SERVIÇO PÚBLICO FEDERAL

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS

PRÓ-REITORIA DE ENSINO

CÂMPUS GOIÂNIA

BACHARELADO EM ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA

ANGELA GABRIELLY PIRES SILVA

Integração de imagens ópticas e SAR com processamento em nuvem no mapeamento de uso e cobertura da terra no Cerrado

Goiânia, 2021.

SERVIÇO PÚBLICO FEDERAL

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS

PRÓ-REITORIA DE ENSINO

CÂMPUS GOIÂNIA

BACHARELADO EM ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA

ANGELA GABRIELLY PIRES SILVA

Integração de imagens ópticas e SAR com processamento em nuvem no mapeamento de uso e cobertura da terra no Cerrado

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Engenharia Cartográfica e de Agrimensura como requisito parcial para obtenção do título de Bacharel em Engenharia Cartográfica e de Agrimensura.

Orientador: Prof. Dr. Édipo Henrique Cremon

Goiânia, 2021.

Si381i Silva, Angela Gabrielly Pires.
Integração de imagens ópticas e SAR com processamento em nuvem no mapeamento de uso e cobertura da terra no Cerrado / Angela Gabrielly Pires Silva. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2021.
70 f. : il.

Orientador: Prof. Dr. Édipo Henrique Cremon.

TCC (Trabalho de Conclusão de Curso) – Curso de Bacharelado em Engenharia Cartográfica e de Agrimensura, Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

1. Sensoriamento remoto. 2. Mapeamento de uso. I. Cremon, Édipo Henrique (orientador). II. Instituto Federal de Educação, Ciência e Tecnologia de Goiás. III. Título.

CDD 621.3678

Ficha catalográfica elaborada pela Bibliotecária Karol Almeida da Silva Abreu CRB1/ 2.740
Biblioteca Professor Jorge Félix de Souza,
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

[ ] Tese
[ ] Dissertação
[ ] Monografia – Especialização
[X] TCC - Graduação
[ ] Artigo Científico
[ ] Capítulo de Livro
[ ] Livro
[ ] Trabalho Apresentado em Evento
[ ] Produto Técnico e Educacional - Tipo: _____

Nome Completo do Autor: Angela Gabrielly Pires Silva
Matrícula: 20161011190040
Título do Trabalho: Integração de imagens ópticas e SAR com processamento em nuvem no mapeamento de uso e cobertura da terra no Cerrado.

Autorização - Marque uma das opções

1. ( ) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. (X) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data 01/08/2021(Embargo);
3. ( ) Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

( ) O documento está sujeito a registro de patente.
(x) O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
( ) Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia - GO, 18/03/2021.

Angela Gabrielly Pires Silva

Assinatura do Autor e/ou Detentor dos Direitos Autorais

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS GOIÂNIA

ATA Nº ECA 03 - TCC/2021.1

ATA DA SESSÃO PÚBLICA DE APRESENTAÇÃO E DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

Aos 18 dias do mês de março do ano de dois mil e vinte e um, às 13:00 horas, na sala virtual meet.google.com/bnr-xtig-szm, foi realizada a sessão pública de apresentação e defesa do Trabalho de Conclusão de Curso da Graduanda **Angela Gabrielly Pires Silva (matrícula 20161011190040)** do curso de Bacharelado em Engenharia Cartográfica e de Agrimensura, do Departamento III, Câmpus Goiânia, do Instituto Federal de Goiás, no primeiro semestre do ano de dois mil e vinte e um. A banca foi composta pelos seguintes membros:

- Édipo Henrique Cremon (Professor-Orientador)

- Fabio Corrêa Alves (Avaliador Externo)

- Elaine Reis Costa Lima

- Giovanni de Araújo Boggione

sob a presidência da(o) primeira(o). O Trabalho de Conclusão de Curso tem como título "**Integração de imagens ópticas e SAR com processamento em nuvem para mapeamento de uso e cobertura da terra no Cerrado**", sob orientação do Prof. **Édipo Henrique Cremon**. Após a apresentação do Projeto de Trabalho de Conclusão de Curso, tendo sido a autora arguida pela Banca Examinadora, a nota obtida foi 9,6 (nove virgula seis).

Encerra-se a sessão às 15 horas e 30 minutos. Eu, Profa. Marina Alberti Macedo, dato e assino a presente ata que segue assinada por todos os membros da Banca Examinadora.

Profa. Marina Alberti Macedo (Assinado eletronicamente)

Prof. Édipo Henrique Cremon (Assinado eletronicamente)

Prof. Fabio Corrêa Alves (Assinado eletronicamente)

Profa. Elaine Reis Costa Lima (Assinado eletronicamente)

Prof. Giovanni de Araújo Boggione (Assinado eletronicamente)

Documento assinado eletronicamente por:

- **Marina Alberti Macedo**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 22/03/2021 13:22:17.
- **Elaine Reis Costa Lima**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 19/03/2021 13:05:24.
- **Fabio Corrêa Alves**, **FABIO CORRÊA ALVES - 332105 - PROFESSOR ASSISTENTE DE REGÊNCIA DE CLASSE - IFG - CÂMPUS GOIÂNIA (10870883000225)**, em 19/03/2021 12:14:14.
- **Giovanni de Araujo Boggione**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 19/03/2021 11:40:42.
- **Edipo Henrique Cremon**, **PROFESSOR ENS BASICO TECN TECNOLOGICO**, em 19/03/2021 11:04:04.

Este documento foi emitido pelo SUAP em 04/03/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse https://suap.ifg.edu.br/autenticar-documento/ e forneça os dados abaixo:

Código Verificador: 137263
Código de Autenticação: 4d79bfdde9

AGRADECIMENTOS

Esse trabalho é fruto de esforços e contribuições de muitas pessoas às quais gostaria de deixar os meus agradecimentos:

Agradeço primeiramente a Deus, pelas oportunidades que sempre me forneceu e pelas bênçãos ao longo de minha vida.

Ao meu orientador Dr. Édipo Cremon, pela dedicação, auxílio, orientação, paciência, confiança e amizade, tornando possível a conclusão desse trabalho.

Aos membros da banca, por aceitarem o convite e pelas contribuições para o aperfeiçoamento do trabalho.

Aos meus amigos que contribuíram com suas amizades e apoio em todos os momentos.

Aos meus colegas de faculdade, que dividiram comigo bons e maus momentos, pelo apoio, experiência e ajuda nos momentos difíceis.

A minha família, pelo apoio e incentivo, foram cruciais para a minha vida, ajudaram-me a chegar até aqui, sei que sem vocês nada disso seria possível.

A todos que não estão listados, mas que participaram de forma direta ou indireta na realização desse trabalho.

Eu quero ser tudo que sou capaz de me tornar.

Katherine Mansfield,
(1888-1923).

RESUMO

Título: Integração de imagens ópticas e SAR com processamento em nuvem no mapeamento de uso e cobertura da terra no Cerrado
Autor: Angela Gabrielly Pires Silva
Orientador: Dr. Édipo Henrique Cremon

O mapeamento de uso e cobertura da terra é de suma importância para o monitoramento ambiental e gestão territorial. Séries temporais de radar de abertura sintética (SAR) do Sentinel-1 (S-1) e sensor óptico MSI/Sentinel-2 (S-2) fornecem condições favoráveis para o mapeamento de uso e cobertura da Terra devido suas resoluções espectrais e espaciais. Este trabalho parte do pressuposto que a combinação entre as séries temporais S-1 e S-2 permite maior exatidão no mapeamento de uso e cobertura da terra no bioma Cerrado em relação às classificações com sensores únicos. As imagens, referentes ao período de janeiro/2019 a dezembro/2019, foram classificadas utilizando o algoritmo Random Forest testando diferentes parâmetros na plataforma de processamento em nuvem *Google Earth Engine*. As classificações obtidas apenas com os dados S-2 (Kappa = 89,99) foram melhores do que as obtidas com os dados S-1 sozinhos (Kappa = 75,78). A eficiência da classificação aumentou ao combinar os dados S-1 e S-2 (Kappa= 93,07). Infravermelho de ondas curtas, a polarização VH dos dados SAR e os índices CAI (*Cellulose absorption index*) e Hall Cover foram as variáveis mais importantes na classificação integrada dos dados S-2 e S-1, respectivamente.

Palavras–chave

Google Earth Engine, sensoriamento remoto, sinergia óptica-SAR, Random Forest, aprendizado de máquina

ABSTRACT

Title: Optical and sar imagery integration in land cover mapping based on cloud computing in the Cerrado
Author: Angela Gabrielly Pires Silva
Advisor: Dr. Édipo Henrique Cremon

Land use and land cover mapping is important for environmental monitoring and land management. Synthetic aperture radar (SAR) from Sentinel-1 (S-1) and optical sensor MSI/Sentinel-2 (S-2) time series provide favorable conditions for land use and land cover mapping at orbital level due to their spectral, spatial and temporal resolutions. This work has been based on the hypothesis that the combination between S-1 and S-2 time series allows for greater accuracy in land use and land cover mapping in the Cerrado biome. The images were classified using the Random Forest algorithm by different hyperparameters in the *Google Earth Engine* cloud processing platform. The classifications obtained with the S-2 data alone (Kappa = 89.99) were better than those obtained with the S-1 data alone (Kappa = 75.78). The classification efficiency increased when combining the S-1 and S-2 data (Kappa = 93.07). Shortwave infrared, VH polarization from SAR data, CAI (Cellulose absorption index) and Hall Cover indices were the most important variables in the integrated classification of the S-2 and S-1 data, respectively.

Keywords

Google Earth Engine, remote sensing, optical-SAR synergy, Random Forest, machine learning

LISTA DE FIGURAS

3.1 Constelação dos satélites Sentinel-1A e 1B. 21
3.2 Exemplo dos processos de treinamento e classificação com o Random Forest. 26
3.3 Exemplo de uma matriz de confusão de uma classificação hipotética . . . 28
3.4 Ambiente de desenvolvimento do Google Earth Engine. 31

4.1 Localização da área de estudo. 32
4.2 Fluxograma da pesquisa. 34

5.1 Cena do Sentinel-1 a) sem filtragem e b)com filtragem do ruído speckle. . 40
5.2 Cena do Sentinel-2 a) sem filtragem e b)com filtragem de nuvem com porcentagem $\leq 20\%$. 41
5.3 Classificação com integração de dados ópticos e radar com diferentes números de árvore. 45
5.4 Importância de cada variável usada na classificação do RF, com 100 árvores, pelo índice de Gini. 45
5.5 Importância de cada variável usada na classificação do RF, com 500 árvores, pelo índice de Gini. 46
5.6 Classificação com dados de radar com diferentes números de árvore. . . . 49
5.7 Importância de cada variável usada na classificação do RF, com *ntree* = 100 e *ntree* = 500, pelo Índice de Gini. 50
5.8 Classificação com dados ópticos com *ntree* 100 e 500. 53
5.9 Importância de cada variável usada na classificação do RF, com 100 árvores, pelo índice de Gini. 54
5.10 Importância de cada variável usada na classificaçãodo RF, com 500 árvores, pelo índice de Gini. 54
5.11 Áreas utilizadas para comparação. 56
5.12 Comparação entre os três cenários com *ntree* = 500. 57

LISTA DE TABELAS

5.1 Matriz de Confusão da classificação utilizando a integração de dados ópticos e radar com o *ntree* = 100. 42
5.2 Exatidão do produtor, exatidão do usuário e kappa condicional de cada classe com *ntree* = 100. 43
5.3 Matriz de Confusão da classificação utilizando a integração de dados ópticos e radar com o *ntree* = 500. 43
5.4 Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com *ntree*= 500. 44
5.5 Matriz de Confusão da classificação utilizando dados de radar com *ntree* = 100. 47
5.6 Exatidão do produtor, Exatidão do usuário e Kappa condicional de cada classe com *ntree*= 100. 47
5.7 Matriz de Confusão da classificação utilizando dados de radar com *ntree* = 500. 48
5.8 Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com *ntree*= 500. 48
5.9 Matriz de Confusão da classificação utilizando dados ópticos com *ntree* = 100. 51
5.10 Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com *ntree*= 100. 51
5.11 Matriz de Confusão da classificação utilizando dados ópticos com *ntree* = 500. 52
5.12 Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com *ntree*= 500. 52

LISTA DE QUADROS

3.1 Bandas utilizadas por sistemas radares com os respectivos comprimentos de onda e frequências e a equação que os relaciona. 20
3.2 Características dos modos de aquisição do sensor Sentinel-1. 22
3.3 Aplicações dos modos de aquisição da missão Sentinel-1. 23
3.4 Organização das faixas espectrais do sensor MSI/Sentinel-2 de acordo com a resolução espacial. 24
3.5 Níveis de processamento para as imagens MSI/Sentinel-2. 24
4.1 Índices utilizados no estudo. 36
4.2 Classes utilizadas na classificação e suas respectivas descrições. 36
4.3 Distribuição das amostras entre treinamento e validação por classe. 37
4.4 Código de legenda utilizada pelo MapBiomas (2019). 39

LISTA DE ABREVIATURAS E SIGLAS

API	Interface de Programas e Aplicativos
ERS	*European Remote Sensing*
ESA	Agência Espacial Europeia
EW	*Extra-Wide Swath Mode*
GEE	Google Earth Engine
GRD	*Ground Range Detected*
IE	Importância Baseada no Erro
IG	Importância de Gini
IW	*Interferometric Wide Swath Mode*
oob	*Out-of-Bag*
RAR	*Real Aperture Radar*
RF	*Random Forest*
S-1	Sentinel-1
SAR	*Synthetic Aperture Radar*
SIG	sistemas de informações geográficas
SM	*Stripmap Mode*
SVM	*Support Vector Machine*
WV	*Wave Mode*

LISTA DE SÍMBOLOS

σ	Coeficiente de Retroespalhamento
q	Número de classes
n	Número total de amostras
$n_k k$	Elementos da diagonal principal da matriz de confusão

SUMÁRIO

LISTA DE FIGURAS **9**

LISTA DE TABELAS **10**

LISTA DE QUADROS **11**

LISTA DE ABREVIATURAS E SIGLAS **12**

LISTA DE SÍMBOLOS **13**

1 INTRODUÇÃO **16**

2 OBJETIVOS **18**

3 REVISÃO DE LITERATURA **19**
- 3.1 Mapeamento do Uso e Cobertura da Terra 19
- 3.2 Radares imageadores e a missão Sentinel-1 19
- 3.3 Dados ópticos do MSI/Sentinel-2 23
- 3.4 Classificação de imagens em sensoriamento remoto 24
  - 3.4.1 Random Forest 25
    - 3.4.1.1 Importância de Gini 27
  - 3.4.2 Validação do desempenho da classificação 27
    - 3.4.2.1 Matriz de Confusão 28
    - 3.4.2.2 Exatidão Global 29
    - 3.4.2.3 Exatidão do Produtor 29
    - 3.4.2.4 Exatidão do usuário 29
    - 3.4.2.5 Índice Kappa 30
- 3.5 Google Earth Engine 30

4 METODOLOGIA **32**
- 4.1 Localização e caracterização da área de estudo 32
- 4.2 Materiais 33
  - 4.2.1 Coleta dos dados e Pré-Processamento 33
- 4.3 Procedimentos Metodológicos 33
  - 4.3.1 Processamento dos dados SAR 34
  - 4.3.2 Processamento dos dados ópticos 35
  - 4.3.3 Amostras de treinamento e validação 35
  - 4.3.4 Classificação do uso e cobertura da terra 37
  - 4.3.5 Validação 38
  - 4.3.6 Pós-Classificação 39

5 RESULTADOS 40
5.0.1 Imagens SAR 40
5.0.2 Imagens Ópticas 41
5.1 Classificação do uso e cobertura da terra 41
5.1.1 Classificação utilizando a integração de dados ópticos e radar 42
5.1.2 Classificação utilizando dados de radar 46
5.1.3 Classificação utilizando dados ópticos 50
5.1.4 Comparação dos melhores modelos de cada cenário 55

6 DISCUSSÕES 58

7 CONCLUSÕES 61

8 RECOMENDAÇÕES 62

REFERÊNCIAS BIBLIOGRÁFICAS 63

1 INTRODUÇÃO

O reconhecimento e detalhamento do espaço físico territorial é de suma importância obras de engenharia, arquitetura, conservação de solos, planejamento territorial e regional ou mapeamentos temáticos, entre outros. O mapeamento do uso e cobertura da terra, por exemplo, pode ser realizado com base em métodos de classificação de imagens(RIBEIRO; CENTENO, 2001).

A utilização do sensoriamento remoto na obtenção de dados relacionados ao uso e cobertura da terra para monitoramento e análises dos recursos naturais tem sido amplamente difundida. Nos dias atuais, os sensores disponíveis possuem diferentes resoluções espaciais, espectrais, radiométricas e temporais, possibilitando maiores níveis de informação a serem extraídos dos dados (NUNES; ROIG, 2015). Uma das aplicações pertinentes para o levantamento de informações territoriais é o uso de imagens orbitais para o mapeamento do uso e cobertura da terra, uma vez que, imagens tomadas destas plataformas conseguem mapear vastas áreas com qualidade e periodicidade a um baixo custo operacional visto que existem imagens que são disponibilizadas gratuitamente.

O monitoramento do uso e cobertura da terra baseado apenas em imagens ópticas por um longo período apresentam insuficiência deinformações desejadas para tal análise, principalmente em períodos chuvosos ou em épocas com presença de queimadas ou ainda com comportamento espectral óptico semelhante, pois nessas condições o mapeamento é comprometido pela impossibilidade de interpretação do produto por causa da confusão de alvos. Uma alternativa são os sensores ativos, como os dados de radar, que têm como principal vantagem a sua independência da radiação solar. Além disso, os sensores ativos possuem certos comprimentos de onda capazes de atravessar as nuvens, e têm a capacidade de imageamento quase independente das condições climáticas, além de interagir com a geometria e constante dielétrica dos alvos na faixa de comprimento de onda das micro-ondas.

O mapeamento, a caracterização e a análise da dinâmica do uso e da cobertura da terra no Cerrado, por meio de um monitoramento preciso e sistemático, utilizando sensoriamento remoto e sistemas de informações geográficas (SIG), são fatores cruciais para lidar com o desafio de ordenar a ocupação de um território de mais de dois milhões de km^2, equivalentes a 23,92% do território nacional (IBGE, 2004).

No entanto, a maioria dos estudos utilizando o sensoriamento remoto no Cerrado, utilizam apenas imagens ópticas dependendo apenas de sensores passivos que tem seus dados prejudicados em situações como nuvens, fumaça e poeira, entre outros. Para

contornar a situação, o recomendado seria utilizar como complemento, dados de radar imageadores, que por serem sensores ativos na faixa espectral das micro-ondas não são tão prejudicados por esses tipos de situações. Adicionalmente, os alvos interagem na faixa espectral das micro-ondas possibilitando uma maior separabilidade de classes que sejam espectralmente semelhantes no espectro óptico.

O uso de imagens combinadas entre dados ópticos e *Synthetic Aperture Radar* (SAR) para mapeamentos de uso e cobertura da terra pode apresentar avanços na discriminação de alvos da superfície da Terra (PEREIRA; S.; S., 2016; SYMEONAKIS *et al.*, 2018; ZHANG *et al.*, 2019; LOPES *et al.*, 2020). Especificamente para o bioma Cerrado, é comum a confusão entre classes de uso e cobertura devido à similaridade espectral entre algumas classes temáticas, como vegetação natural de formação gramíneo-lenhosas e pastagens (SANO; JESUS; BEZERRA, 2001). A avaliação da integração de dados SAR e ópticos, com os satélites Sentinel 1 e 2, por exemplo, para melhorar a detecção de classes de uso e cobertura da terra por meio de classificadores ainda é um campo pouco pesquisado e pode auxiliar na reprodutibilidade metodológica de projetos de mapeamento de biomas como o que foi executado pelo projeto TerraClass Cerrado (BRASIL, 2015) e MapBiomas (SOUZA *et al.*, 2020).

O armazenamento e processamento de dados de sensoriamento remoto em disco local, seja óptico e/ou de radar, demandam infraestrutura computacional. Plataformas baseadas na nuvem, como o Google Earth Engine (GEE), permitem amenizar e contornar limitações de infraestrutura computacional. As capacidades existentes dentro do GEE, tais como uma biblioteca consolidada de imagens de satélite gratuitas, possibilitam o monitoramento contínuo da superfície terrestre por diferentes resoluções espectrais, espaciais e temporais, inclusive para a integração entre dados ópticos e de radar com armazenamento e processamento em nuvem (GHORBANIAN *et al.*, 2020).

2 OBJETIVOS

O objetivo deste trabalho foi avaliar a integração de imagens do Sentinel 1 e 2 no processo de classificação supervisionada de imagens para uso e cobertura da terra em uma região contida no bioma Cerrado dentro da plataforma do Google Earth Engine.

Os objetivos específicos são:

- Avaliar se a integração de dados SAR e ópticos melhoram a classificação de imagens de sensoriamento remoto no mapeamento de uso e cobertura da terra no bioma Cerrado;
- Verificar quais as bandas, índices espectrais e polarizações, entre dados ópticos e SAR, mais eficientes e adequados na classificação de uso e cobertura da terra por classe;
- Analisar quais métricas estatísticas temporais (mínimo, mediana, variância e desvio padrão) das bandas, polarizações e índices espectrais de dados ópticos e SAR são mais eficientes na classificação de uso e cobertura da terra no Cerrado;
- Inferir o custo computacional necessário para o armazenamento e processamento local comparado com processamento em nuvem pelo Google Earth Engine.

3 REVISÃO DE LITERATURA

Neste capítulo é feita uma breve revisão sobre o mapeamento de uso e cobertura da terra, dados SAR e ópticos dos satélites do Sentinel-1 e Sentinel-2, respectivamente. Também é abordado a classificação supervisionada, o algoritmo de classificação *Random Forest* e validação da classificação. Por fim, é apresentada a plataforma do Google Earth Engine.

3.1 Mapeamento do Uso e Cobertura da Terra

Mudanças no uso da terra estão entre os principais impulsionadores das mudanças ambientais globais e o sensoriamento remoto é a ferramenta mais importante para monitorar os efeitos do uso da terra nos ecossistemas (SONNENSCHEIN *et al.*, 2011).

Segundo Lambin e Geist (2008), o mapeamento da cobertura terrestre, está relacionado às propriedades físicas de uma superfície, enquanto o uso da terra está relacionado às atividades desenvolvidas pelos seres humanos naquele ambiente. O mapeamento é o processo pelo qual se geram mapas temáticos, através das classes de cobertura de terra.

Vegetação, superfícies construídas, e corpos d´água, são alguns exemplos de cobertura da terra, enquanto turismo, uso residencial e conservação, são exemplos de usos da terra. Segundo Oliveira (2015): "A mudança no uso da terra é a mudança dos usos que o homem faz sobre o solo mantendo uma floresta, ou derrubando-a para transformá-la em um campo de cultivo ou pastagem; tornando áreas rurais em urbanas; inundando vales com a construção de barragens; criando unidades de conservação."

Esse monitoramento de uso e cobertura da terra podem ser feitos por meio do sensoriamento remoto orbital utilizando sensores ópticos ou sensores de radar, com imagens de diferentes resoluções e integração de múltiplos sensores.

3.2 Radares imageadores e a missão Sentinel-1

Os radares são sistemas ativos que operam na região das micro-ondas. Radar é um acrônimo para radio detection and ranging, que significa detecção e medição de distâncias por ondas de rádio. Segundo Ponzoni, Shimabukuro e Kuplich (2015), os radares podem ser divididos em duas classes: *Real Aperture Radar* (RAR) conhecido

como radar de abertura real e *Synthetic Aperture Radar* (SAR) que é o radar de abertura sintética.

De acordo com Ponzoni, Shimabukuro e Kuplich (2015), os radares são classificados de acordo com seus parâmetros, que são, de modo geral, o comprimento de onda/frequência, tipo de polarização e ângulo de incidência. A maioria dos radares operam apenas com uma uma banda, embora existam exceções de radares que operam em até três bandas, essa limitação existe devido ao fato do sistema gerar a sua própria energia. O Quadro 3.1 apresenta as bandas utilizadas pelo sistema SAR e seus respectivos comprimentos de onda e frequências, cada banda permite uma penetração distinta nos objetos que aumenta com o comprimento de onda (PONZONI; SHIMABUKURO; KUPLICH, 2015).

Quadro 3.1: *Bandas utilizadas por sistemas radares com os respectivos comprimentos de onda e frequências e a equação que os relaciona.*

Banda de radar	**Comprimento de onda – (cm)**	**Frequência – f (MHz)**
P	136-77	220-390
UHF	100-30	300-1.000
L	30-15	1.000-2.000
S	15-7,5	2.000-4.000
C	7,5-3,75	4.000-8.000
X	3,75-2,40	8.000-12.500
Ku	2,40-1,67	12.500-18.000
K	1,67-1,18	18.000-26.500
Ka	1,18-0,75	26.500-40.000
$\lambda(\text{cm}) = \frac{c}{f} = \frac{30.000}{f(MHz)} = \frac{30}{f(GHz)}$	1 hertz = 1 ciclo s^{-1} 1 mega-hertz = 10^6 hertz 1 giga-hertz = 10^9 hertz	

Fonte: Ponzoni, Shimabukuro e Kuplich (2015).

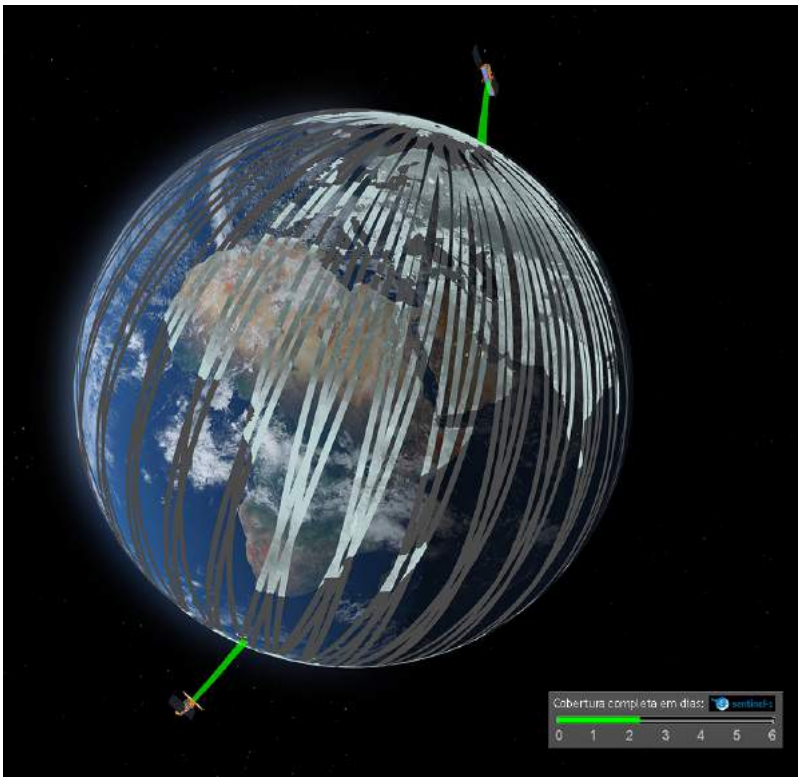
A missão Sentinel-1 (S-1) é baseada em uma constelação de satélites SAR operando com a banda C idênticos, que atualmente são compostos pelas unidades A e B, para fornecer continuidade de dados para a Agência Espacial Europeia (ESA) e teve como precursores as missões *European Remote Sensing* (ERS) e ENVISAT SAR. As imagens

de radar tem sido utilizadas para diversos estudos: mapeamento de uso e cobertura do solo (CARNEIRO *et al.*, 2020; DINIZ, 2019), monitoramento da cobertura de neve (RONDEAU-GENESSE; TRUDEL; LECONTE, 2016), áreas de inundação (GRAOSQUE, 2018), entre outros.

A missão S-1 fornece uma capacidade operacional independente para mapeamento de radar contínuo da Terra e foi projetada para fornecer frequência, cobertura e confiabilidade de revisita aprimoradas para serviços operacionais e aplicações que requerem longas séries temporais. A missão também foi concebida para fornecer uma capacidade de interferometria (InSAR) operacional por meio de requisitos rigorosos colocados em precisão de atitude, conhecimento de atitude e órbita e precisão de tempo de coleta de dados (ESA, 2020).

A constelação S-1 cobre todas as massas do globo a cada duas semanas, zonas de gelo marinho, zonas costeiras, rotas de navegação diariamente e oceano aberto continuamente por imagens SAR (ESA, 2020). A missão S-1 está em uma órbita quase polar, heliossíncrona, com um ciclo de repetição de 12 dias e 175 órbitas por ciclo para um único satélite. Ambos os satélites, Sentinel-1A e Sentinel-1B, compartilham o mesmo plano orbital com uma diferença de fase orbital de 180 °. Com os dois satélites operando, o ciclo de repetição é de 6 dias (Figura 3.1).

Figura 3.1: *Constelação dos satélites Sentinel-1A e 1B.*



Fonte: Adaptado de ESA (2020).

Os satélites da missão Sentinel-1 têm incorporado um instrumento SAR que opera na banda C a uma frequência de 5.405 GHz. O sistema tem a capacidade de operar

nos modos de dupla polarização (HH/HV ou VV/VH) e polarização única (HH ou VV) e possui quatro modos de aquisição distintos: *Stripmap Mode* (SM), *Interferometric Wide Swath Mode* (IW), *Extra-Wide Swath Mode* (EW) e *Wave Mode* (WV) (SARAIVA, 2015; ESA, 2020). No Quadro 3.2 pode-se verificar um resumo das características dos quatro modos de aquisição da missão Sentinel-1.

Quadro 3.2: *Características dos modos de aquisição do sensor Sentinel-1.*

Modo	Ângulo	Resolução Espacial (m)	Largura da faixa (km)	Polarização
SM	20° - 45°	5x5	80	HH/HV, VV/VH. HH, VV
IW	29° - 46°	5x20	250	HH/HV, VV/VH. HH, VV
EW	19° - 47°	20x40	400	HH/HV, VV/VH. HH, VV
WV	22° - 35°	5x5	20x20	HH, VV
	35° - 38°			

Fonte: Saraiva (2015).

O modo SM pode operar com uma das seis elevações de feixe pré-definidas, cada uma com um ângulo incidente distinto. Este modo destina-se apenas a casos excepcionais, como apoio a ações de gestão de emergência (ESA, 2020)

No Quadro 3.3 está representado as aplicações de cada modo de aquisição.

Quadro 3.3: *Aplicações dos modos de aquisição da missão Sentinel-1.*

Aplicação	Modo			
	SW	**IW**	**EW**	**Wv**
Ártico e gelo marítimo		X	X	
Vigilância de navios em mar aberto		X	X	
Monitoramento de poluição de óleo		X	X	
Ventos marítimos				X
Florestas	X	X		
Agricultura	X	X		
Cartografia de deformação urbana	X	X		
Monitoramento de inundações		X		
Análise de sismos		X		
Monitoramento de deslizamentos de terras e vulcões		X		

Fonte: Adaptado de Saraiva (2015).

3.3 Dados ópticos do MSI/Sentinel-2

Sentinel-2 é uma missão européia de imageamento multiespectral óptico que atualmente é composta por dois satélites com sensores idênticos. A especificação completa da missão dos satélites gêmeos voando na mesma órbita, mas em fase de 180°, foi projetada para fornecer uma alta frequência de revisita de 5 dias no Equador (SENTINEL, 2015). O sensor MultiSpectral Instrument (MSI), abordo do Sentinel-2, fornece 13 bandas espectrais (Quadro 3.4) para aplicações no monitoramento agrícola, detecção de desastres, estudo de material suspenso na atmosfera etc.

Quadro 3.4: *Organização das faixas espectrais do sensor MSI/Sentinel-2 de acordo com a resolução espacial.*

Número da banda	Nome da banda	Comprimento de onda central (nm)	Largura de banda (nm)	Resolução espacial (metros)
1	Aerossol costeiro	443	20	60
2	Azul	490	65	10
3	Verde	560	35	10
4	Vermelho	665	30	10
5	Borda Vermelha da Vegetação	705	15	20
6	Borda Vermelha da Vegetação	740	15	20
7	Borda Vermelha da Vegetação	783	20	20
8	NIR	842	115	10
8b	Narrow NIR	865	20	20
9	Vapor d'água	945	20	60
10	SWIR - Cirrus	1380	30	60
11	SWIR	1610	90	20
12	SWIR	2190	180	20

Fonte: Adaptado de Drusch *et al.* (2012).

Os dados adquiridos, a cobertura da missão e a alta frequência de revisitas, possibilitam a geração de geoinformação em escalas diversas. Os dados dos satélites Sentinel-2 são complementares às missões existentes em questão de informação, incluindo Landsat e SPOT. As imagens são disponibilizadas em dois níveis de correção para os usuários (Quadro 3.5).

Quadro 3.5: *Níveis de processamento para as imagens MSI/Sentinel-2.*

Nome	Descrição de cada nível	Produção e Distribuição	Volume de Dados
Nível-1C	Dados georreferenciados fornecidos em refletância aparente (topo da atmosfera)	Geração sistemática e distribuição on-line	600 MB (cada 100 x 100 km2)
Nível 2A	Dados georreferenciados fornecidos em refletância de superfície	Geração sistemática e distribuição on-line e geração local (usando a caixa de ferramentas Sentinel-2 no software SNAP)	800 MB (cada 100 x 100 km2)

3.4 Classificação de imagens em sensoriamento remoto

Classificação de imagens é o processo de extração de informações em imagens a qual associa cada pixel da imagem a um "rótulo" descrevendo assim um objeto real (p.ex. solo, água, vegetação, etc.) (CROSTA, 1999). Os valores de número digital ou reflectância de cada pixel são agrupados em classes com valores semelhantes, este agrupamento envolve a análise dos dados matriciais e a aplicação de regras de decisão baseadas em estatísticas e algoritmos (PASSO, 2013).

A classificação de imagens subdivide-se em supervisionada e não supervisionada. Na classificação não-supervisionada um algoritmo de agrupamento (*clustering*) decide quais as classes serão separadas e agrupadas em conjuntos homogêneos de pixels. Já a classificação de imagens supervisionada requer amostras de treinamento para cada classe, ou seja, uma coleção de pontos de dados que representam as classes existentes na área (REDDY; REDDY, 2008).

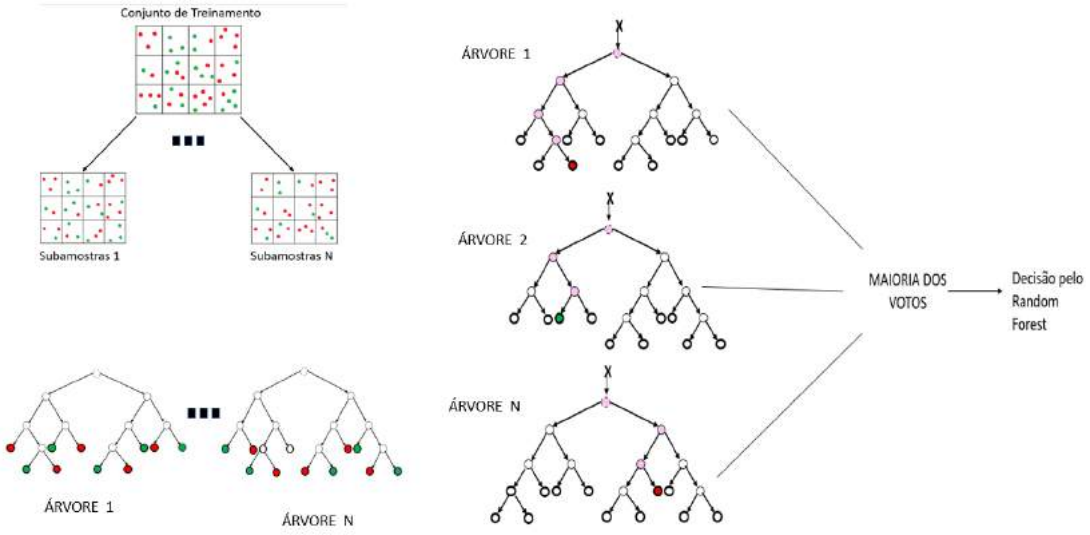
O treinamento consiste no reconhecimento da comportamento espectral associada aos objetos reais. Para alguns classificadores esse reconhecimento envolve parâmetros estatísticos (média, matriz de covariância, etc.), enquanto outros baseiam-se apenas no nível mínimo e máximo de cinza do pixel.

O processo de mapeamento do uso e cobertura, os principais procedimentos incluem: a seleção das classes a representar, a coleta de amostras de treinamento, a escolha e utilização do algoritmo classificador supervisionado, a estimativa de parâmetros do classificador a ser utilizado através das amostras, produção de mapas temáticos ou tabelas de síntese dos resultados obtidos e a avaliação da exatidão do produto final utilizando amostras de validação (SARAIVA, 2015).

3.4.1 Random Forest

O Classificador *Random Forest* (RF) é uma combinação de árvores de decisão preditoras. Essas árvores de decisão são criadas de forma independentes através de um subconjunto de amostras de treinamento (BREIMAN, 2001; DINIZ, 2019).Esse algoritmo cria um conjunto de árvores de decisão que, a partir de vetores aleatórios amostrados, de forma independente, são igualmente distribuídos em todas as árvores de decisão. Ele tem como resposta a classe que recebeu o maior número de votos dentre todas as árvores consideradas pelo algoritmo (HAN; KAMBER; PEI, 2011). Na Figura 3.2, está um exemplo de treinamento e os processos de classificação utilizando o RF. Cada árvore de decisão no conjunto é construída sobre uma subamostra aleatória dos dados originais, que contém a classe verde e vermelha. A predição de classe para as novas instâncias usando o RF é baseada em um processo de votação por maioria entre todas as árvores individuais. O procedimento realizado por cada árvore é: para cada novo dado (ou seja, X), o algoritmo começa no nó raiz de uma árvore de decisão e percorre a árvore pelos ramos testando os valores das variáveis em cada um dos nós (representado pela cor rosa), de acordo com cada nó, ele seleciona o ramo seguinte. Esse procedimento é repetido até que um nó folha (último nó do ramo) seja alcançado, o que atribui uma classe a esta instância (verde ou vermelho). No final do processo, cada árvore tem um voto para o rótulo da classe, e o voto majoritário da floresta é escolhido como a previsão final.

Figura 3.2: *Exemplo dos processos de treinamento e classificação com o Random Forest.*



Fonte: Adaptado de Machado, Mendoza e Corbellini (2015).

De acordo com Rodriguez-Galiano *et al.* (2012), o classificador denominado Random Forest apresenta muitas vantagens para aplicações em sensoriamento remoto:

- É executado de forma eficiente em bancos de dados extensos;
- Pode lidar com milhares de variáveis de entrada sem exclusão de variável;
- Fornece estimativas de quais variáveis são importantes na classificação;
- Gera uma estimativa imparcial interna do erro de generalização (erro *Out-of-bag*);
- Calcula proximidades entre pares de casos que podem ser usados na localização de outliers;
- É relativamente robusto para outliers e ruído;
- É computacionalmente mais leve do que outros métodos de conjunto de árvores.

Há dois hiperparâmetros de entrada para produzir as árvores de decisão: o número de árvores a serem criadas (*ntree*) e o número de variáveis que serão testadas (*mtry*)(DINIZ, 2019). Cada subconjunto selecionado usando *bagging* para fazer cada *ntree* crescer geralmente contém 2/3 do conjunto de dados de calibração. As amostras que não estão presentes no subconjunto de calibração são incluídas como parte de outro subconjunto denominado *Out-of-Bag* (oob) (RODRIGUEZ-GALIANO *et al.*, 2012).

A proporção entre as classificações erradas e o número total de elementos oob contribui para uma estimativa não enviesada do erro de generalização (BREIMAN, 2001). O erro de generalização converge conforme o número de árvores aumenta, portanto, o RF não se ajusta aos dados.

Além disso, o RF permite calcular a importância das variáveis utilizando duas medidas: Importância Baseada no Erro (IE) e Importância de Gini (IG), a primeira mede o aumento quando se permutam os valores do atributo de interesse enquanto a segunda baseada na soma dos decréscimos do índice de Gini em todos os nós rotulados pelo atributo (BASTOS; NASCIMENTO; LAURETTO, 2014).

O algoritmo RF utiliza o índice de Gini para avaliar a quantidade relativa que cada variável contribuiu para a diferença entre as classes na classificação. Valores mais altos de índice de Gini representam maior importância na exatidão da classificação final e uma maior contribuição para maximizar a diferença entre as classes.

3.4.1.1 Importância de Gini

A Importância de Gini ou Impureza de Gini, desenvolvida por Conrado Gini em 1912, diz que se escolhermos ao acaso uma amostra arbitrária da base de dados e essa amostra sempre for atribuída a mesma classe, isso significa que a amostra de dados é dita como pura, resultando em um coeficiente de Gini igual a um (ONODA, 2001) . Caso contrário, a amostra de dados é dita como impura, resultando em um coeficiente de Gini igual a zero.

Este índice num determinado nó é dado por:

$$IG = 1 - \sum_{k=1}^{n} p_k^2 \tag{3-1}$$

onde:

$p_k{}^2$ i é a freqüência relativa de cada classe em cada nó

n é p número de classes

3.4.2 Validação do desempenho da classificação

Uma etapa vital no processo de classificação supervisionada é a avaliação da exatidão do resultado das imagens classificadas. Isso envolve a identificação de um conjunto de amostras tidas como verdadeiras, muitas vezes que são visitadas no campo ou fotointerpretadas por especialistas. A amostra de validação é então comparada a classe que foi mapeada na imagem para o mesmo local. Avaliações estatísticas de exatidão podem então serem derivadas para toda a área de estudo, bem como para as classes individuais (RICHARDS; JIA, 2006).

Muitos métodos de avaliação de imagens foram discutidos na literatura de sensoriamento remoto (ARONOFF *et al.*, 1982; ARONOFF, 1985; JR; MILLONES, 2011; OLOFSSON *et al.*, 2014). A maioria é calculado com base em uma matriz confusão ou matriz de erro (FOODY, 2002).

3.4.2.1 Matriz de Confusão

De acordo com Congalton e Green (1999), a matriz de contingência é considerado um modo eficaz para representar a exatidão do mapeamento. A leitura da matriz deve ser realizada da seguinte forma: na diagonal principal estará expressa a quantidade de pixels que foram classificados corretamente, enquanto os elementos fora da diagonal expressarão a quantidade de pixels que foram incluídos erroneamente em uma classe (erros de inclusão), bem como, a quantidade de pixels que foram excluídos erroneamente de uma classe (erros de omissão).

A matriz de confusão é expressa por meio da correlação dos dados tidos como verdade e os dados classificados (preditos) (Figura 3.3). A matriz de confusão estabelece a base para descrever a qualidade da classificação e descrever seus erros e acertos, fator que contribui no refinamento da classificação e/ou as estimativas derivadas dela (FOODY, 2002).

Figura 3.3: *Exemplo de uma matriz de confusão de uma classificação hipotética*

Classificação \ Referência	A	B	C	D	Σ
A	n_{AA}	n_{AB}	n_{AC}	n_{AD}	n_{A+}
B	n_{BA}	n_{BB}	n_{BC}	n_{BD}	n_{B+}
C	n_{CA}	n_{CB}	n_{CC}	n_{CD}	n_{C+}
D	n_{DA}	n_{DB}	n_{DC}	n_{DD}	n_{D+}
Σ	n_{+A}	n_{+B}	n_{+C}	n_{+D}	n

Fonte: Adaptado de Foody (2002).

Os elementos na 3.3 representam a diagonal principal da matriz que contém os casos em que os rótulos de classe descritos na classificação da imagem e no conjunto de dados do solo estão os acertos, enquanto os elementos fora da diagonal contêm os casos em que há discordância nos erros. Muitas métricas de exatidão de classificação podem ser derivadas de uma matriz de confusão, como por exemplo a exatidão global e o índice Kappa, além de métricas por classe como as exatidões do produtor e usuário, além do Kappa condicional.

3.4.2.2 Exatidão Global

A taxa de acerto ou Exatidão Global (Po), proposta por Helldén (1980) (Equação 3-2), é calculada dividindo-se a soma dos valores da diagonal principal, que representa o número de amostras classificadas corretamente n_{++}, pelo número total de amostras coletadas n, ou seja:

$$Po = \frac{\sum_{k=1}^{q} n_{kk}}{n} * 100 \tag{3-2}$$

onde:

q = número de classes;
n_{kk}= elementos da diagonal principal; e,
n = número total de amostras.

3.4.2.3 Exatidão do Produtor

Segundo Tung e LeDrew (1988) a exatidão do produtor é o número de amostras classificadas corretamente de uma categoria particular dividido pelo número total de amostras de referência para essa categoria (Equação 3-3). É uma medida do erro de omissão (STORY; CONGALTON, 1986).

$$A_p = \frac{n_{kk}}{n_{k+}} \tag{3-3}$$

onde:

n_{kk}= elementos da diagonal principal; e ,
n_{k+} = total de elementos classificados para uma categoria k.

3.4.2.4 Exatidão do usuário

Esta medida, também chamada de confiabilidade, é um indicativo da probabilidade de que um pixel classificado no mapa / imagem realmente representar essa categoria

no terreno (Equação 3-4) (STORY; CONGALTON, 1986).

$$A_u = \frac{n_{kk}}{n_{+k}} \tag{3-4}$$

onde:

n_{kk}= elementos da diagonal principal; e ,

n_{k+} = total de elementos de referência amostrados para uma categoria k.

3.4.2.5 Índice Kappa

Cohen (1960) define Kappa (K) como um coeficiente de concordância entre dados da classificação e verdade de campo para escalas nominais. De forma similar a exatidão global, o índice kappa também leva elementos não diagonais em consideração (BANKO, 1998).

O kappa pode ser determinado pela equação 3-5.

$$K = \frac{\sum_{k=1}^{q} n_{kk} - \sum_{k=1}^{q} n_{k+} n_{+k}}{n^2 - \sum_{k=1}^{q} n_{k+} n_{+k}} \tag{3-5}$$

onde:

q = número de classes;

n_{k+} = total de elementos classificados para uma categoria k;

n_{+k} = total de elementos de referência amostrados para uma categoria k;

n_{kk}= elementos da diagonal principal; e,

n = número total de amostras.

3.5 Google Earth Engine

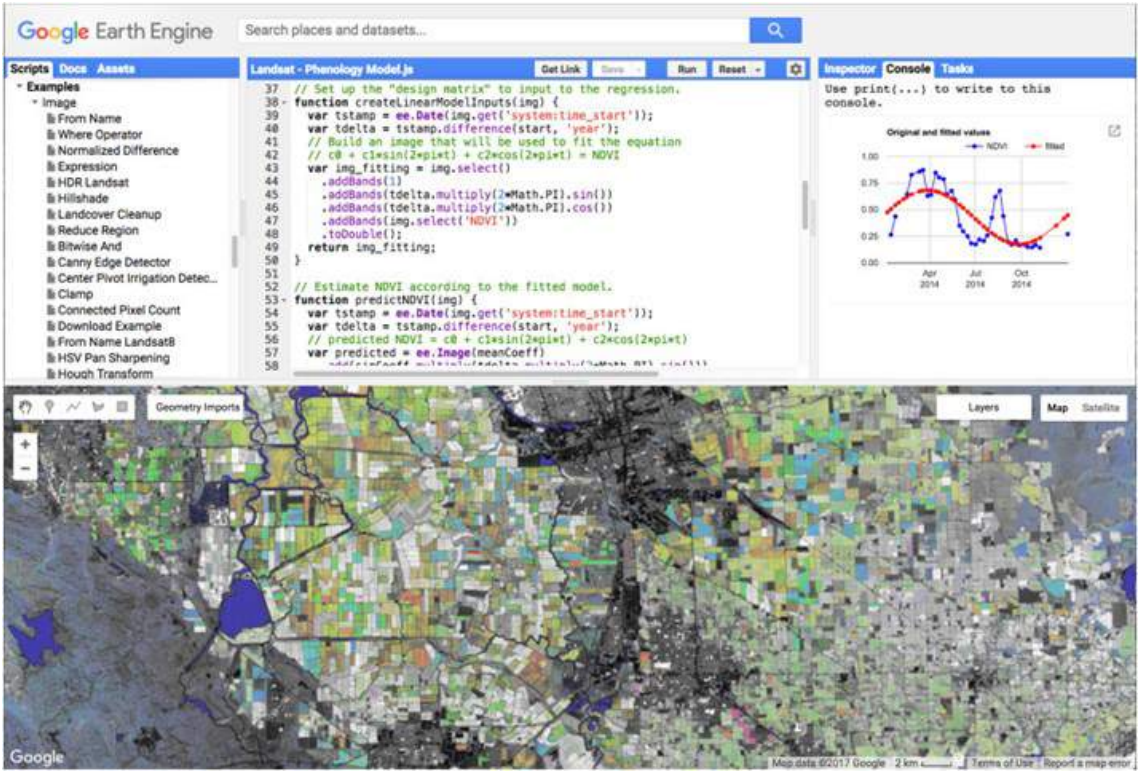
O GEE é uma plataforma de processamento de dados geo-espacial desenvolvida para análise geoespacial baseada em nuvem que facilita o acesso a recursos de computação de alto desempenho para o processamento de conjuntos de dados geoespaciais muito grandes (GORELICK *et al.*, 2017).

O GEE consiste em um catálogo de dados pronto para análise de vários petabytes combinados com um serviço de computação intrinsecamente paralelo de alto desempenho. Ele é acessado e controlado por meio de uma interface de programação de aplicativos (API) (Figura 3.4) acessível pela Internet e um ambiente de desenvolvimento interativo (IDE) associado à web que permite a prototipagem rápida e a visualização dos resultados (GORELICK *et al.*, 2017).

O catálogo de dados públicos do GEE é uma repositório de vários petabytes de conjuntos de dados geoespaciais amplamente usados pela comunidade acadêmica para vi-

sualização e análise de dados geoespaciais. A maior parte do catálogo é composta por imagens de sensoriamento remoto de observação da Terra, incluindo todo o arquivo Landsat, bem como arquivos completos de dados do Sentinel-1 e Sentinel-2, bem comoprevisões climáticas, dados de cobertura do solo e muitos outros conjuntos de dados ambientais, geofísicos e socioeconômicos (GORELICK *et al.*, 2017). A plataforma GEE oferece recursos suficientes para lidar com grandes volumes de imagens de sensoriamento remoto que podem ser usadas, por exemplo, para fins de classificação, como mapeamento de culturas para grandes territórios (SHELESTOV *et al.*, 2017; GHORBANIAN *et al.*, 2020). Para isso é possível customizar processamentos com base em linguagem de programação *JavaScript* ou *Python*.

Figura 3.4: *Ambiente de desenvolvimento do Google Earth Engine.*



Fonte: Gorelick *et al.* (2017).

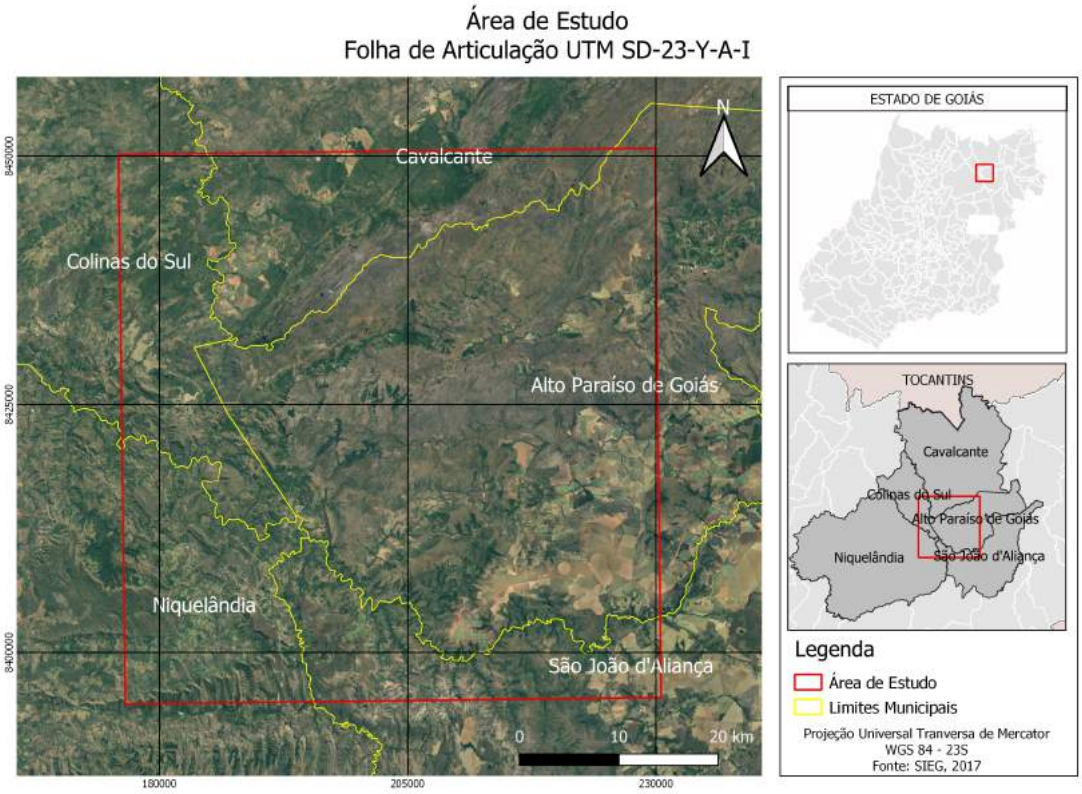
4 METODOLOGIA

Neste capítulo são apresentadas a localização e caracterização da área de estudo, bem como materiais utilizados para realização do trabalho e os procedimentos metodológicos.

4.1 Localização e caracterização da área de estudo

Como área de estudo foi escolhida uma área que fosse representativa dos aspectos fitofisionômicos do Cerrado bem como de seu uso antrópico por agropecuária. Sendo assim, foi definida a quadrícula de articulação SD-23-Y-A-I, uma área aproximada de 3.230 km^2 e perímetro de 230 km. A área abrange os municípios de Alto Paraíso de Goiás, Cavalcante, Colina do Sul, Niquelândia e São João d'Aliança localizados na região Nordeste do Estado de Goiás, na Microrregião da Chapada dos Veadeiros (Figura 4.1).

Figura 4.1: *Localização da área de estudo.*



Fonte: A autora (2021).

4.2 Materiais

Para a realização desse estudo foi necessária a aquisição dos seguintes materiais:

- Imagens do sensor Sentinel-1 com modo de imageamento IW, referentes ao período de Janeiro a Dezembro de 2019;
- Imagens do sensor Sentinel-2 utilizando as Bandas B02 a B08A e B10 a B12, referentes ao período de Janeiro a Dezembro de 2019;
- Plataforma de análise geoespacial baseada na nuvem Google Earth Engine (GORELICK *et al.*, 2017); e
- Mapa de uso e cobertura do solo do MapBiomas (2019), disponível online em: `http://mapbiomas.org/pages/database/mapbiomas_collection`.

4.2.1 Coleta dos dados e Pré-Processamento

Inicialmente foi realizada a aquisição das imagens Sentinel-1 e Sentinel 2 no período de Janeiro a Dezembro de 2019. As imagens do Sentinel -1, no Google Earth Engine, consistem em cenas *Ground Range Detected* (GRD) de nível 1, com as polarizações VH e VV disponíveis e processadas para o coeficiente de retroespalhamento ($\sigma°$) em decibéis (dB). O coeficiente de retroespalhamento representa a área de retroespalhamento do alvo (seção transversal do radar) por unidade de área de solo. Os dados SAR na plataforma GEE já vem com as seguintes correções:

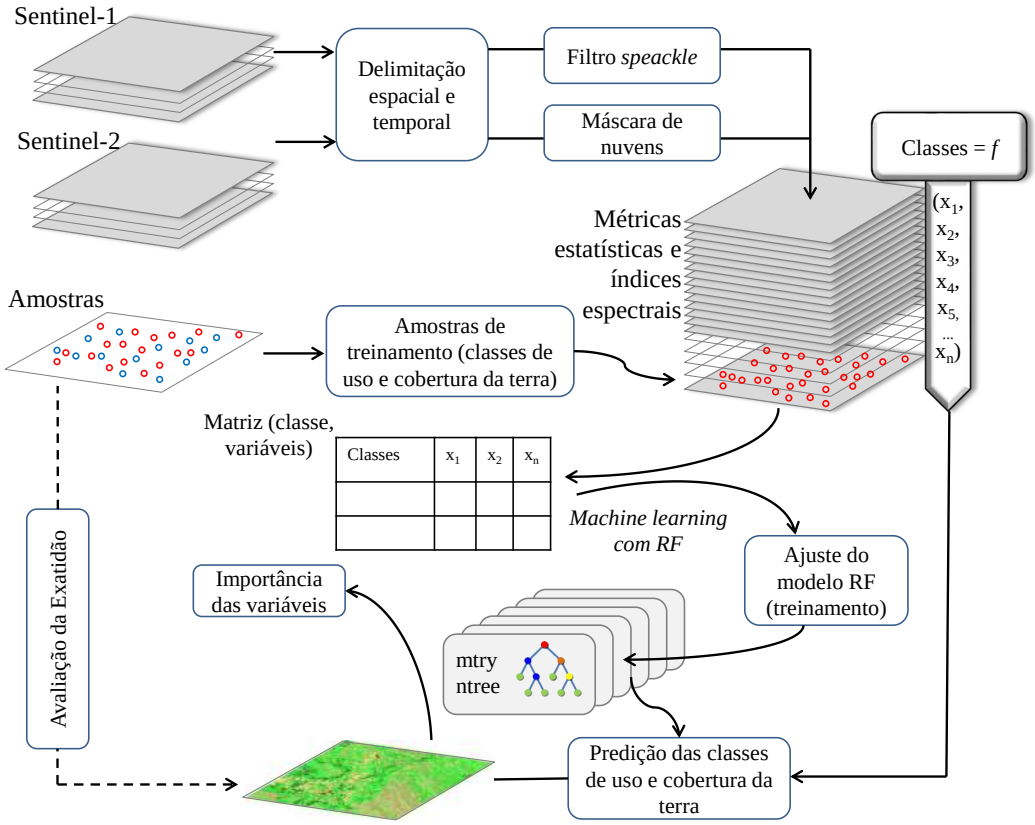
- Arquivos de órbita aplicados;
- Remoção de ruído de borda GRD;
- Remoção de ruído térmico;
- Calibração radiométrica; e,
- Correção do terreno (ortorretificação).

As imagens da Sentinel-2 foram processadas como Level-1C, o que significa que elas são imagens ortorretificadas e que contêm dados de reflectância no topo da atmosfera (ToA). Em seguida, são apresentadas as demais etapas da pesquisa.

4.3 Procedimentos Metodológicos

Os procedimentos metodológicos realizados para a execução do presente estudo estão apresentados na Figura 4.2:

Figura 4.2: *Fluxograma da pesquisa.*



Fonte: A autora (2021)..

4.3.1 Processamento dos dados SAR

A primeira etapa consistiu em aplicar o filtro do tipo passa-baixa para reduzir o efeito speckle e atenuar o efeito do ruído que degrada a qualidade da imagem SAR. Segundo Sant'Anna, Fernandes e Lacava (2001), o ruído dificulta a interpretação visual dos dados, assim como a classificação destas imagens. Por isso, é necessário reduzir este ruído para obter uma análise precisa e detalhada da cena.

Para tanto, foi utilizado o filtro *Refined Lee* (LEE *et al.*, 2008) 7x7, para que qualquer ruído granular (speckle) presente na imagem, causado pela interferência de microondas transmitidas e recebidas na antena do radar, fossem reduzidos. Na última etapa foram realizados os cálculos da média, da mediana, variância e do desvio padrão das série temporal das imagens SAR, isto é, para cada pixel foi calculado a média, mediana, variância e desvio padrão em relação aos valores de $\sigma°$ para cada polarização (VH e VV) ao longo do período de tempo analisado (janeiro a dezembro de 2019).

4.3.2 Processamento dos dados ópticos

Segundo Novo (2010), as nuvens apresentam altíssima reflectância (ρ= 0,7) entre 0,38 μm e 2,5μm, com bandas de absorção amplas em torno de 1,0μm, 1,3μm, e 2μm. As nuvens apresentam alta reflectância no visível, influênciando os índices de vegetação, como por exemplo o NDVI que tende a ter seu valor negativo com essa influência, podendo degradar o processo de classificação supervisionada.

Portanto, das imagens selecionadas, foi aplicado o filtro para que fossem selecionadas apenas imagens com percentual de nuvens menor que 20%. Do mesmo modo, foi empregada a *bitmask* QA60 (banda de correção de nuvens e cirrus do Sentinel-2), para mascarar nuvens opacas e cirros.

As imagens selecionadas foram agrupadas por redutores de média, de mediana, variância e por desvio-padrão. Posteriormente, foram calculados os diversos índices espectrais, conforme expressos no Quadro 4.1. Para cada banda e índice espectral foram calculados a média, mediana, variância e desvio padrão do pixel para todo o período de tempo analisado.

4.3.3 Amostras de treinamento e validação

A seleção das amostras foi realizada a partir de interpretação visual de imagens ópticas do Sentinel-2 e conferência no mapa de uso e cobertura do terra do MapBiomas (MAPBIOMAS, 2019). As amostras coletadas representam 7 (sete) classes de uso e cobertura do solo, as mesmas utilizadas no mapeamento feito pelo MapBiomas (SOUZA *et al.*, 2020). Neste trabalho, a classe denominada "agricultura"é entendida como uma classe de uso da terra (não cobertura) e que essa classe pode ser identificada ao longo de métricas de séries temporais assim como no projeto MapBiomas (SOUZA *et al.*, 2020), por isso é adotado o termo de mapeamento de uso e cobertura da terra.

As amostras receberam um código identificador (ID) sequenciados de 0 a 6, cada ID corresponde a uma classe. As classes de uso e ocupação da terra utilizadas e seus respectivos ID, estão descritas no Quadro 4.2.

Como regra geral, Congalton (1991) recomenda pelo menos 50 amostras por classe. Se a área exceder 500 km^2 ou o número de classes for superior a 12, pelo menos 75–100 amostras devem ser coletadas por classe. Além disso, a amostragem pode ser alocada com respeito à variabilidade dentro de cada categoria.

A distribuição das amostras para cada classe está representada no Quadro 4.3.

Quadro 4.1: *Índices utilizados no estudo.*

Índice	Fórmula	Referência
NDVI	(NIR - RED) / (NIR + RED)	Deering (1978)
EVI2	2.5 * ((NIR - RED) / (NIR + 2.4*RED + 1))	Jiang *et al.* (2008)
CAI	SWIR2 / SWIR1	Nagler *et al.* (2003)
NDWI	(NIR - SWIR1) / (NIR + SWIR1)	Gao (1996)
GCVI	(NIR / GREEN - 1)	Gitelson *et al.* (2003)
Hall Cover	(-RED*0.017 - NIR*0.007 - SWIR2*0.079 + 5.22)	Hall *et al.* (2006)
PRI	(BLUE - GREEN) / (BLUE + GREEN)	Gamon, Serrano e Surfus (1997)
SAVI	(1 + L) * (NIR - RED) / (NIR + RED + 0.5)	Huete (1988)
NDRE	(NIR-REDEDGE) / (NIR+REDEDGE)	Gitelson e Merzlyak (1996)
NDBI	(SWIR1-NIR) / (SWIR1+NIR)	Zha, Gao e Ni (2003)

Nota: *Sendo BLUE=banda com comprimento de onda na faixa do azul; GREEN=verde; RED=vermelho; NIR=infravermelho próximo; REDEDGE=borda do vermelho; SWIR1=infravermelho de ondas curtas; SWIR2=infravermelho de ondas curtas.*[0]

Fonte: A autora (2021).

Quadro 4.2: *Classes utilizadas na classificação e suas respectivas descrições.*

ID	Classes	Descrição
0	Água	Rios, lagos, represas, reservatórios e outros corpos d'água
1	Urbano	Áreas urbanizadas com predomínio de superfícies não vegetadas, incluindo estradas, vias e construções.
2	Floresta	Tipos de vegetação com predomínio de espécies arbóreas
3	Agricultura	Áreas predominantemente ocupadas com cultivos anuais, perenes e semi-perenes.
4	Pastagem	Áreas de pastagens, naturais ou plantadas, vinculadas à atividade agropecuária.
5	Savana	Formações savânicas com estratos arbóreo e arbustivo-herbáceos definidos.
6	Campo	Formações campestres com predominância de estrato herbáceo e algumas áreas de formações savânicas.

Fonte: A autora (2021).

Quadro 4.3: *Distribuição das amostras entre treinamento e validação por classe.*

Classe	Amostras
Floresta	104
Savana	109
Campo	115
Água	100
Agricultura	105
Pastagem	100
Urbano	100

Fonte: A autora (2021).

4.3.4 Classificação do uso e cobertura da terra

Para classificar as imagens foi utilizado o algoritmo classificador RF de aprendizado em máquina (*machine learning*). O algoritmo foi utilizado em conjunto com a Interface de Programas e Aplicativos (API) do Google Earth Engine (GEE) e com o pacote *Classifier* em linguagem *JavaScript*.

As amostras (Quadro 4.3) foram particionadas em amostras de treinamento e validação com a condição de 70% ser destinada ao treinamento e 30% a validação. Para a classificação foi utilizada a função *ee.Classifier.smileRandomForest* tendo como argumento as amostras de treinamento e os dados matriciais. Foram abordados três cenários no estudo:

1. classificação com a integração dos dados ópticos e SAR;
2. classificação utilizando apenas os dados ópticos; e
3. classificação apenas com os dados SAR.

Para a aplicação desse classificador foram necessários a inserção de alguns hiperparâmetros, como o número de árvores construídas (*ntree*) e o número de variáveis amostradas aleatoriamente como candidatas em cada árvore de decisão (*mtry*).

As classificações feitas em todos os cenários tiveram os mesmos hiperparâmetros. Como *mtry* em todos os cenários foram usados a raiz quadrada do número de variáveis (e.g. métricas estatísticas das bandas, polarizações e índices espectrais), como sugerido por Breiman (2001). Para cada cenário ainda foram realizadas duas classificações com diferentes *ntree* (número de árvores). Foram utilizadas o *ntree* 100 e o *ntree* 500, o primeiro *ntree* foi escolhido pois número de árvores menores tinham valores de exatidão

global menores que 80% e maior confusão entre as classes. A partir do *ntree* 100 a exatidão global melhora em todos os cenário, o *ntree* 500 foi escolhido pois após 500 árvores criadas há pouca diferença nas exatidões globais, ou seja, os valores de exatidão global se estabilizam a partir de 500 árvores construídas.

Após a análise dos resultados de kappa e exatidão global em função do *ntree* dos três cenários foi feita a análise dos resultados entre eles com o objetivo de avaliar se a classificação usando os dados ópticos e radar integrados foi melhor do que as classificações individualmente. Além de avaliar qual o melhor cenário, foi avaliado também quais variáveis foram mais importantes para a classificação e a partir da matriz de confusão analisou-se em quais classes tiveram maiores erros e avaliar qual cenário foi mais assertivo para cada classe.

4.3.5 Validação

A avaliação da exatidão do classificador *Random Forest* foi realizada a partir da matriz de confusão ou de contingência. Essa avaliação foi realizada no GEE utilizando o API *ConfusionMatrix*, que possibilitou a tabulação cruzada das classes reais (amostras de validação) e mapeadas (preditas pelo classificador).

O *Random Forest* também fornece uma avaliação da importância relativa dos diferentes recursos ou variáveis durante o processo de classificação. Esse aspecto é útil para estudos de múltiplas fontes, onde a dimensionalidade dos dados é muito alta, e é importante saber como cada variável preditiva influencia o modelo de classificação para poder selecionar as melhores variáveis. Para avaliar a importância de cada característica (por exemplo, banda de imagem de satélite), o RF muda uma das variáveis aleatórias de entrada, mantendo o resto constante, e mede a diminuição na exatidão que ocorreu por meio da estimativa de erro oob e de Impureza de Gini (BREIMAN, 2001).

Matrizes de confusão foram usadas para avaliar a exatidão da classificação. A exatidão global, a exatidão do usuário e do produtor e as estatísticas Kappa e por categoria Kappa foram derivadas das matrizes de confusão, para o cálculo dessas estatísticas foram utilizadas as ferramentas do API *ConfusionMatrix (accuracy (Equação 3-2), consumersAccuracy (Equação , kappa (Equação 3-5), producucerAccuracy (Equação 3-3)).*

Para avaliar a eficácia do sistema radar, a classificação feita por meio da integração dos sensores SAR e óptico foi comparada com uma classificação feita apenas com sensor óptico. Por meio da Impureza de Gini, foi avaliado quais variáveis preditivas tiveram maior influência na classificação de uso e cobertura da Terra, e para isso os parâmetros da classificação (número de árvore, classes, estimativa de out-of-bags e as estatísticas) foram utilizados para construir um dicionário (*ee.Dictionary.get()*) e a partir

desse dicionário é possível construir um gráfico de barras que represente a importância de cada variável na classificação.

4.3.6 Pós-Classificação

Após realizada a classificação, ainda na plataforma do GEE, foi aplicado um filtro de mediana de janela 3x3 com o objetivo de eliminar os pixels espúrios, ou seja, eliminar os pixels isolados na imagem tornando a imagem mais homogênea e com menos ruídos.

No software QGIS foi realizada a reclassificação de todas as imagens classificadas, com o intuito de reclassificar os valores de pixel das imagens, que variam de 0 a 6, para os valores adotados do projeto Mapbiomas (Quadro 4.4).

Quadro 4.4: *Código de legenda utilizada pelo MapBiomas (2019).*

ID	Classe
33	Água
24	Urbano
3	Floresta
18	Agricultura
15	Pastagem
4	Savana
12	Campo

Fonte: A autora (2021).

Após a reclassificação dos pixels foi elaborado os mapas referentes a cada cenário e a cada *ntree* e feita as devidas análises apresentadas no capítulo 5.

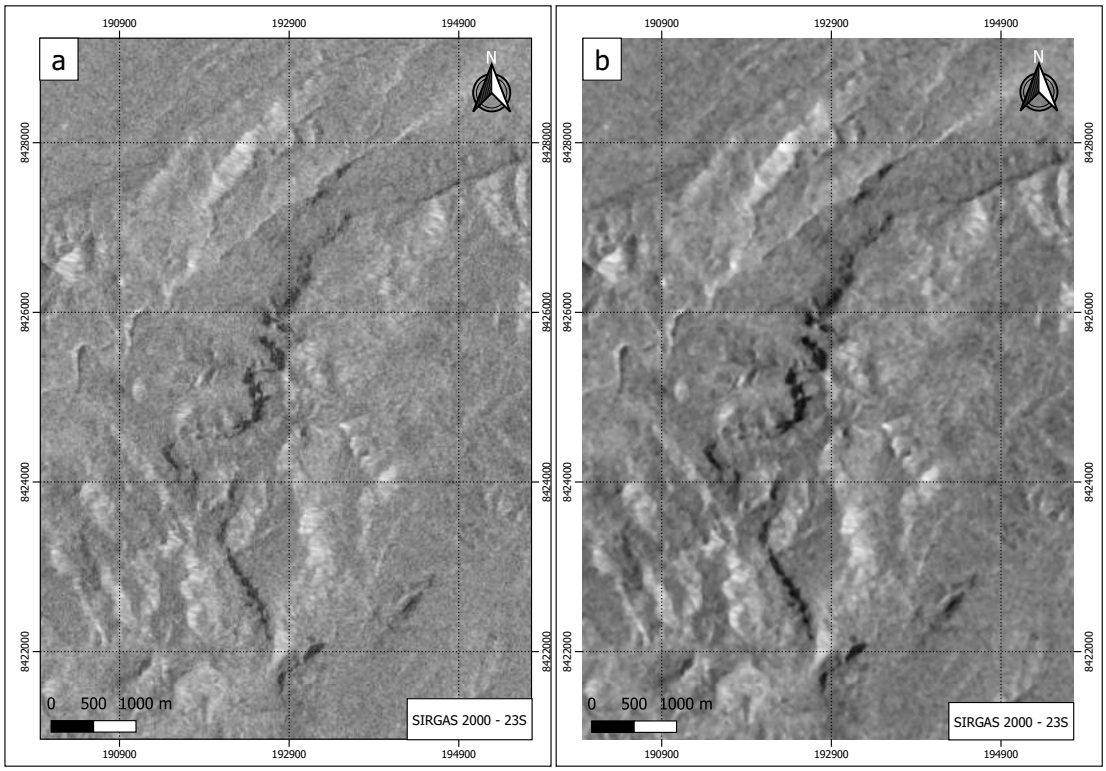
5 RESULTADOS

Neste capítulo são abordados os resultados obtidos utilizando os procedimentos metodológicos propostos. Todo o script usado para este trabalho pode ser acessado pelo endereço: `https://bit.ly/3b6BJUE`.

5.0.1 Imagens SAR

Na Figura 5.1 é possível observar a diferença entre a imagem sem filtro e a imagem filtrada, esta diferença é notada mais facilmente nos corpos d'água, onde a imagem sem filtro apresenta ruídos (pequenos pontos de cor clara) enquanto na imagem filtrada o corpo d'água tem a cor uniforme e totalmente escura.

Figura 5.1: *Cena do Sentinel-1 a) sem filtragem e b)com filtragem do ruído speckle.*

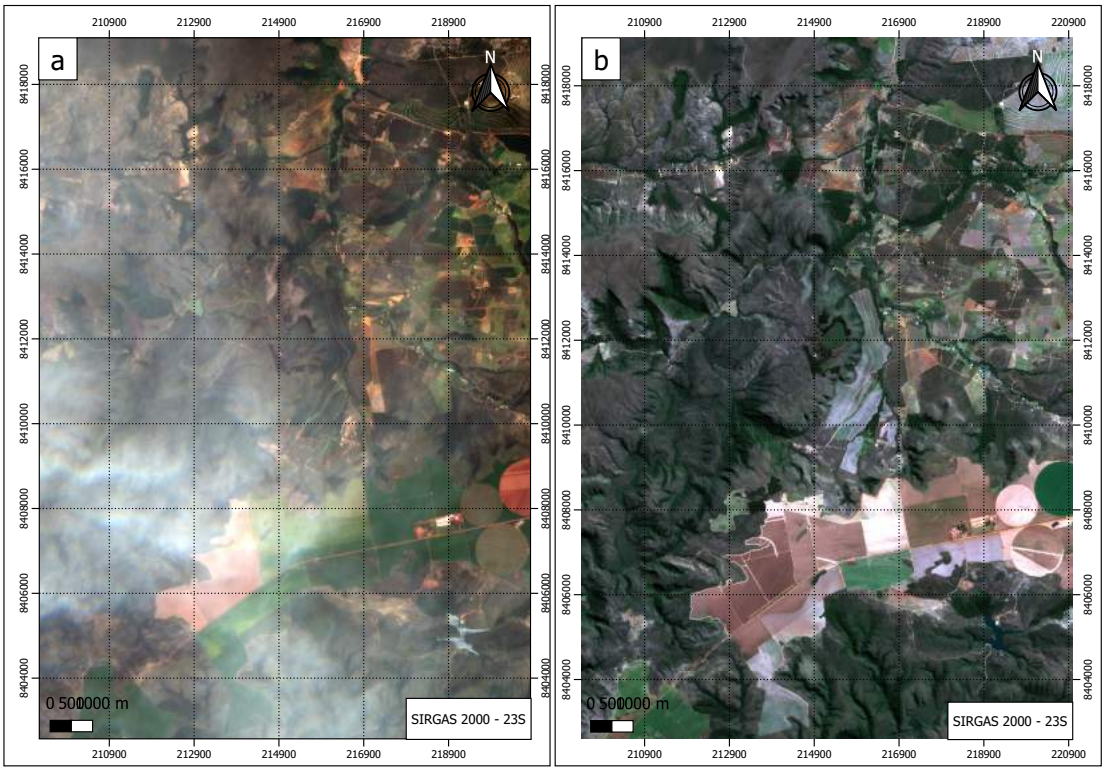


Fonte: A autora (2021).

5.0.2 Imagens Ópticas

Na plataforma do GEE estão disponíveis 248 imagens do S-2 do período de 01/01/2019 a 31/12/2019 na área de estudo. Essas imagens não tiveram nenhum tipo de filtragem de nuvens (Figura 5.2). As nuvens causam grande interferência nas imagens pela sua alta reflectância e por "esconder"os alvos da superfície terrestre para o sensor óptico, com esse fenômeno a resposta de reflectância de cada banda do sensor tem um comportamento alterado quando comparado com imagens sem nuvens.

Figura 5.2: *Cena do Sentinel-2 a) sem filtragem e b)com filtragem de nuvem com porcentagem $\leq$ 20%.*



Fonte: A autora (2021).

Com a filtragem dessas imagens, foram escolhidas apenas imagens com porcentagem de nuvem abaixo de 20%, com isso o número total de imagens que antes eram de 248 passou para 130 imagens. Houve redução do número de imagens principalmente nos meses chuvosos como janeiro, fevereiro, novembro e dezembro.

5.1 Classificação do uso e cobertura da terra

Nesta seção são apresentados os mapas de classificação de uso e cobertura da terra e a avaliação das suas exatidões para os três cenários considerados no presente

estudo.

Para obter os melhores resultados das classificações foi configurado para a classificação limiares de *ntree*, sendo que os melhores valores de exatidão global foram encontrados dentro do intervalo de 100 a 500 *ntree*.

Sendo assim, os resultados apresentados e posteriormente discutidos são referentes as classificações usando os limiares *ntree* 100 e *ntree* 500.

5.1.1 Classificação utilizando a integração de dados ópticos e radar

Nas Tabelas 5.1 e 5.3 são apresentadas as matrizes de confusão tendo como base os limiares de *ntree* de 100 e 500. A classificação com o *ntree*=100 teve exatidão global e kappa iguais a 90,29% e 88,66%, respectivamente. Analisando apenas a matriz de confusão (Tabela 5.1), é possível perceber que a classe água foi a que teve menor confusão entre as classes, enquanto a classe savana foi a que mais se confundiu.

Tabela 5.1: *Matriz de Confusão da classificação utilizando a integração de dados ópticos e radar com o* ntree = *100.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	32	0	1	0	0	0	0
Urbano	0	25	0	1	2	0	0
Floresta	1	0	28	0	0	3	0
Agricultura	0	0	0	23	3	0	0
Pastagem	0	0	0	0	27	1	1
Savana	0	0	1	0	0	26	5
Campo	0	0	0	0	0	1	25

Fonte: A autora (2021).

Analisando as exatidões por classe, confirmou-se que a classe savana foi a classe que teve o menor valor de kappa, e portanto, maior confusão entre as classes temáticas, particularmente, com a classe campo. As classes de água, campo e pastagem obtiveram kappa superior a 90%, todas as classes tiveram kappa acima de 75%.

Tabela 5.2: *Exatidão do produtor, exatidão do usuário e kappa condicional de cada classe com* ntree = *100.*

Classe	**Exat. do produtor (%)**	**Exat. do usuário (%)**	**Kappa cond. (%)**
Água	96,97	96,97	96,39
Inf. Urbana	89,28	100	87,81
Floresta	87,50	96,55	85,37
Agricultura	92,31	96,00	86,94
Pastagem	93,10	87,10	91,84
Savana	90,62	85,29	77,93
Campo	96,15	86,20	95,47

Fonte: A autora (2021).

A classificação realizada com o *ntree*= 500 (Tabela 5.3) apresentou melhora nas classes savana e floresta se comparada com a classificação usando um *ntree* = 100.. A classificação obteve Exatidão Global e Kappa iguais a 94,12% e 93,07%, respectivamente.

Tabela 5.3: *Matriz de Confusão da classificação utilizando a integração de dados ópticos e radar com o* ntree = *500.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	21	0	1	0	0	0	0
Urbano	0	31	0	0	1	0	0
Floresta	0	0	30	0	0	2	0
Agricultura	0	0	0	28	3	0	0
Pastagem	0	0	0	1	11	1	0
Savana	0	0	1	0	0	30	0
Campo	0	0	0	0	0	1	25

Fonte: A autora (2021).

Analisando a Tabela 5.4, observa-se a melhora significativa da classe savana em termos de kappa condicional e uma aumento do kappa de todas as outras classes,

exceto as classes de água, campo, pastagem e agricultura que não apresentaram diferenças significativas.

Com base na matriz de confusão, os valores do kappa geral e das estatísticas por classe, conclui-se que a classificação com *ntree*= 500 analisada estatísticamente a exatidão global e o kappa geral não foram significativamente diferentes, porém analisando os resultados por classe, nota-se que melhorou significamente o kappa de algumas classes.

Tabela 5.4: *Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com* ntree= *500.*

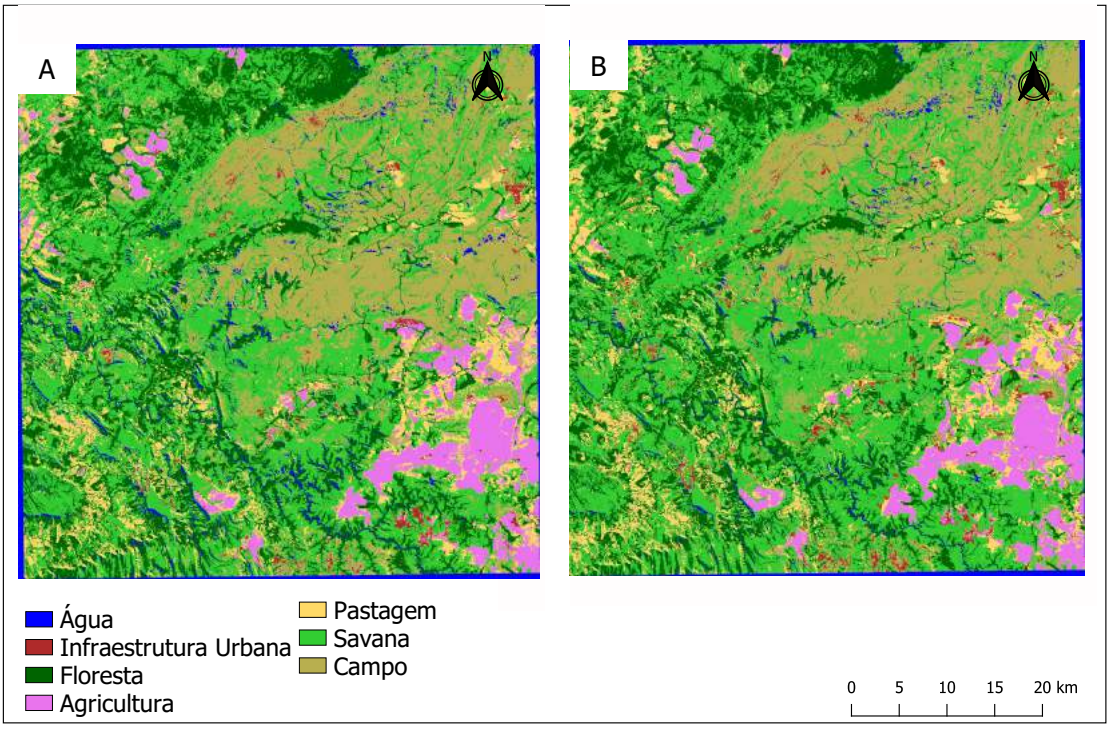
Classe	**Exat. do Produtor (%)**	**Exat. do Usuário (%)**	**Kappa cond. (%)**
Água	93,33	96,43	94,88
Inf. Urbana	96,87	100	96,25
Floresta	96,87	94,59	92,46
Agricultura	97,06	100	88,55
Pastagem	95,0	75,00	83,27
Savana	81,25	81,33	96,06
Campo	96,55	92,00	95,56

Fonte: A autora (2021).

Assim como na comparação das estatísticas gerais, na comparação visual (Figura 5.3) também não há diferenças significativas na classificação.

A variável de maior importância na classificação feita com 100 árvores foi o valor mínimo e a mediana do índice CAI que é a razão entre as bandas de infravermelho de ondas curtas, região onde a reflectância da vegetação é dominada pelo conteúdo de absorção pela água.

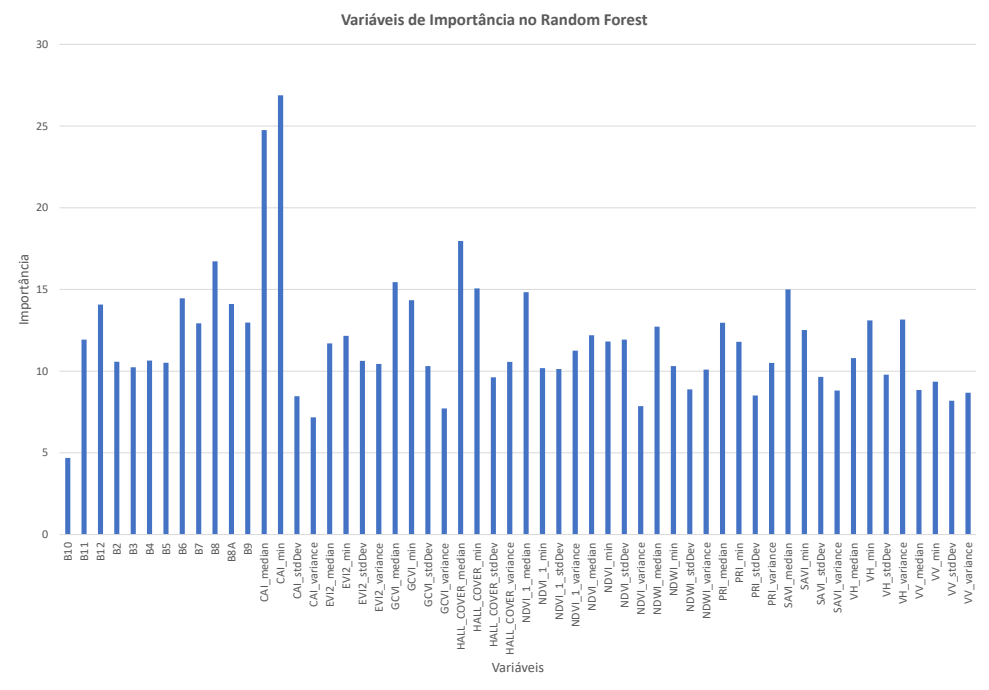
Figura 5.3: *Classificação com integração de dados ópticos e radar com diferentes números de árvore.*



Nota: *Classificação A) com* ntree= *100 e B) com* ntree=*500.*[0]

Fonte: A autora (2021).

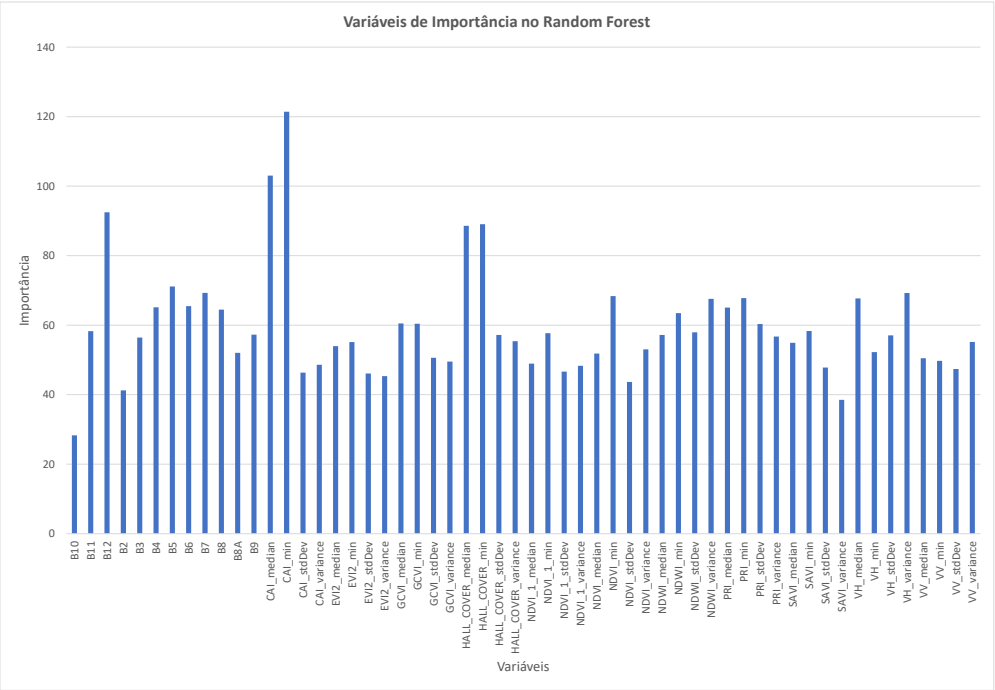
Figura 5.4: *Importância de cada variável usada na classificação do RF, com 100 árvores, pelo índice de Gini.*



Fonte: A autora (2021).

Na classificação com 500 árvores o índice CAI também se mostrou importante, variáveis como a banda 12 e os valores mínimos e medianos do índice Hall Cover, também apresentaram alta importância. A banda 12 corresponde a banda do infravermelho de ondas curtas (SWIR2), o índice Hall Cover tem em sua formulação as bandas no vermelho, infravermelho e o infravermelho de ondas curtas.

Figura 5.5: *Importância de cada variável usada na classificação do RF, com 500 árvores, pelo índice de Gini.*



Fonte: A autora (2021).

5.1.2 Classificação utilizando dados de radar

Na classificação baseada em dados de radar, o índice kappa e a exatidão global da classificação com *ntree*=100 foram de 46,15% e 53,81%, respectivamente. As classes savana e campo na matriz de confusão (Tabela 5.5) foram as que mais se confundiram. Essa confusão também ocorreu em outras classes, tal como savana, pastagem, campo, floresta e urbano.

Analisando o kappa condicional de cada classe (Tabela 5.6), apenas as classes água e agricultura tiveram resultados maiores que 80%, enquanto classes como campo, savana, floresta e infraestrutura urbana não alcançaram 50%.

Tabela 5.5: *Matriz de Confusão da classificação utilizando dados de radar com* ntree = *100.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	26	0	0	0	0	1	3
Urbano	1	16	8	0	2	2	1
Floresta	2	9	14	0	0	3	5
Agricultura	0	0	0	24	5	0	0
Pastagem	0	2	2	1	19	1	6
Savana	0	6	7	0	9	9	7
Campo	2	1	3	1	4	3	5

Fonte: A autora (2021).

Tabela 5.6: *Exatidão do produtor, Exatidão do usuário e Kappa condicional de cada classe com* ntree= *100.*

Classe	**Exat. do Produtor (%)**	**Exat. do Usuário (%)**	**Kappa cond. (%)**
Água	86,67	95,73	84,36
Inf. Urbana	53.33	93,75	44,32
Floresta	42,42	53,33	31,30
Agricultura	82,75	64,70	80,32
Pastagem	61,29	100,0	52,46
Savana	23,68	66,67	16,09
Campo	26,31	55,55	15,44

Fonte: A autora (2021).

Na classificação das imagens de radar utilizando 500 árvores de decisões (Tabela **??**, obteve-se exatidão global e kappa iguais a 79,29% e 75,78%, respectivamente. Em comparação com a classificação realizada com *ntree*=100, a classificação com *ntree* = 500 teve resultados melhores em relação a exatidão global e também com os resultados por classe. Apenas a classe água teve seu kappa reduzido, as outras classes tiveram um aumento do valor de kappa condicional, principalmente as classes de savana, campo e

floresta. A classe com o menor valor de Kappa foi a classe referente as infraestruturas urbanas, o que é esperado por ser uma classe com uma mistura espectral grande pelo grande números de elementos presentes nas cidades.

Tabela 5.7: *Matriz de Confusão da classificação utilizando dados de radar com* ntree = *500.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	25	1	1	0	0	2	3
Urbano	0	14	6	0	2	5	1
Floresta	1	1	28	0	0	1	0
Agricultura	0	0	0	24	2	0	0
Pastagem	0	0	1	2	25	2	2
Savana	2	0	0	0	3	14	0
Campo	0	0	1	0	1	1	27

Fonte: A autora (2021).

Tabela 5.8: *Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com* ntree= *500.*

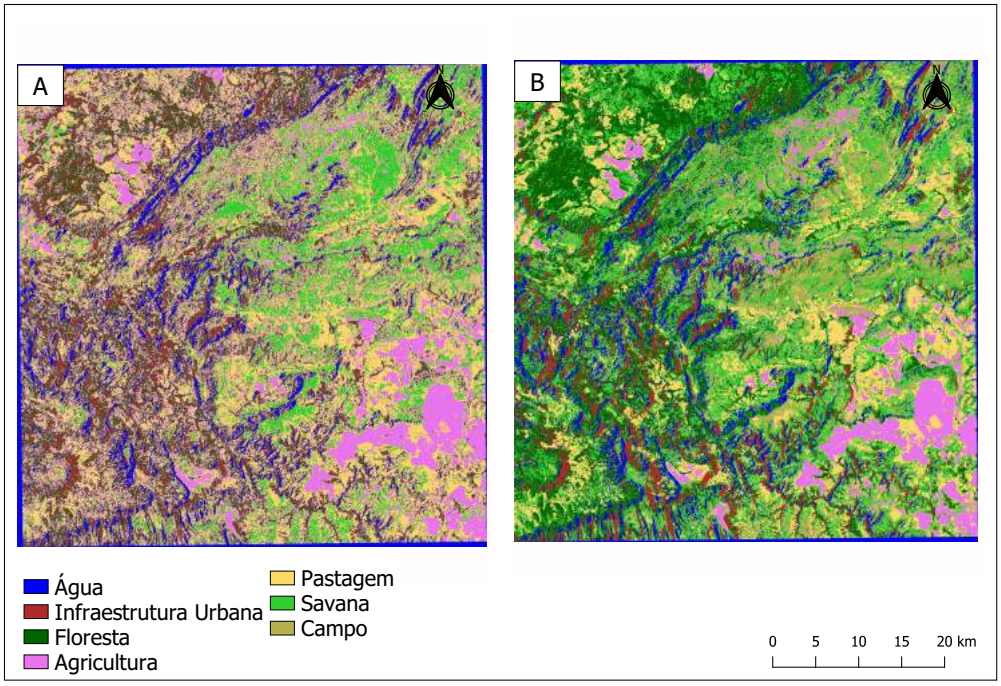
Classe	**Exat. do Produtor (%)**	**Exat. do Usuário (%)**	**Kappa cond. (%)**
Água	78,12	89,28	74,52
Inf. Urbana	50.0	87,50	45,60
Floresta	90,32	75,67	88,10
Agricultura	92,30	92,30	91,14
Pastagem	78,12	75,75	73,75
Savana	73,68	56,0	69,88
Campo	90,0	81,81	88,0

Fonte: A autora (2021).

Na Figura 5.6 estão apresentadas as classificações realizadas com as imagens de radar utilizando 100 e 500 números de árvores de decisão. Diferentemente das

classificações realizadas com a integração dos dados ópticos e radar, as classificações realizadas apenas com dados de radar tiveram muitas mudanças entre uma e outra, não só analisando as estatísticas de validação mas também quando se compara as imagens resultantes das classificações.

Figura 5.6: *Classificação com dados de radar com diferentes números de árvore.*



Nota: *Classificação A) com* ntree= *100 e B) com* ntree=*500.*[0]

Fonte: A autora (2021).

As áreas de agricultura presentes na classificação com *ntree*=100 também são apresentadas na classificação com *ntree*=500, o que justifica os altos valores das exatidões do produtor e kappa condicional dessa classe nas duas classificações. Na classificação com *ntree*=100 é possível perceber grandes áreas classificadas como pastagem enquanto nessas mesmas regiões na classificação de *ntree*=500 foram identificadas outras classes. O aumento de árvores melhorou a classificação das classes de pastagem quando analisadas visualmente e consequentemente melhorou as classes de campo e savana que estavam se confundindo entre elas e com a pastagem.

Em áreas com maior declividade, em ambas as imagens, o que seriam áreas de sombra foram classificadas como água, o que justifica o grande número de áreas classificadas como água quando comparadas com a classificação dos dados combinados (Figura 5.3)

Na classificação utilizando os dados ópticos e de radar combinados a variável de radar mais importante foi a mediana da polarização cruzada VH. Na classificação

Figura 5.7: *Importância de cada variável usada na classificação do RF, com* ntree = *100 e* ntree = *500, pelo Índice de Gini.*

Fonte: A autora (2021).

utilizando apenas dados de radar, além da mediana VH a mediana da polarização VV também se destacou em ambas as classificações variando o parâmetro *ntree* entre 100 e 500.

5.1.3 Classificação utilizando dados ópticos

A classificação utilizando apenas os dados ópticos com *ntree* = 100 teve como Exatidão Global e Índice Kappa valores iguais a 87,98% e 85,95%, respectivamente. Analisando a matriz de confusão (Tabela 5.9) e os resultados das métricas de validação (Tabela 5.10) é possível perceber que houve poucas confusões no âmbito geral. As classes que mais se confundiram entre si foram pastagem e agricultura, enquanto a classe água teve o melhor resultado dentre as classes. Todas as classes tiveram kappa condicional maior que 70%.

Tabela 5.9: *Matriz de Confusão da classificação utilizando dados ópticos com* ntree = *100.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	28	0	0	0	0	0	0
Urbano	0	28	0	2	1	0	0
Floresta	2	0	24	0	0	1	0
Agricultura	0	5	0	25	1	0	0
Pastagem	0	0	0	4	23	4	0
Savana	0	0	1	0	0	32	2
Campo	0	2	0	0	0	0	23

Fonte: A autora (2021).

Tabela 5.10: *Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com* ntree= *100.*

Classe	**Exat. do Produtor (%)**	**Exat. do Usuário (%)**	**Kappa cond. (%)**
Água	100,0	96,55	100,0
Inf. Urbana	90.32	90,32	88,36
Floresta	92,59	96,15	87,37
Agricultura	93,55	87,88	77,26
Pastagem	87,10	93,10	70,67
Savana	97,14	91,89	89,57
Campo	92,0	100,0	90,91

Fonte: A autora (2021).

A Exatidão Global e o Índice Kappa da classificação utilizando ao *ntree = 500* é de 91,44% e 89,99%. Analisando a matriz de confusão (Tabela 5.11) e os resultados das métricas de validação (Tabela 5.12) é possível perceber que houve poucas confusões no âmbito geral. As classes que mais se confundiram entre si foram pastagem e agricultura. A classe floresta teve confusão com a classe savana, esta por sua vez teve kappa igual

a 100% juntamente com a classe de campo. Todas as classes tiveram kappa condicional acima de 75%.

Tabela 5.11: *Matriz de Confusão da classificação utilizando dados ópticos com* ntree = *500.*

	Água	Urbano	Floresta	Agricultura	Pastagem	Savana	Campo
Água	31	0	1	0	0	0	0
Urbano	0	33	0	0	1	0	0
Floresta	0	0	31	0	0	5	0
Agricultura	0	1	0	27	4	0	0
Pastagem	0	0	0	4	33	3	0
Savana	0	0	0	0	0	24	0
Campo	0	0	0	0	0	0	24

Fonte: A autora (2021).

Tabela 5.12: *Exatidão do Produtor, Exatidão do Usuário e Kappa Condicional de cada classe com* ntree= *500.*

Classe	**Exat. do Produtor (%)**	**Exat. do Usuário (%)**	**Kappa cond. (%)**
Água	96,87	100,0	96,37
Inf. Urbana	97.06	100,0	96,53
Floresta	86,11	100,0	83,77
Agricultura	84,37	96,55	81,84
Pastagem	82,50	89,74	78,89
Savana	100,0	80,0	100,0
Campo	100,0	100,0	100,0

Fonte: A autora (2021).

Os valores de Kappa e Exatidão Global usando *ntree* de 100 e 500 não são significamente diferentes em um nível de confiança de 5%. As classes que apresentaram um aumento significativo no índice kappa na classificação *ntree* = 500 foram as áreas de

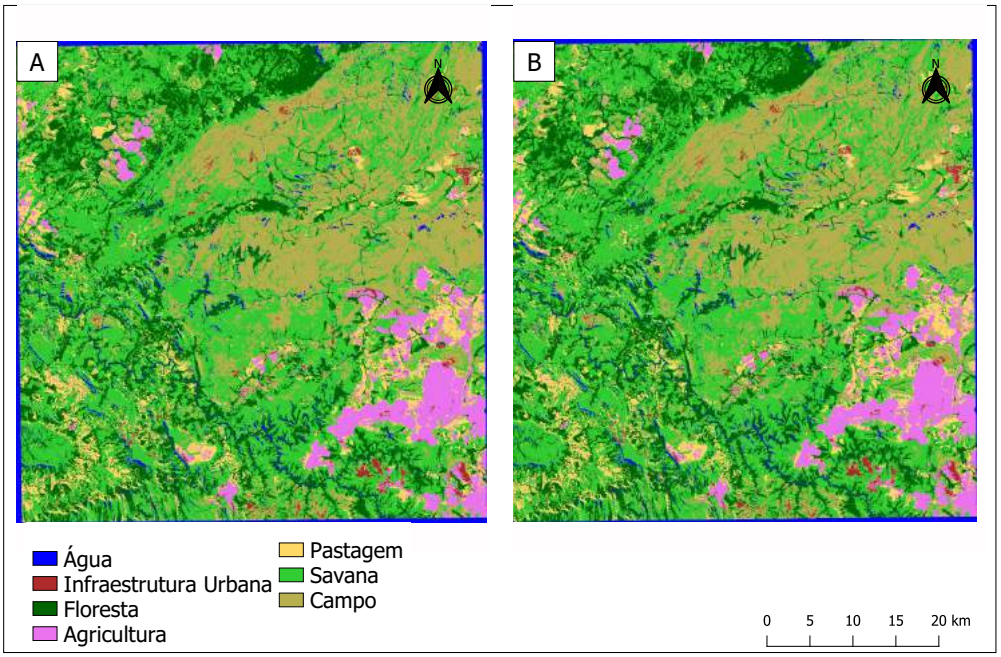
infraestrutura urbana, pastagem, savana e campo. As classes de água e floresta tiveram aumento não muito significativos. A classe floresta teve uma confusão com a classe savana, consequentemente, reduziu seu índice Kappa, em comparação com a classificação utilizando *ntree* = 100.

Assim como na classificação feita com a integração dos dados, visualmente as classificações usando apenas dados ópticos também não apresentaram diferenças significativas ao utilizar um ntree de 100 e 500 (Figura 5.8).

Assim como na classificação utilizando os sensores integrados, na classificação utilizando apenas o sensor óptico, as variáveis mais importantes foram os valores mínimos e medianos do índice CAI. Para a classificação realizada com 100 árvores (Figura 5.9), após o índice CAI aparecem as bandas 05 e 09, banda Red Edge 1 e a banda de correção atmosférica para vapor d'água, e o índice Hall Cover.

Já na Figura 5.10 as variáveis mais importantes foram os valores mínimos e medianos do índice CAI seguidos do valor mínimo do índice Hall Cover e da banda 12 (SWIR2 - infravermelho de ondas curtas).

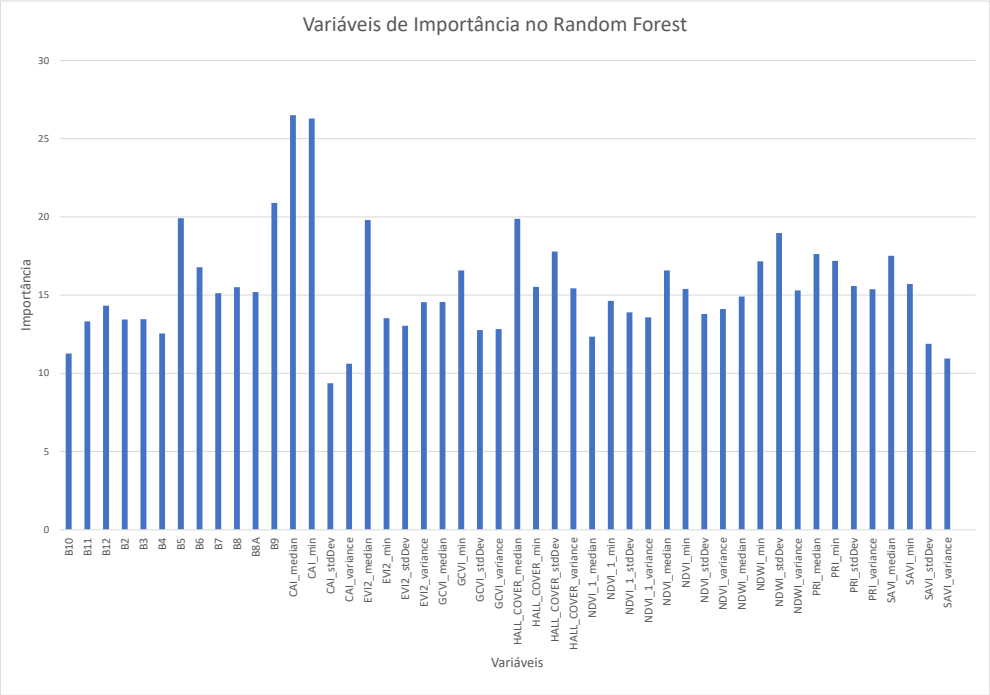
Figura 5.8: *Classificação com dados ópticos com* ntree *100 e 500.*



Nota: *Classificação A) com* ntree= *100 e B) com* ntree=*500.*[0]

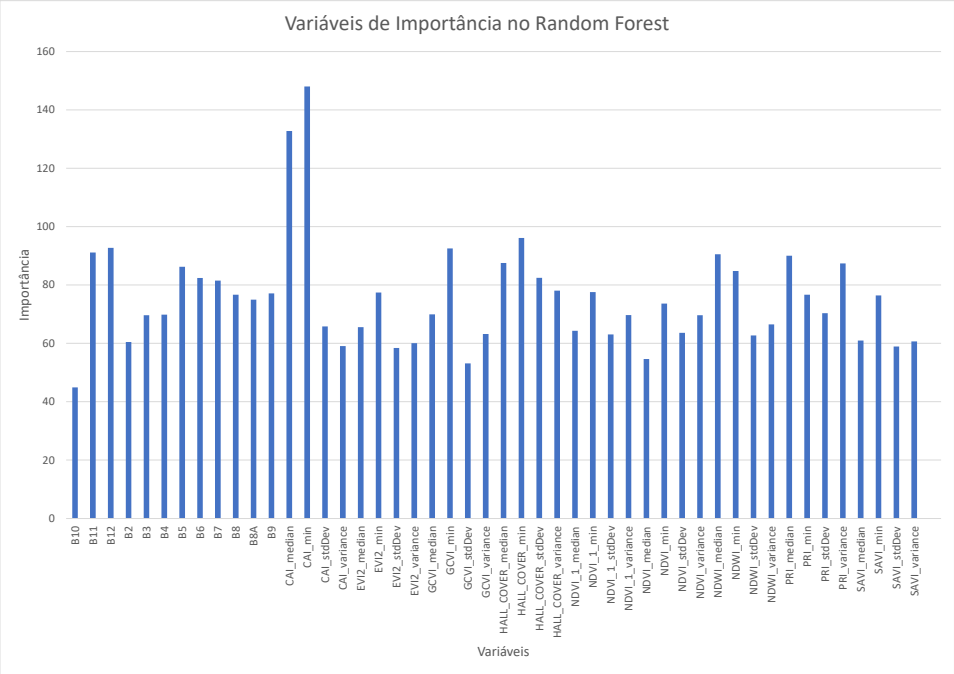
Fonte: A autora (2021).

Figura 5.9: *Importância de cada variável usada na classificação do RF, com 100 árvores, pelo índice de Gini.*



Fonte: A autora (2021)..

Figura 5.10: *Importância de cada variável usada na classificaçãodo RF, com 500 árvores, pelo índice de Gini.*



Fonte: A autora (2021)..

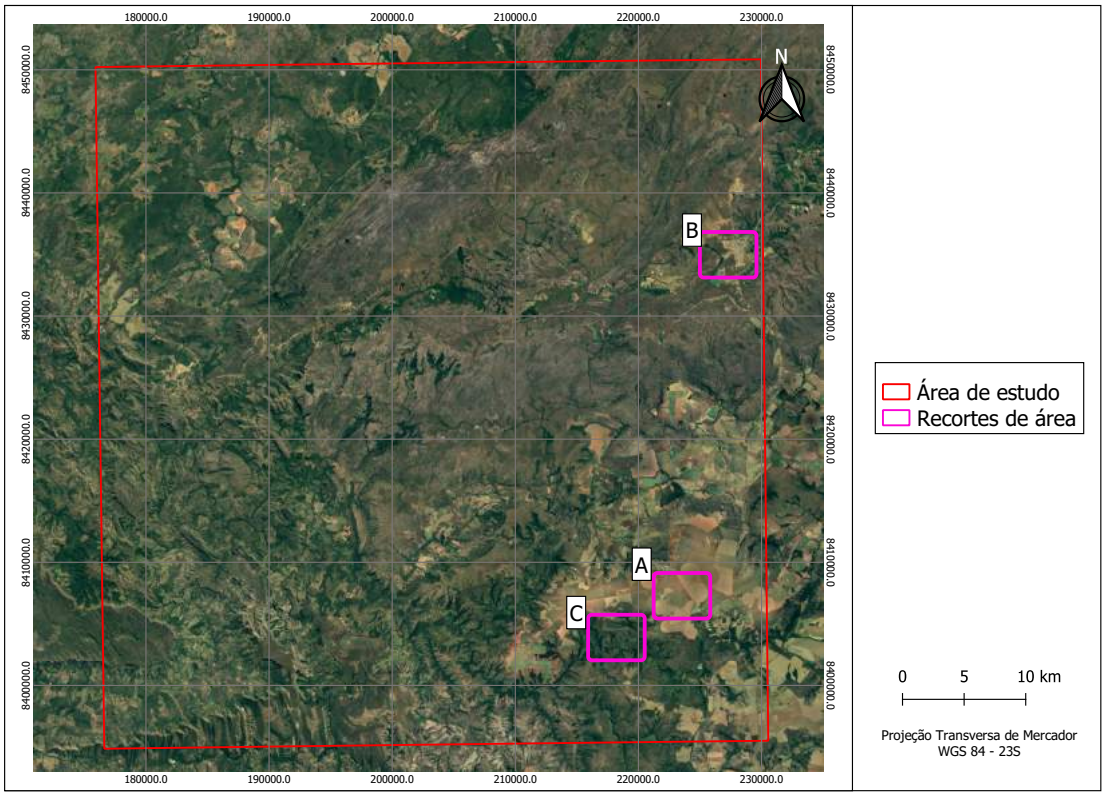
5.1.4 Comparação dos melhores modelos de cada cenário

Para os três cenários, o maior valor de índice kappa e exatidão global foram resultados decorrentes da classificação com 500 árvores construídas. A classificação utilizando apenas as imagens de radar foi inferior as demais classificações. Em um nível de significância de 5% as classificações feitas com a integração das imagens ópticas e radar e a classificação realizada apenas com as imagens ópticas não foram estatisticamente diferentes.

Na Figura 5.12 está representada uma comparação visual do melhor modelo (*ntree*=500) de cada cenário (classificação com integração, utilizando apenas o sensor óptico e utilizando apenas radar).Essas áreas em destaque são representadas na Figura 5.11. Analisando a Figura 5.12 é possível notar que a classificação utilizando apenas os dados de radar, comparando com a imagem de referência (Google Earth) e as demais classificações, teve bastante confusão em diversas classes principalmente a de infraestrutura urbana. Áreas de floresta e savana foram confundidas com água nas áreas de maior declividade e área urbana nas áreas com vegetação rasteira. As classes melhor discriminadas na classificação da imagem de radar forma as de pastagem e agricultura.

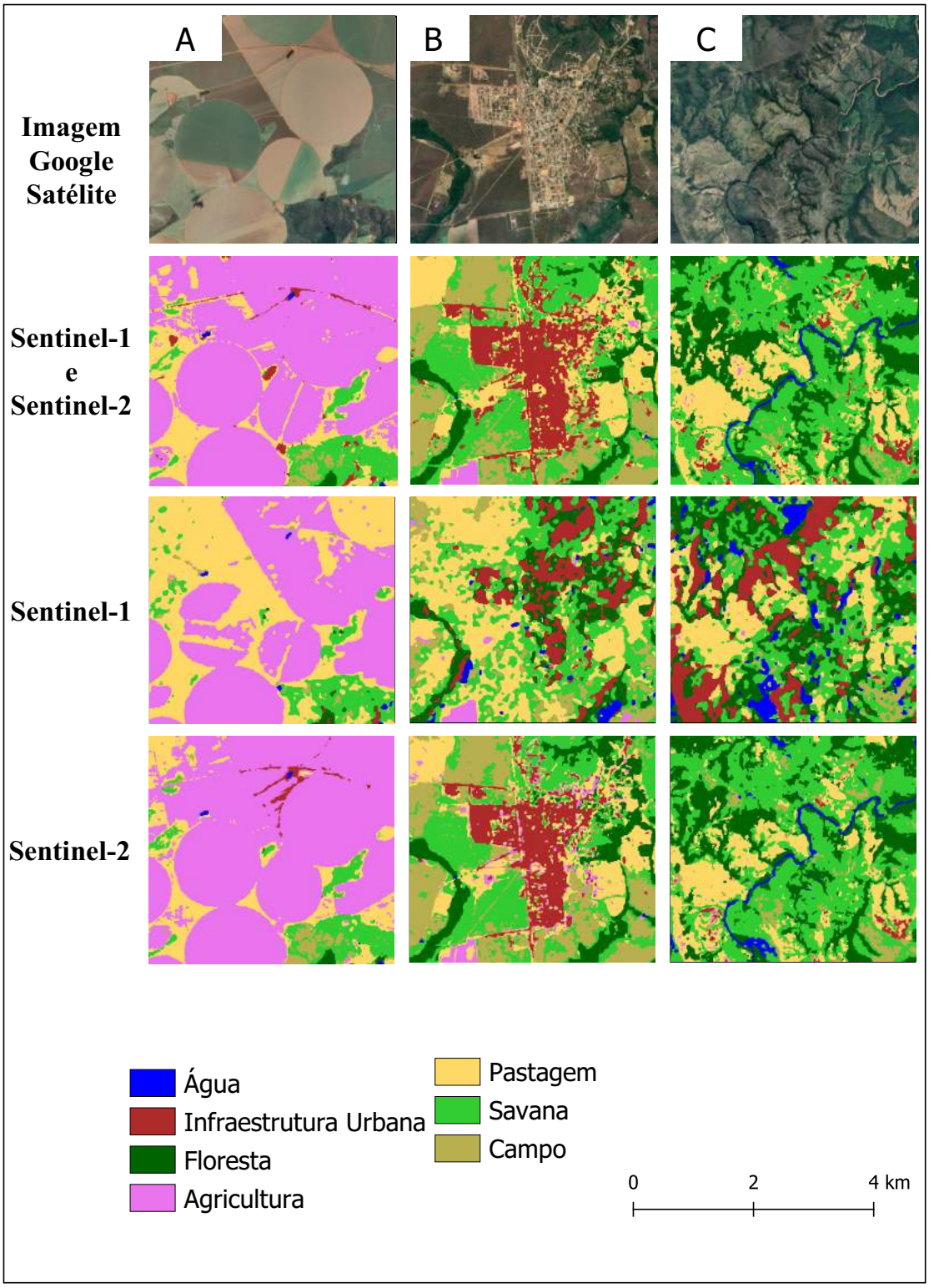
Nas imagens referentes as classificações utilizando apenas o sensor óptico e utlizando a integração dos dados de radar e óptico é possível notar a similaridade entre elas, com poucas diferenças visuais. Nessas classificações é possível notar que elas conseguiram delinear bem, pelo menos visualmente, todas as classes.

Figura 5.11: *Áreas utilizadas para comparação.*



Fonte: A autora (2021)..

Figura 5.12: *Comparação entre os três cenários com* ntree = *500.*



Fonte: A autora (2021)..

6 DISCUSSÕES

Os resultados mostram que os cenários que usam apenas os dados do Sentinel-1 apresentam desempenho inferior ao Sentinel-2 na maioria das classes. O melhor cenário utilizando o Sentinel-2 (*ntree*=500) produziu uma exatidão global de 91,44% em comparação com 79,29% para o melhor modelo utilizando apenas o Sentinel-1 (*ntree*=500), uma diferença significativa (teste z com nível de confiança de 95%). Esses resultados são compatíveis com o estudo de Lopes *et al.* (2020) que comparou imagens ópticas com imagens SAR nas regiões subsaarianas e com Borges *et al.* (2020) que comparou a integração dos dados ópticos e SAR em diferentes estações em uma área de conservação na Tanzânia.

Os resultados também mostraram que a combinação dos dados do Sentinel-1 e Sentinel-2 alcançaram maior exatidão global em comparação com os cenários de classificação com um único sensor. A integração desses dados aumentou a exatidão para as classes de floresta, agricultura e pastagem. Sano *et al.* (2010) também obteve maior exatidão na classificação geral onde usou sensores ópticos e radar em regiões de savana e pastagem.

Chatziantoniou, Psomiadis e Petropoulos (2017) obtiveram exatidão global e kappa iguais a 94,82% e 93,62% no mapeamento em zonas úmidas no Mediterrâneo utilizando a integração do Sentinel-1 e Sentinel-2 classificando com o algoritmo de aprendizado de máquina *Support Vector Machine* (SVM). Zhang *et al.* (2019) sugerem a integração SAR a dados ópticos, pois eles podem minimizar erros com classificações, particularmente em áreas altamente produtivas (por exemplo, pântanos, irrigação, campos e gramíneas perenes) que podem ser erroneamente classificados como árvores.

A variável de radar que mais se destacou na classificação foi a polarização cruzada VH do Sentinel-1, resultado que está em concordância com o trabalho de Mendes *et al.* (2019) que fez classificações utilizando Sentinel-1, Sentinel-2, TanDEM-X e ALOS-PALSAR2. Porém, em seu trabalho a polarização cruzada HV que mais se destacou foi a do ALOS-PALSAR2, com banda L que tem maior potencial de penetração nos alvos terrestres do que a banda C.

Em relação a importância das variáveis no processo de classificação, a razão das bandas SWIR2 pela SWIR1 dado pelo índice CAI foi a variável mais importante, tanto pela métrica estatística da mediana, quanto pela mínima. Isso provavelmente ocorreu pelo fato dessa razão de bandas no infravermelho de ondas curtas realçar o conteúdo de água na folha (PONZONI; SHIMABUKURO; KUPLICH, 2015), tendo em vista que a

vegetação no Cerrado é caracterizada por apresentar sazonalidade em suas características fitofisionômicas em períodos de estiagem e chuva. Por sua vez, o índice Hall Cover foi outra variável importante na separação de classes com o uso do algoritmo RF. O Hall Cover foi originalmente proposto para estimativa de biomassa acima do solo (HALL *et al.*, 2006) e tem sido utilizado pela equipe do MapBiomas (MAPBIOMAS, 2019) como uma das variáveis no projeto de mapeamento de uso e cobertura da terra no Brasil e também no Chaco na América do Sul.

Individualmente, a banda espectral mais importante foi a banda 12 (infravermelho de ondas curtas). No comprimento de onda dessa banda ($\tilde{2}$.190 nm) é possível ter uma nítida separação entre solo exposto e vegetação (PONZONI; SHIMABUKURO; KUPLICH, 2015; NOVO, 2010), tal característica pode ter contribuído para auxiliar na separação das classes de uso e cobertura da terra.

Já em relação aos dados SAR, a polarização mais importante foi a VH com as métricas estatísticas de meadiana e variância. Em estudo efetuado por Saraiva (2015) a polarização VH permitiu melhor discriminação de culturas agrícolas do que a polarização VV. O retroespalhamento volumétrico entre os diferentes alvos analisados, dado pela polarização VH, pode ter contribuido para essa variável ser mais importante entre os dados SAR.

Mercier *et al.* (2019) ao usarem o Sentinel-1 na classificação teve como melhor variável a polarização VV o que está em discordância com o atual trabalho que encontrou como variável de radar mais importante a variável VH. Por outro lado, Mercier *et al.* (2019) em uma classificação utilizando apenas o Sentinel-2 também tiveram como variável mais importante a banda 12 (infravermelho de ondas curtas), e ao classificar utilizando a integração dos dados do Sentinel - 1 e 2 as variáveis mais importantes foram as variáveis derivadas do sensor óptico, resultados esses que estão em conformidade com os resultados da presente pesquisa.

Em relação ao uso do processamento em nuvem, neste estudo foram utilizadas 31 imagens SAR e 130 imagens ópticas. Na plataforma Copernicus as imagens ópticas e SAR são disponibilizadas para download sem nenhuma ou pouca correção com tamanho de 1,6 Gb e 500 Mb em formato compactado, respectivamente, totalizando em mais 120 Gb de imagens compactadas. Além do grande consumo da memória de disco, processar esses dados demandaria muito tempo e um alto esforço de processamento da máquina. Kumar e Mutanga (2018) compararam o processamento em nuvem com o processamento em disco local *desktop* utilizando o trabalho de Hansen *et al.* (2013) que identificaram mudanças na cobertura florestal em escala global durante 12 anos (2000-2012) usando 654.178 cenas do Landsat-7, totalizando em 70 terabytes de dados. O processamento durou 100 horas no GEE em comparação com aproximadamente 1.000.000 horas que teria levado em um computador desktop. Tais números realçam o potencial e vantagens

da capacidade de armazenamento e processamento em nuvem especialmente em estudos que dependem de séries temporais de imagem, como o fornecido pelo GEE.

De acordo com Magidi *et al.* (2021), há uma grande vantagem na combinação da técnica do RF com plataformas de gerenciamento de *big data*, como o GEE, que é o processamento de grandes conjuntos de dados com inúmeras variáveis em um curto espaço de tempo. Porém, é apontado como desvantagem do algoritmo sua tendência de se ajustar excessivamente aos dados de treinamento (*overfitting*), razão principal para não fornecerem uma exatidão ainda melhor. Contudo, Oshiro (2013) explica a técnica de poda que evita o problema de *overfitting* por meio da geração de uma hipótese mais genérica a partir do conjunto de treinamento.

7 CONCLUSÕES

A integração entre dados ópticos (MSI/Sentinel-2) e dados de radar (Sentinel-1) para o mapeamento de uso e cobertura da Terra por classificação supervisionada em parte do Bioma Cerrado usando o algoritmo RF pela plataforma GEE permitiu concluir:

- A integração entre dados ópticos e SAR obteve resultados melhores do que o uso individual desses dados na discriminação da vegetação, áreas de agricultura e pastagem e também nas áreas urbanas (classe que apresentou confusão pelo mistura espectral de diversos alvos) e consequentemente essa integração se mostra capaz de melhor mapear o bioma Cerrado e ser utilizada para o monitoramento da cobertura e uso da terra, monitoramento de queimadas, desmatamentos e monitoramento agrícola;
- A combinação de dados multitemporais e multissensores para mapear regiões heterogêneas, como o Cerrado, sugere que em áreas ou períodos com alta cobertura de nuvens onde a coleta de dados ópticos não é viável, o uso dados com diferentes características espectrais, incluindo dados de radar e ópticos tem potencial para um mapeamento confiável e contínuo;
- A banda 12 (infravemelho de ondas curtas), a polarização VH, os índices CAI e o Hall Cover foram as variáveis mais importantes na separação das classes propostas de uso e cobertura neste trabalho;
- As métricas estatísticas de valor mínimo, mediana e desvio padrão foram as mais importantes entre as bandas, polarizações e índices espectrais utilizados e puderam capturar a sazonalidade espectral dos alvos ao longo do ano permitindo sua separabilidade de classes;
- A utilização do Google Earth Engine para todo o processo, desde a aquisição das imagens até a classificação, foi satisfatório e eficaz pelo fato de diminuir o tempo e o custo computacional para tal execução. Porém, a implementação exige conhecimento prévio de lógicas de programação e funcionamento da plataforma, porém o algoritmo pode ser utilizado mudando poucas linhas de código, tornando mais eficiente que o modo convencional de classificação feito em softwares de processamento de imagem. Combinando o processamento em nuvem com o uso de algoritmos *Machine Learning*, possibilitou o uso de um conjunto de imagens disponíveis durante o ano de 2019 de ambos os sensores de modo acessível e ágil.

8 RECOMENDAÇÕES

O uso de multissensores, bem como de dados geográficos auxiliares, podem auxiliar na melhoria da classificação de imagens para uso e cobertura da terra. Neste trabalho foi avaliado apenas a integração entre dados ópticos do sensor MSI/Sentinel-2 e do sensor SAR do Sentinel-1.

Como recomendação, sugere-se o uso de modelos digitais de elevação, bem como suas derivações geomorfométricas (declividade, curvaturas vertical e horizontal, orientação de vertente, altura em relação à drenagem, entre outros) e dados auxiliares, como a base de arruamentos do projeto *Open Street Map*, setores censitários (urbano e rural) e dados orbitais de luzes noturnas como variáveis potenciais para o incremento da exatidão de classificações para uso e cobertura do solo. Adicionalmente, o aumento do número de amostras de treinamento, a definição de subclasses de uso e cobertura, como para diferenciar culturas agrícolas, seria um desafio para ser melhor explorado na integração de dados para o mapeamento de uso e cobertura da terra.

REFERÊNCIAS BIBLIOGRÁFICAS

ARONOFF, Stan. The minimum accuracy value as an index of classification accuracy. **Photogrammetric Engineering and Remote Sensing**, v. 51, n. 1, p. 99–111, 1985. 28

ARONOFF, Stan *et al.* Classification accuracy: a user approach. **Photogrammetric Engineering and Remote Sensing**, v. 48, n. 8, p. 1299–1307, 1982. 28

BANKO, Gebhard. A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory. IR-98-081, 1998. 30

BASTOS, Denise Gasques; NASCIMENTO, Patricia Santos; LAURETTO, Marcelo de Souza. Análise empírica de desempenho de quatro métodos de seleção de características para random forests. **iSys-Revista Brasileira de Sistemas de Informação**, v. 7, n. 2, p. 25–47, 2014. 27

BORGES, Joana; HIGGINBOTTOM, Thomas P; SYMEONAKIS, Elias; JONES, Martin. Sentinel-1 and sentinel-2 data for savannah land cover mapping: Optimising the combination of sensors and seasons. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 12, n. 23, p. 3862, 2020. 58

BRASIL, M. Mapeamento do uso e cobertura do cerrado: Projeto terra class cerrado 2013. **Ministério do Meio Ambiente (MMA), Brasília**, v. 67, 2015. 17

BREIMAN, Leo. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001. 25, 27, 37, 38

CARNEIRO, A. F.; OLIVEIRA, W. V.; SANT'ANNA, S. J. S.; DOBLAS, J.; VAZ, D. V. Exploiting sentinel-1 sar time series to detect grasslands in the northern brazilian amazon. **The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, Copernicus GmbH, v. 43, p. 259–265, 2020. 21

CHATZIANTONIOU, Andromachi; PSOMIADIS, Emmanouil; PETROPOULOS, George P. Co-orbital sentinel 1 and 2 for lulc mapping with emphasis on wetlands in a mediterranean setting based on machine learning. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 9, n. 12, p. 1259, 2017. 58

COHEN, Jacob. A coefficient of agreement for nominal scales. **Educational and psychological measurement**, Sage Publications Sage CA: Thousand Oaks, CA, v. 20, n. 1, p. 37–46, 1960. 30

CONGALTON, RG; GREEN, K. Assessing the accuracy of remotely sensed data: Principles and practices. Lewis Publishers, Boca Raton, FL (United States), 1999. 28

CONGALTON, Russell G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote sensing of environment**, Elsevier, v. 37, n. 1, p. 35–46, 1991. 35

CROSTA, Alvaro Penteado. **Processamento digital de imagens de sensoriamento remoto**. [S.l.]: UNICAMP/Instituto de Geociências, 1999. 24

DEERING, Donald Wayne. **Rangeland reflectance characteristics measured by aircraft and spacecraft sensors**. Tese (Doutorado) — Texas A&M University. Libraries, 1978. 36

DINIZ, Juliana Maria Ferreira de Souza. **Avaliação do potencial dos dados polarimétricos Sentinel-1A para mapeamento do uso e cobertura da terra na região de Ariquemes - RO**. 132 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2019-02-15 2019. Disponível em: `http://urlib.net/rep/8JMKD3MGP3W34R/3SL65N8`. Acesso em: 26 nov. 2020. 21, 25, 26

DRUSCH, Matthias; BELLO, Umberto Del; CARLIER, Sébastien; COLIN, Olivier; FERNANDEZ, Veronica; GASCON, Ferran; HOERSCH, Bianca; ISOLA, Claudia; LABERINTI, Paolo; MARTIMORT, Philippe *et al.* Sentinel-2: Esa's optical high-resolution mission for gmes operational services. **Remote sensing of Environment**, Elsevier, v. 120, p. 25–36, 2012. 24

EUROPEAN SPACE AGENCY (ESA). **Sentinel-1**. 2020. Disponível em: `https://sentinel.esa.int/web/sentinel/missions/sentinel-1`. Acesso em: 03 set 2020. 21, 22

FOODY, Giles M. Status of land cover classification accuracy assessment. **Remote sensing of environment**, Elsevier, v. 80, n. 1, p. 185–201, 2002. 28

GAMON, JAm; SERRANO, L; SURFUS, JS. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. **Oecologia**, Springer, v. 112, n. 4, p. 492–501, 1997. 36

GAO, Bo-Cai. Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. **Remote sensing of environment**, Elsevier, v. 58, n. 3, p. 257–266, 1996. 36

GHORBANIAN, Arsalan; KAKOOEI, Mohammad; AMANI, Meisam; MAHDAVI, Sahel; MOHAMMADZADEH, Ali; HASANLOU, Mahdi. Improved land cover map of iran using sentinel imagery within google earth engine and a novel automatic workflow for land cover classification using migrated training samples. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 167, p. 276–288, 2020. ISSN 0924-2716. Disponível em: https://www.sciencedirect.com/science/article/pii/S0924271620302008. 17, 31

GITELSON, Anatoly A; MERZLYAK, Mark N. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. **Journal of plant physiology**, Elsevier, v. 148, n. 3-4, p. 494–500, 1996. 36

GITELSON, Anatoly A; VIÑA, Andrés; ARKEBAUER, Timothy J; RUNDQUIST, Donald C; KEYDAN, Galina; LEAVITT, Bryan. Remote estimation of leaf area index and green leaf biomass in maize canopies. **Geophysical research letters**, Wiley Online Library, v. 30, n. 5, 2003. 36

GORELICK, Noel; HANCHER, Matt; DIXON, Mike; ILYUSHCHENKO, Simon; THAU, David; MOORE, Rebecca. Google earth engine: Planetary-scale geospatial analysis for everyone. **Remote sensing of Environment**, Elsevier, v. 202, p. 18–27, 2017. 30, 31, 33

GRAOSQUE, Jones Zamboni. Mapeamento das áreas de inundação utilizando imagens c–sar e srtm, nas províncias de santa fé e entre ríos, argentina. 2018. 21

HALL, R.J.; SKAKUN, R.S.; ARSENAULT, E.J.; CASE, B.S. Modeling forest stand structure attributes using landsat etm+ data: Application to mapping of aboveground biomass and stand volume. **Forest Ecology and Management**, v. 225, n. 1, p. 378–390, 2006. ISSN 0378-1127. Disponível em: https://www.sciencedirect.com/science/article/pii/S0378112706000235. 36, 59

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. Data mining concepts and techniques third edition. **The Morgan Kaufmann Series in Data Management Systems**, v. 5, n. 4, p. 83–124, 2011. 25

HANSEN, Matthew C; POTAPOV, Peter V; MOORE, Rebecca; HANCHER, Matt; TURUBANOVA, Svetlana A; TYUKAVINA, Alexandra; THAU, David; STEHMAN, SV; GOETZ, Scott J; LOVELAND, Thomas R *et al.* High-resolution global maps of 21st-century forest cover change. **Science**, American Association for the Advancement of Science, v. 342, n. 6160, p. 850–853, 2013. 59

HELLDÉN, U. A test of landsat-2 imagery and digital data for thematic mapping illustrated by an environmental study in northern kenya, lund university. **Natural Geography Institute Report No. 47**, 1980. 29

HUETE, Alfredo R. A soil-adjusted vegetation index (savi). **Remote sensing of environment**, Elsevier, v. 25, n. 3, p. 295–309, 1988. 36

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Mapas de biomas e vegetação**. 2004. 16

JIANG, Zhangyan; HUETE, Alfredo R; DIDAN, Kamel; MIURA, Tomoaki. Development of a two-band enhanced vegetation index without a blue band. **Remote sensing of Environment**, Elsevier, v. 112, n. 10, p. 3833–3845, 2008. 36

JR, Robert Gilmore Pontius; MILLONES, Marco. Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, Taylor & Francis, v. 32, n. 15, p. 4407–4429, 2011. 28

KUMAR, Lalit; MUTANGA, Onisimo. Google earth engine applications since inception: Usage, trends, and potential. **Remote Sensing**, v. 10, n. 10, p. 1509, 2018. 59

LAMBIN, Eric F; GEIST, Helmut J. **Land-use and land-cover change: local processes and global impacts**. [S.l.]: Springer Science & Business Media, 2008. 19

LEE, Jong-Sen; WEN, Jen-Hung; AINSWORTH, Thomas L; CHEN, Kun-Shan; CHEN, Abel J. Improved sigma filter for speckle filtering of sar imagery. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE, v. 47, n. 1, p. 202–213, 2008. 34

LOPES, Mailys; FRISON, Pierre-Louis; DURANT, Sarah M; BÜHNE, Henrike Schulte to; IPAVEC, Audrey; LAPEYRE, Vincent; PETTORELLI, Nathalie. Combining optical and radar satellite image time series to map natural vegetation: savannas as an example. **Remote Sensing in Ecology and Conservation**, Wiley Online Library, v. 6, n. 3, p. 316–326, 2020. 17, 58

MACHADO, Gustavo; MENDOZA, Mariana Recamonde; CORBELLINI, Luis Gustavo. What variables are important in predicting bovine viral diarrhea virus? a random forest approach. **Veterinary research**, Springer, v. 46, n. 1, p. 1–15, 2015. 26

MAGIDI, James; NHAMO, Luxon; MPANDELI, Sylvester; MABHAUDHI, Tafadzwanashe. Application of the random forest classifier to map irrigated areas using google earth engine. **Remote Sensing**, v. 13, n. 5, 2021. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/13/5/876. 60

MAPBIOMAS, Projeto. Projeto mapbiomas: Coleção 5.0 da série anual de mapas de cobertura e uso de solo do brasil. **Retrieved from Projeto MapBiomas: Coleção**, v. 5, 2019. 11, 33, 35, 39, 59

MENDES, Flávia de Souza; BARON, Daniel; GEROLD, Gerhard; LIESENBERG, Veraldo; ERASMI, Stefan. Optical and sar remote sensing synergism for mapping vegetation types in the endangered cerrado/amazon ecotone of nova mutum—mato grosso. **Remote Sensing**, Multidisciplinary Digital Publishing Institute, v. 11, n. 10, p. 1161, 2019. 58

MERCIER, Audrey; BETBEDER, Julie; RUMIANO, Florent; BAUDRY, Jacques; GOND, Valéry; BLANC, Lilian; BOURGOIN, Clément; CORNU, Guillaume; CIUDAD, Carlos; MARCHAMALO, Miguel; POCCARD-CHAPUIS, René; HUBERT-MOY, Laurence. Evaluation of sentinel-1 and 2 time series for land cover classification of forest–agriculture mosaics in temperate and tropical landscapes. **Remote Sensing**, v. 11, n. 8, 2019. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/11/8/979. 59

NAGLER, Pamela L; INOUE, Y; GLENN, EP; RUSS, AL; DAUGHTRY, CST. Cellulose absorption index (cai) to quantify mixed soil–plant litter scenes. **Remote Sensing of Environment**, Elsevier, v. 87, n. 2-3, p. 310–325, 2003. 36

NOVO, Evlyn ML de Moraes. **Sensoriamento Remoto: princípios e aplicações**. [S.l.]: Editora Blucher, 2010. 35, 59

NUNES, Juliana Faria; ROIG, Henrique Llacer. Análise e mapeamento do uso e ocupação do solo da bacia do alto do descoberto, df/go, por meio de classificação automática baseada em regras e lógica nebulosa. **Revista árvore**, SciELO Brasil, v. 39, n. 1, p. 25–36, 2015. 16

OLIVEIRA, C. P. **Modelagem dinâmica da cobertura florestal do Município de Floresta–PE. 2015. 68 f**. Dissertação (Mestrado) — Dissertação de Mestrado, Universidade Federal Rural de Pernambuco, Recife, 2015. 19

OLOFSSON, Pontus; FOODY, Giles M; HEROLD, Martin; STEHMAN, Stephen V; WOODCOCK, Curtis E; WULDER, Michael A. Good practices for estimating area and

assessing accuracy of land change. **Remote Sensing of Environment**, Elsevier, v. 148, p. 42–57, 2014. 28

ONODA, M. **Estudo sobre um algoritmo de árvores de decisão acoplado a um sistema de banco de dados relacional. 2001. 110p**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001. 27

OSHIRO, Thais Mayumi. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para classificação de bases de expressão gênica**. Tese (Doutorado) — Universidade de São Paulo, 2013. 60

PASSO, Denilson Pereira. **Análise da qualidade de classificadores para identificação de alvos urbanos em imagens de alta resolução espacial: uma aplicação com as imagens do satélite Worldview II**. Dissertação (Mestrado) — Universidade de Brasília (UnB), Brasília, 2013. Disponível em: `https://repositorio.unb.br/handle/10482/13435`. Acesso em: 26 nov. 2020. 24

PEREIRA, L. O. Freitas C. C.; S., Sant'Anna S. J.; S., Reis M. ALOS/PALSAR data evaluation for land use and land cover mapping in the amazon region. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 12, p. 5413–5423, 2016. 17

PONZONI, Flávio Jorge; SHIMABUKURO, Yosio Edemir; KUPLICH, Tatiana Mora. **Sensoriamento remoto da vegetação**. [S.l.]: Oficina de Textos, 2015. 19, 20, 58, 59

REDDY, M Anji; REDDY, Anji. **Textbook of remote sensing and geographical information systems**. [S.l.]: BS publications Hyderabad, 2008. 25

RIBEIRO, Selma Regina Aranha; CENTENO, Jorge Silva. Classificação do uso do solo utilizando redes neurais e o algoritmo maxver. *In*: KRUG, Thelma; FONSECA, Leila Maria Garcia (Ed.). **Anais...** São José dos Campos: INPE, 2001. p. 1341–1348. ISBN 85-17-00016-1. Disponível em: `http://urlib.net/rep/dpi.inpe.br/lise/2001/09.20.17.56`. Acesso em: 19 nov. 2020. 16

RICHARDS, John A; JIA, Xiuping. Remote sensing digital image analysis. **Remote Sensing Digital Image Analysis**, p. 197, 2006. 27

RODRIGUEZ-GALIANO, Victor Francisco; GHIMIRE, Bardan; ROGAN, John; CHICA-OLMO, Mario; RIGOL-SANCHEZ, Juan Pedro. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 67, p. 93–104, 2012. 26

RONDEAU-GENESSE, Gabriel; TRUDEL, Mélanie; LECONTE, Robert. Monitoring snow wetness in an alpine basin using combined c-band sar and modis data. **Remote Sensing of Environment**, Elsevier, v. 183, p. 304–317, 2016. 21

SANO, Edson E; JESUS, Eristelma T; BEZERRA, Heleno S. Mapeamento e quantificação de áreas remanescentes do cerrado através de um sistema de informações geográficas. **Sociedade & Natureza**, v. 13, n. 25, 2001. 17

SANO, Edson E; ROSA, Roberto; BRITO, Jorge LS; FERREIRA, Laerte G. Land cover mapping of the tropical savanna region in brazil. **Environmental monitoring and assessment**, Springer, v. 166, n. 1, p. 113–124, 2010. 58

SANT'ANNA, SJS; FERNANDES, D; LACAVA, JC da S. Análise comparativa do desempenho de filtros redutores de "speckle" em imagens sar. **Simpósio Brasileiro de Telecomunicações (SBT)**, v. 9, 2001. 34

SARAIVA, Cátia Sofia Alves. **Avaliação o Potencial das Imagens Sentinel-1 para Identificação de Culturas Agrícolas**. 95 p. Dissertação (Mestrado) — Universidade de Lisboa, Lisboa, 2015. Disponível em: `https://repositorio.ul.pt/handle/10451/20682`. Acesso em: 26 nov. 2020. 22, 23, 25, 59

SENTINEL, ESA. User handbook. **ESA Standard Document**, v. 64, 2015. 23

SHELESTOV, Andrii; LAVRENIUK, Mykola; KUSSUL, Nataliia; NOVIKOV, Alexei; SKAKUN, Sergii. Exploring google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. **frontiers in Earth Science**, Frontiers, v. 5, p. 17, 2017. 31

SONNENSCHEIN, Ruth; KUEMMERLE, Tobias; UDELHOVEN, Thomas; STELLMES, Marion; HOSTERT, Patrick. Differences in landsat-based trend analyses in drylands due to the choice of vegetation estimate. **Remote sensing of environment**, Elsevier, v. 115, n. 6, p. 1408–1420, 2011. 19

SOUZA, Carlos M; SHIMBO, Julia Z; ROSA, Marcos R; PARENTE, Leandro L; ALENCAR, Ane A; RUDORFF, Bernardo FT; HASENACK, Heinrich; MATSUMOTO, Marcelo; FERREIRA, Laerte G; SOUZA-FILHO, Pedro WM *et al.* Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, v. 12, n. 17, 2020. ISSN 2072-4292. Disponível em: `https://www.mdpi.com/2072-4292/12/17/2735`. 17, 35

STORY, Michael; CONGALTON, Russell G. Accuracy assessment: a user's perspective. **Photogrammetric Engineering and remote sensing**, Falls Church, v. 52, n. 3, p. 397–399, 1986. 29, 30

SYMEONAKIS, Elias; HIGGINBOTTOM, Thomas P.; PETROULAKI, Kyriaki; RABE, Andreas. Optimisation of savannah land cover characterisation with optical and sar data. **Remote Sensing**, v. 10, n. 4, 2018. Disponível em: https://www.mdpi.com/2072-4292/10/4/499. 17

TUNG, F; LEDREW, E. The determination of optimal threshold levels for change detection using various accuracy indexes. **Photogrammetric Engineering and Remote Sensing**, AMER SOC PHOTOGRAMMETRY 5410 GROSVENOR LANE SUITE 210, BETHESDA, MD 20814-2160, v. 54, n. 10, p. 1449–1454, 1988. 29

ZHA, Yong; GAO, Jay; NI, Shaoxiang. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. **International journal of remote sensing**, Taylor & Francis, v. 24, n. 3, p. 583–594, 2003. 36

ZHANG, Wenmin; BRANDT, Martin; WANG, Qiao; PRISHCHEPOV, Alexander V; TUCKER, Compton J; LI, Yunmei; LYU, Heng; FENSHOLT, Rasmus. From woody cover to woody canopies: How sentinel-1 and sentinel-2 data advance the mapping of woody plants in savannas. **Remote Sensing of Environment**, Elsevier, v. 234, p. 111465, 2019. 17, 58