Seringueira	AP	73,3	73,3	73,3
	AU	91,7	91,7	91,7
Nativa	AP	92,6	96,3	96,2
	AU	86,2	89,7	86,2
Eucalipto	AP	88	88,5	88,5
	AU	81,5	85,2	85,2
Outros Usos	AP	93,3	96,6	93,3
	AU	96,6	96,6	96,6
Acurácia Global (AG)		88,7	90,7	89,7

Por meio dos resultados obtidos para a fonte de informação NDVI, percebe-se que a maior AG foi encontrada quando foram utilizados apenas os quatro atributos espaciais do primeiro grupo (Classificação 8), no valor de 90,7%. A AG intermediária apresentou o valor 89,7% e foi obtida quando foram utilizados os sete atributos espaciais do segundo grupo (Classificação 9). Já a AG mais baixa foi obtida quando foram utilizados os 11 atributos (Classificação 7) e apresentou o valor de 88,7%.

Esses resultados indicam que houve interação negativa entre os atributos do primeiro e segundo grupos, que refletiu no valor da AG. Esse resultado é sustentado pelo aumento evidenciado na AG da Classificação 8, que foi obtida com a retirada dos atributos do segundo grupo da análise. Outra forte evidência é o fato da Classificação 7, que foi obtida quando foram utilizados os 11 atributos espaciais, ter apresentado a menor AG das três classificações que utilizaram a fonte NDVI para extração de atributos espaciais, que pode indicar os atributos do segundo grupo tenham atrapalhado a Classificação 7.

Os resultados obtidos para a variação dos atributos extraídos da fonte de informação NDVI também estão de acordo com a conclusão obtida no Capítulo 2 e com os resultados obtidos nas duas subseções anteriores de que não é a quantidade de

atributos que importa e sim a qualidade. A Classificação 8 utilizou apenas 8 atributos e obteve resultado superior às Classificações 7 e 9, que utilizaram concomitantemente 15 e 11 atributos, e apresentaram AGs mais baixas.

As diferenças entre as acurácias das três classificações foi bem pequena, com as classificações com maior e menor AG apresentando diferença de apenas 2. O pior resultado em questão apresentou AG com valor satisfatório para a identificação de quatro classes em estudo com particular dificuldade, no valor de 90,7%.

Nas Tabelas 24, 25 e 26 se encontram os *rankings* de importância dos atributos para cada uma das três classificações, seguido da discussão dos resultados.

Tabela 24. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 7 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.62	372
2°	NDVI (2)	0.61	373
3°	NIR (1)	0.56	389
4°	NIR (2)	0.53	343
5°	RVF (1° grupo)	0.47	155
6°	AFM (2° grupo)	0.43	42
7°	FDO (1° grupo)	0.42	120
8°	RSF (1° grupo)	0.42	119
9°	SDT (1° grupo)	0.4	80
10°	MFM (2° grupo)	0.39	81
11°	FML (2° grupo)	0.39	81
12°	VFM (2° grupo)	0.36	79
13°	RMM (2° grupo)	0.35	76
14°	DMF (2° grupo)	0.3	70
15°	SDF (2° grupo)	0.27	62

Através do *ranking* de importância dos atributos percebe-se que o padrão identificado nas duas subseções anteriores foi mantido, com os quatro primeiros atributos sendo ocupados por atributos espectrais (com exceção da Classificação 5), com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. Assim como nas Classificações 1, 2 e 4, o único atributo espacial que se encontra entre os cinco atributos mais importantes é o RVF, que

pertence ao primeiro grupo de atributos. De modo geral os atributos espaciais do primeiro grupo estão melhores colocados no *ranking* do que os do segundo grupo, sendo que apenas o atributo AFM do segundo grupo se encontra entre os atributos do primeiro grupo.

Os atributos do primeiro grupo ocuparam os primeiros lugares do *ranking* da Classificação 7, e assim como identificado na subseção anterior, foi observado que a maior AG foi obtida ao serem utilizados os primeiros atributos do *ranking*, que no caso foram os atributos do primeiro grupo (Classificação 7).

Tabela 25. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 8 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.61	324
2°	NIR (2)	0.54	339
3°	NDVI (2)	0.53	317
4°	NIR (1)	0.52	429
5°	RVF	0.46	103
6°	FDO	0.34	89
7°	RSF	0.31	97
8°	SDT	0.28	73

Através do *ranking* de importância dos atributos percebe-se que os quatro atributos mais importantes são os atributos espectrais, mas assim como o ocorrido na Classificação 5, os atributos se alternaram entre extraídos da fonte de informação NDVI e NIR. Mas diferente da Classificação 5 e similar ao identificado nas Classificações 1, 2, 4 e 7 o único atributo espacial do primeiro grupo que se encontra entre os cinco atributos mais importantes é o RVF.

Tabela 26. *Ranking* da importância dos atributos, redução média das impurezas causada por cada atributo e número de nós em que foram utilizados para a Classificação 9 (Os atributos espectrais vêm acompanhados do número um ou dois entre parênteses, que indicam se o atributo foi extraído da primeira ou da segunda data de análise).

Posição	Atributo	Redução média das impurezas	Número de nós que utilizaram o atributo
1°	NDVI (1)	0.6	396
2°	NDVI (2)	0.57	346
3°	NIR (2)	0.55	341
4°	NIR (1)	0.53	409
5°	FML	0.43	96
6°	MFM	0.42	104
7°	VFM	0.39	107
8°	DMF	0.38	71
9°	RMM	0.38	91
10°	AFM	0.34	59
11°	SDF	0.26	77

Através do *ranking* de importância dos atributos percebe-se que, assim como nas Classificações 1, 2, 3, 4, 6, 7 e 8 os quatro atributos mais importantes são os atributos espectrais, com os dois atributos extraídos da fonte do NDVI nos dois primeiros lugares, seguidos dos atributos extraídos da fonte NIR. O único atributo espacial do segundo grupo que se encontra entre os cinco atributos mais importantes é o FML, assim como identificado na Classificação 6, que também utilizou os atributos do segundo grupo, mas utilizou a fonte de informação NIR para extração desses atributos.

Devido aos resultados obtidos nas Classificações 5 e 8 pode-se sugerir que os atributos espaciais mais interessantes, do ponto de vista de descrição das classes, são os atributos do primeiro grupo de análise, por que tanto para a fonte de informação NIR quanto para a fonte NDVI, os melhores resultados foram obtidos quando foram utilizados os atributos espaciais do segundo grupo.

Ao serem utilizados os 11 atributos espaciais, ou seja, entre as Classificações 1, 4 e 7, a maior AG foi encontrada na Classificação 4, que utilizou a fonte de informação NIR para extrair os atributos, com o valor de 92,8%, seguida da Classificação 1, que utilizou a fonte de informação R para extrair os atributos, com o valor de 90,7%, e por último a Classificação 7, que utilizou a fonte de informação NDVI para extrair os atributos, com o valor de 90,7%.

Ao serem utilizados os quatro atributos espaciais do primeiro grupo, ou seja, entre as Classificações 2, 5 e 8, a maior AG foi encontrada na Classificação 5, que também utilizou a fonte de informação NIR para extrair os atributos, com o valor de 92,8%, seguida da Classificação 8, que utilizou a fonte de informação NDVI para extrair os atributos, com o valor de 90,7%, e por último a Classificação 2, que utilizou a fonte de informação R para extrair os atributos, com o valor de 89,7%.

Ao serem utilizados os sete atributos espaciais do segundo grupo, ou seja, entre as Classificações 3, 6 e 9, a maior AG foi encontrada na Classificação 6, que também utilizou a fonte de informação NIR para extrair os atributos, com o valor de 90,7%, seguida das Classificações 3 e 9, que utilizaram as fontes de informação R e NDVI, respectivamente, para extrair os atributos, com o valor de 89,7%.

3.3.1.4. Melhores classificações

Nesta subseção ilustramos e discutimos as classificações que apresentaram os melhores resultados do presente capítulo. As classificações 3, 4 e 5 obtiveram AG de 92,8%, mas as classificações 4 e 5 apresentaram também mesmas APs e AUs. Então foram gerados mapas e tabelas de área das classificações 3 e 5, que se encontram respectivamente nas Figuras 17 e 18.

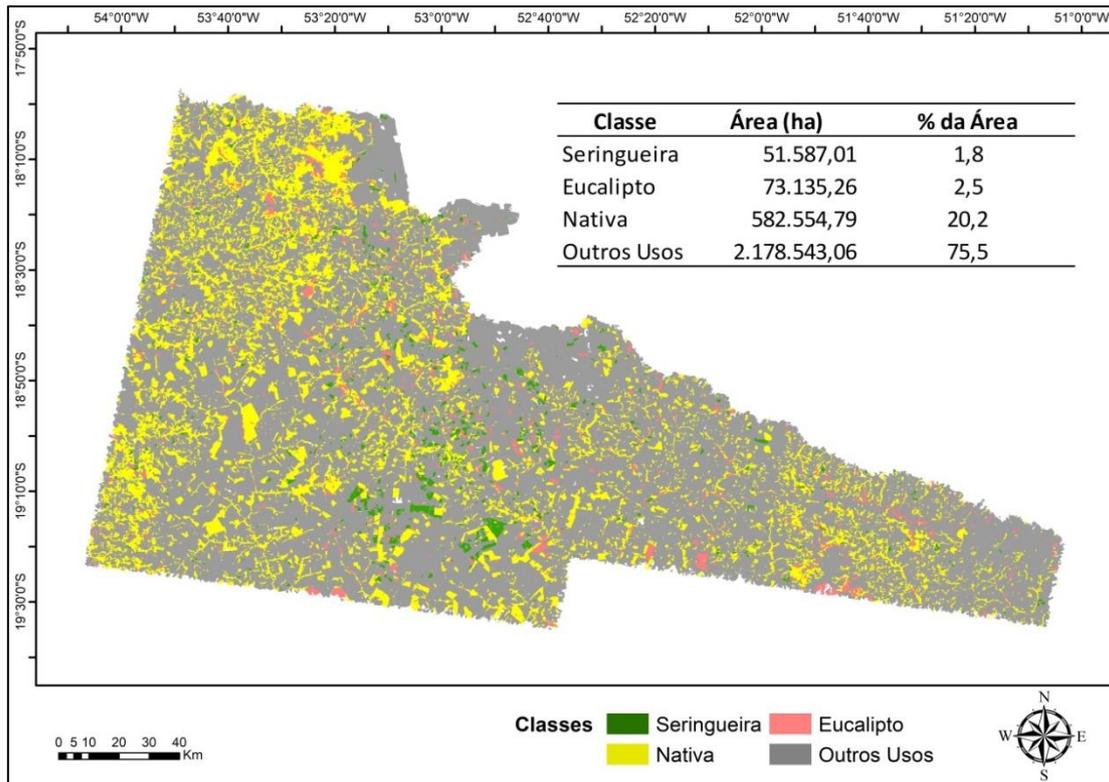


Figura 17. Classificação 3 e tabela com a área ocupada por cada uma das classes, nesta classificação.

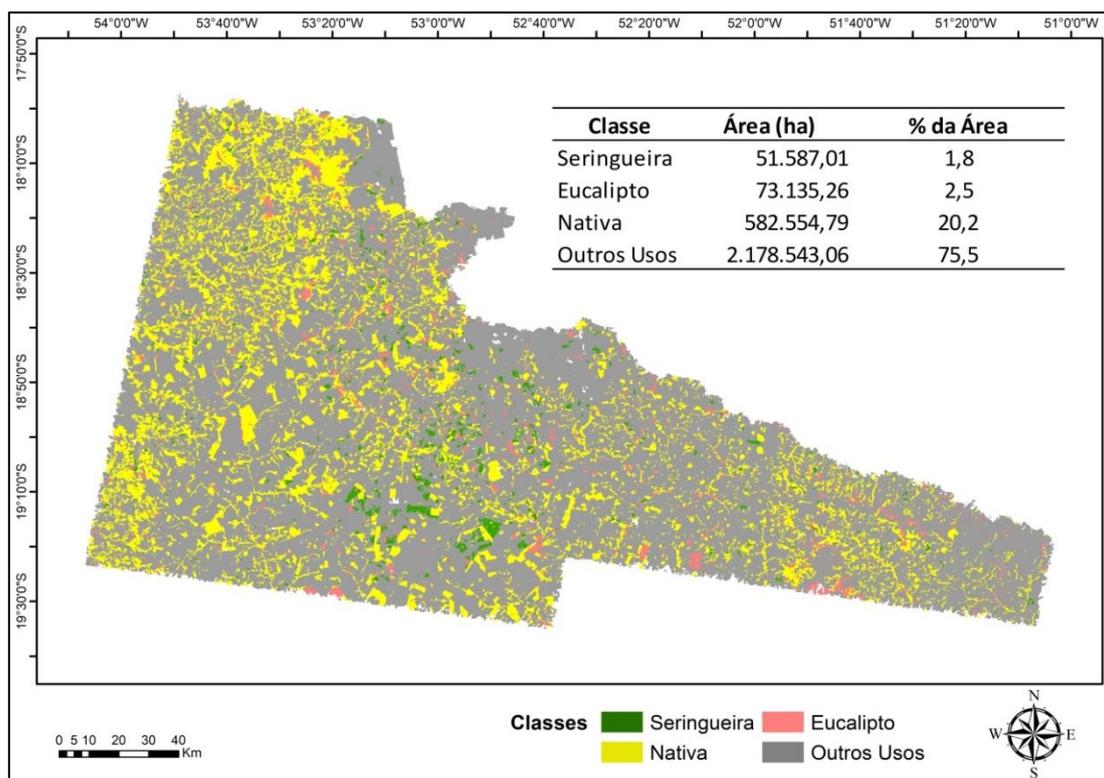


Figura 18. Classificação 5 e tabela com a área ocupada por cada uma das classes, nesta classificação.

O que mais chama a atenção nestas duas classificações são as tabelas de área. Em ambas as classificações foram obtidas os mesmos valores de área para as quatro classes de estudo, mesmo utilizando diferentes atributos espaciais e fontes de informações para extração desses atributos. A Classificação 3 utilizou os atributos do segundo grupo extraídos da fonte de informação R e a Classificação 5 utilizou os atributos do primeiro grupo extraídos da fonte de informação NIR.

Isso pode indicar que não há necessariamente uma melhor fonte de informação para a extração de atributos espaciais, mas sim que cada uma das fontes de informação é representada da melhor forma pelo primeiro ou segundo grupo de atributos. Isso acontece por que cada fonte de informação tem um comportamento diferente e característico, e os grupos propostos por Balaguer *et al.* (2010) foram criados exatamente por utilizarem valores de diferentes partes das curvas dos semivariogramas. Em outras palavras, cada fonte de informação apresenta sua peculiaridade, e portanto a contribuição à classificação, em uma determinada porção do semivariograma, e por esse motivo são melhores ou piores representadas pelos atributos espaciais de determinado grupo. Por exemplo, se a fonte de informação NIR for melhor descrita pela parte do semivariograma mais próxima a origem do gráfico, ou seja, pela primeira parte do mesmo, provavelmente os atributos espaciais que vão agregar mais ao processo de classificação são os atributos gerados a partir dos valores da primeira parte do gráfico, que pertencem ao grupo de 'Atributos extraídos próximos a origem'.

Pode-se iniciar o debate também a respeito da capacidade de descrição das fontes de informação e dos grupos dos atributos espaciais. A partir dos resultados obtidos na presente subseção podemos começar a especular a respeito da equivalência de descrição das fontes de informação, por que como a AG das classificações 3 e 5 foram iguais, mas as APs e AUs não, a conclusão mais sensata seria de que houve apenas coincidência entre as AGs. Mas, ao analisarmos as tabelas de áreas das duas classificações essa coincidência não se mostra tão óbvia por que, pelo que parece, os polígonos foram classificados nas mesmas classes em ambas as classificações, mesmo utilizando diferentes fontes de informação e grupos de atributos espaciais. Isso pode indicar que os atributos espaciais do segundo grupo extraídos da fonte de informação R possuem a mesma capacidade descritiva dos atributos espaciais do primeiro grupo extraídos da fonte de informação NIR. Mas, embora esse seja um resultado interessante

e intrigante na mesma medida, necessita ser explorado mais a fundo para que se possa entender essa interação e equivalência, o que foge do escopo proposto pela presente dissertação.

3.4. Conclusões

Atributos espectrais são mais importantes para a identificação das classes estudadas do que os atributos espaciais, em classificações espectrais-espaciais.

A fonte de informação NDVI origina atributos espectrais mais importantes, em comparação com a fonte de informação NIR.

O grupo a qual os atributos espaciais utilizados pertencem é mais importante do que a quantidade de atributos espaciais e do que a fonte de informação utilizada.

De maneira geral os atributos do primeiro grupo ocuparam os primeiros lugares do *ranking* das classificações, logo após os atributos espectrais.

Os atributos espaciais mais interessantes a serem empregados em classificações espectrais-espaciais são os atributos do primeiro grupo, por que apresentaram melhores resultados tanto para a fonte de informação NIR quanto para a NDVI.

De maneira geral, o atributo espacial mais importante é o atributo do primeiro grupo RVF.

Houve interação negativa entre os atributos do primeiro e segundo grupos, das classificações geradas utilizando atributos espaciais extraídos das fontes de informação R e NDVI.

As melhores classificações foram obtidas quando foram utilizados os atributos espaciais do segundo grupo extraídas da fonte R, ou quando foram utilizados os atributos espaciais do primeiro grupo da fonte NIR, em conjunto com atributos espectrais extraídos das fontes de informação R e NIR.

REFEREÊNCIAS

- BALAGUER, A.; RUIZ, L. A.; HERMOSILLA, T.; RECIO, J. A. Definition of a comprehensive set of texture semivariograma features and their evaluation for object-oriented image classification. **Computers & Geoscience**, Houston-TX, v. 36, p. 231-240, 2010.
- BELGIU, M.; DRAGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 114, p. 24–31, 2016.
- BLASCHKE, T. Object based image analysis for remote sensing. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 65, p. 2–16, 2010.
- BREIMAN, L. Random forests. **Machine Learning**, Bristol-UK, v. 45, n. 1, p. 5–32, 2001.
- CHEN, B.; CAO, J.; WANG, J.; WU, Z.; TAO, Z.; CHEN, J.; YANG, C.; XIE, G. Estimation of rubber stand age in typhoon and chilling injury afflicted area with Landsat TM data: A case study in Hainan Island, China. **Forest Ecology and Management**, Flagstaff-AZ, v. 274, p. 222–230, 2012a.
- CURRAN, P. The semivariograma in remote sensing: an introduction. **Remote Sensing of Environment**. Toronto-ON, v. 24, p. 493–507, 1988.
- DEVRIES, B.; PRATIHAST, A. K.; VERBESSELT J.; KOOISTRA L.; HEROLD, M. Characterizing forest change using community-based monitoring data and Landsat time series. **Plos One**, College Park-MD, v. 11, n. 3, 2016.
- GAMA, J. **Árvores de Decisão**. Porto: Universidade do Porto, 2004.
- GHOSH, A.; JOSHI, P. K. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. **International Journal of Applied Earth Observation Geoinformation**, Enschede-Overijssel, v. 26, p. 298–311, 2014.
- GUAN, H.; LI, J.; CHAPMAN, M.; DENG, F.; JI, Z.; YANG, X. Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. **International Journal of Geographical Information Science**, London-UK, v. 34, p. 5166–5186, 2013.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção Agrícola Mundial**, 2017. Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=ms>> . Acesso em: 28 jun. 2018.
- INMET – INSTITUTO NACIONAL DE METEOROLOGIA. **Série histórica** – Dados diários, 2018. Disponível em: <<http://www.inmet.gov.br/porta/>>. Acesso em: 01 jul. 2018.

LAWRENCE, R. L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). **Remote Sensing**, v. 100, p. 356–362, 2006.

LEAL, S. T. **A heveicultura na mesorregião leste do Estado de Mato Grosso do Sul: Aspectos técnicos e econômicos**. 2017. 108 p. Tese (Mestrado em Agronomia) - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, Ilha Solteira, 2017.

MIRANDA, F. P., FONSECA, L. E. N., CARR, J. R. Semivariogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest. **IEEE International Journal of Remote Sensing**, London-UK, v. 19, n. 3, p. 549–556, 1998.

MITCHELL, T. M. **Machine Learning**. USA: McGraw-Hill, USA, 1997. 432 p.

MUI, A.; HE, Y.; WENG, Q. An object-based approach to delineate wetlands across landscapes of varied disturbance with high spatial resolution satellite imagery. **International Journal of Geographical Information Science**, London-UK, v. 109, p. 30–46, 2015.

ONODA, M. 2001. 110 p. **Estudo sobre um algoritmo de árvore de decisão acoplado a um sistema de banco de dados relacional**. Dissertação (Mestrado) Universidade Federal de São João Del Rei – UFRJ, São João Del Rei-MG, 2001.

OSHIRO, T. M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. 101 p. Dissertação (Mestrado em Bioinformática) – Universidade de São Paulo – USP, Ribeirão Preto-SP, 2013.

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

RUIZ, L. A.; RECIO, J.; FERNÁNDEZ-SARRÍA, A.; HERMOSILLA, T. A Feature extraction software tool for agricultural object-based image analysis. **Computers and Electronics in Agriculture**, Amsterdã-Holanda do Norte, v. 76, n. 2, p. 284–296, 2011.

SCALOPPI JUNIOR, E. J. FREITAS, R. S, GONÇALVES, P. S. **O Agrônomo - Boletim Técnico-Informativo do Instituto Agrônomo**, Campinas-SP, v. 69 p. 56 – 2017.

SILVA, J. S. V.; SPERANZA, E. A.; VENDRUSCULO, L. G.; ESQUERDO, J. C. D. M.; MAURO, R. A.; BIANCHINI, S. L.; FLORENCE, R. O. **Projeto GeoMS: melhorando o Sistema de Licenciamento Ambiental do Estado do Mato Grosso do Sul**. Campinas/SP: Embrapa Informática Agropecuária, 2011. 64 p.

SILVA, L. M. O. **Uma aplicação de árvores de decisão, redes neurais e KNN para identificação de modelos ARMA não-sazonais e sazonais**. 2005. 138 p. Tese

(Doutorado em Engenharia Elétrica) – Pontífica Universidade Católica – PUC, Rio de Janeiro- RJ, 2005.

SILVEIRA, E. M. O.; MENEZES, M. M.; ACERBI JÚNIOR, F. W.; TERRA, M. C. N. S.; MELLO, J. M. Assessment of geostatistical features for object-based image classification of contrasted landscape vegetation cover. **Journal of Applied Remote Sensing**, Orlando-FL, v. 11, n. 3, 2017.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing I: scene models and simulated images. **Remote Sensing of Environment**. Toronto-ON, v. 25, p. 323–348, 1988a.

WOODCOCK, C. E.; STRAHLER, A. H.; JUPP, D. L. B. The use of variograms in remote sensing II: real digital images. **Remote Sensing of Environment**. Toronto-ON, v. 25, p. 323–34349–3798, 1988b.

WU, X.; PENG, J.; SHAN, J.; CUI, W. Evaluation of semivariogram features for object based image classification. **Geo-spatial Information Science**, Londres-UK, v. 18, n. 4, p. 159-170, 2015.