

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE TECNOLOGIAS E RECURSOS NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL E AMBIENTAL
ÁREA DE CONCENTRAÇÃO: RECURSOS HÍDRICOS E SANEAMENTO

JOÃO MARIA DE ANDRADE

**CLASSIFICAÇÃO DA COBERTURA DO SOLO DA CAATINGA: AVALIAÇÃO DE
ABORDAGENS EM MACHINE LEARNING UTILIZANDO DADOS LANDSAT
MONO E MULTI-TEMPORAIS**

CAMPINA GRANDE - PB
FEVEREIRO DE 2021

JOÃO MARIA DE ANDRADE

**CLASSIFICAÇÃO DA COBERTURA DO SOLO DA CAATINGA: AVALIAÇÃO DE
ABORDAGENS EM MACHINE LEARNING UTILIZANDO DADOS LANDSAT
MONO E MULTI-TEMPORAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil e Ambiental (PPGECA) da Universidade Federal de Campina Grande (UFCG), para a obtenção do título de Mestre em Engenharia Civil e Ambiental. Área de concentração: Recursos Hídricos e Saneamento.

Orientadores: Prof. Dr. John Elton de Brito Leite Cunha
Prof.^a Dr.^a Iana Alexandra Alves Rufino

CAMPINA GRANDE - PB
FEVEREIRO DE 2021

A553c

Andrade, João Maria de.

Classificação da cobertura do solo da caatinga : avaliação de abordagens em Machine Learning utilizando dados landsat mono e multi- temporais / João Maria de Andrade. - Campina Grande, 2021.

69 f. : il. Color

Dissertação (Mestrado em Engenharia Civil e Ambiental) - Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, 2021.

"Orientação: Prof. Dr. John Elton de Brito Leite Cunha, Profa. Dra. Iana Alexandra Alves Rufino".

Referências.

1. Sensoriamento Remoto. 2. Random-Forest. 3. Semiárido. 4. Florestas Tropicais Sazonais Secas. 5. NDVI. 6. Multiespectral. I. Cunha, John Elton de Brito Leite. II. Rufino, Iana Alexandra Alves. III. Título.

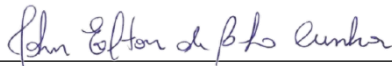
CDU 631.4(043)

JOÃO MARIA DE ANDRADE

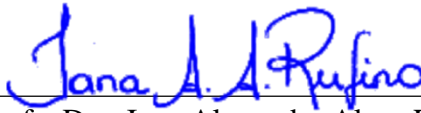
**CLASSIFICAÇÃO DA COBERTURA DO SOLO DA CAATINGA: AVALIAÇÃO DE
ABORDAGENS EM MACHINE LEARNING UTILIZANDO DADOS LANDSAT
MONO E MULTI-TEMPORAIS**

Dissertação aprovada em 11 de Fevereiro de 2021.

BANCA EXAMINADORA



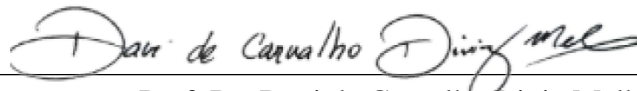
Prof. Dr. John Elton de Brito Leite Cunha
(Orientador - Universidade Federal de Campina Grande – UFCG)



Profa. Dra. Iana Alexandra Alves Rufino
(Coorientadora - Universidade Federal de Campina Grande – UFCG)



Prof. Dr. Carlos de Oliveira Galvão
(Examinador interno - Universidade Federal de Campina Grande – UFCG)



Prof. Dr. Davi de Carvalho Diniz Melho
(Examinador externo - Universidade Federal da Paraíba - UFPB)

CAMPINA GRANDE - PB

FEVEREIRO DE 2021

Dedico este trabalho aos meus pais, Manoel e Benedita, por todo o amor que sempre me deram.

AGRADECIMENTO

Agradeço a Deus, primeiramente, por me conduzir com saúde, esperança e resiliência ao longo desta jornada.

Aos meus pais, Benedita e Manoel, por toda a força e perseverança para criar os seus filhos e por todo o apoio e carinho. Às minhas irmãs, Polianna e Moanna, pelo cuidado e atenção.

À minha querida namorada Carol, pelo companheirismo em todos os momentos e à sua família por todo o acolhimento, amor, incentivo e paciência.

Ao meu orientador John pelos ensinamentos, parceria e apoio ao longo de toda a minha trajetória acadêmica; à professora Iana Rufino pelo incentivo e orientação; e aos professores Carlos Galvão e Rodolfo Nóbrega que contribuíram de forma significativa com minha formação como pesquisador.

Aos meus amigos e companheiros do laboratório de Hidráulica 2 (Lab-Hid2), por todas as conversas, reflexões e risos, em especial a Ulisses e Benits/Benito (*in memoriam*).

À Universidade Federal de Campina Grande pelo espaço físico e recursos disponíveis para o desenvolvimento da minha pesquisa e à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro dedicado à minha formação.

A todos os cientistas, pesquisadores e professores que sempre lutaram e que persistem todos os dias para iluminar a humanidade no meio de todo o obscurantismo que nos cerca.

“O sentido da vida é viver em busca de sentido. É no ato da busca, na experiência do novo e do inesperado, que damos sentido à nossa existência. “

A simples beleza do inesperado, Marcelo Gleiser

RESUMO

Informações precisas sobre a cobertura do solo são fundamentais para o monitoramento e desenvolvimento eficientes de estudos ambientais na Caatinga brasileira. É uma das maiores e mais biodiversas florestas secas do planeta. Distinguir diferentes padrões de cobertura do solo por meio de sensoriamento remoto de resolução espacial média, como a série de imagens Landsat, é um desafio para a Caatinga devido à cobertura heterogênea do solo, complexas interações clima-solo-vegetação e distúrbios antrópicos. Duas abordagens de sensoriamento remoto têm alto potencial para o mapeamento preciso e eficiente da cobertura do solo na Caatinga: imagens mono e multi-temporais. A heterogeneidade da cobertura do solo desse ambiente pode contribuir para um melhor desempenho de abordagens multiespectrais que são normalmente aplicadas para imagens mono-temporais. Em um esforço de mapeamento da cobertura do solo na Caatinga, o fator temporal ganha relevância, e o uso de séries temporais pode trazer vantagens. Este estudo avalia as precisões e vantagens das abordagens multi-temporal do *Normalized Difference Vegetation Index* (NDVI) e mono-temporal multiespectral na classificação da cobertura do solo. Ambas as abordagens usam o método *Random Forest*, e os resultados são avaliados com base em amostras coletadas durante levantamentos de campo. Os resultados indicam que a classificação da cobertura do solo obtida a partir de multi-temporal NDVI tem melhor desempenho (acurácia geral de 88,8% e coeficiente de concordância kappa de 0,86) do que dados mono-temporal multiespectrais (acurácia geral de 81,4% e coeficiente kappa de 0,78). O teste Z indicou que uma diferença estatisticamente significativa no desempenho entre as duas abordagens. O desempenho inferior observado para a classificação multiespectral mono-temporal deve-se às semelhanças nas respostas espectrais para alvos de vegetação decídua que perdem sua folhagem e podem ser interpretados incorretamente como áreas sem vegetação. Enquanto isso, a classificação multi-temporal NDVI dos agrupamentos de vegetais nas florestas sazonais permite incorporar a variabilidade sazonal das classes de cobertura do solo durante as estações chuvosa e seca, bem como as transições entre as estações, por isso apresenta maior precisão. Além disso, as variáveis mais importantes que contribuíram para a precisão foram as bandas *Red*, *Near Infrared* (NIR) e *Short-Wave Infrared* (SWIR) na classificação multiespectral mono-temporal, e os meses na estação seca foram os mais relevantes para classificação multi-temporal NDVI.

Palavras-chave: Sensoriamento remoto; Random-Forest; Semiárido, Florestas Tropicais Sazonais Secas, NDVI, multiespectral.

ABSTRACT

Accurate information on the land cover is crucial for efficient monitoring and development of environmental studies in the Brazilian Caatinga forest, one of the largest and most biodiverse dry forests on the planet. Distinguishing different patterns of land cover through medium spatial-resolution remote sensing, such as the Landsat image series, is challenging to Caatinga due to heterogeneous land cover, complex climate-soil - vegetation interactions, and anthropogenic disturbance. Two remote sensing approaches have a high potential for accurate and efficient land-cover mapping in Caatinga: single and multi-date imagery. The heterogeneity of the land cover of this environment can contribute to a better performance of multispectral approaches that are usually applied for single-date images. In a land-cover mapping effort in Caatinga, the temporal factor gains relevance, and the use of time series can bring advantages, but, in general, this approach uses vegetation index, losing multispectral information. This manuscript assesses the accuracies and advantages of single-date multispectral and multi-date Normalized Difference Vegetation Index (NDVI) approaches in land-cover classification. Both approaches use the Random Forest method, and the results are evaluated based on samples collected during field surveys. Results indicate that land-cover classification obtained from multi-date NDVI performs better (overall accuracy of 88.8% and kappa of 0.86) than single-date multispectral data (overall accuracy of 81.4% and kappa coefficient of 0.78). The Z-test indicated that the difference in performance between the two approaches was statistically significant. The lower performance observed for single-date multispectral classification is due to similarities in spectral responses for targets of deciduous vegetation that lose their foliage and can be misread as non-vegetated areas. Meanwhile, an accurate classification by time series of plant clusters in seasonal forests allows incorporating seasonal variability of land-cover classes during the rainy and dry seasons, as well as transitions between seasons. The most important variables that contributed to the accuracy were the Red, Near Infrared (NIR) and Short-Wave Infrared (SWIR) bands in single-date multispectral classification, and the months in the dry season were the most relevant in multi-date NDVI classification.

Keywords: Remote sensing; Random-Forest; Semi-arid, Seasonal Dry Tropical Forests, NDVI, multispectral.

LISTA DE FIGURAS

Figura 1: Relação entre dois conjuntos de dados utilizados pelas técnicas de ML com a finalidade de realizar previsão ou classificação.....	19
Figura 2: Ciclo de desenvolvimento dos modelos de ML a partir do avanço computacional, métodos estatísticos e disponibilidade de dados. Fonte: Lantz (2013).....	20
Figura 3: Modelo de classificação no Random Forest. A) Cada árvore de decisão no conjunto é construída sobre uma amostra de bootstrap aleatória dos dados originais, que contém exemplos positivos (rótulos verdes) e negativos (rótulos vermelhos). B) Classe prevista para novas instâncias usando um modelo de floresta aleatório é baseada em um procedimento de votação por maioria entre todas as árvores individuais. Fonte: Machado et. (2015).....	22
Figura 4: Descrição do gráfico boxplot com resumo de seis informações de um conjunto de dados (potências outlines, mínimo, quartil inferior, mediana, quartil superior e máximo). Fonte: Siervi (2019).	25
Figura 5: Interação entre superfície e radiação eletromagnética. Fonte: Esri (2020).	26
Figura 6: Linha do tempo das missões Landsat, Fonte: USGS (2020).	27
Figura 7: Mapa esquemático global de savanas tropicais e florestas secas. Fonte: Pennington et al. (2018).	32
Figura 8: Fisionomias da Caatinga: A) Arbórea, B) Subarbórea, C) Arbustivo, D) Subarbustivo.	33
Figura 9: Localização da área de estudo na Caatinga. Os padrões verdes na imagem são as localizações dos dados de referência do solo.....	35
Figura 10: Climograma com base em dados climáticos da estação Monteiro de 1987 a 2016. Fonte: INMET (2020), código da estação # 82792. A) Vegetação de Caatinga na estação chuvosa, B) Vegetação de Caatinga na estação seca.	36
Figura 11: Fluxograma das etapas metodológicas.	36
Figura 12: Abordagens de classificação: Mono-temporal multiespectral e multi-temporal NDVI.....	39
Figura 13: Gráfico de radar que representa as precisões do usuário e do produtor para a classificação NDVI multiespectral e de data única.....	42
Figura 14: Gráficos de boxplot das amostras de treinamento para classes de cobertura da terra A) Bandas espectrais 2 (azul), 3 (verde), 4 (vermelha), 5 (NIR), 6 (SWIR 1), 7 (SWIR 2) da imagem Landsat 8 de 29 de setembro de 2016; B) 24 Composições Mensais de NDVI. Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva	

Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Solo Exposto (SE).....	43
Figura 15: A) Classificação geral de importância variável usando o MDG: A) Classificação multiespectral de dados única, bandas espectrais 2 (azul), 3 (verde), 4 (vermelha), 5 (NIR), 6 (SWIR 1), 7 (SWIR 2) da imagem Landsat 8; B) Classificação Multi-temporal NDVI.	44
Figura 16: Área (km ²) derivada das classificações mono-temporal multiespectral e multi-temporal NDVI.	45
Figura 17: Mapas de cobertura do solo classificados por: A) Mono-temporal multiespectral e B) Multi-temporal NDVI.	46
Figura 18: Refletância média da amostra de treinamento para classificação mono-temporal multiespectral.	48
Figura 19: Valores médios de NDVI para cada classe da amostra de treinamento da classificação Multi-temporal NDVI.	49

LISTA DE TABELAS

Tabela 1: Exemplo da Matriz de Confusão.....	23
Tabela 2: Principais contribuições das faixas espectrais do Landsat 8 para o mapeamento. ..	28
Tabela 3: Classificação da vegetação da Caatinga.	37
Tabela 4: Conjuntos de dados de treinamento e teste usados no estudo.....	38
Tabela 5: Matriz de Confusão (%) da classificação multi-temporal de (NDVI).	41
Tabela 6: Matriz de Confusão (%) para classificação mono-temporal multiespectral.	42

LISTA DE ABREVIATURAS E SIGLAS

AG	–	Acurácia Geral
ASD	–	Arbórea Subarbórea Densa
ASMD	–	Arbórea Subarbórea Muito Densa
ATSA	–	Arbustiva Subarbórea Aberta
ETM+	–	Enhanced Thematic Mapper Plus
EVI	–	Enhanced Vegetation Index
FTSS	–	Floresta Tropical Sazonal Seca
IAF	–	Índice de Área Foliar
INMET	–	Instituto Nacional de Meteorologia
INSA	–	Instituto Nacional do Semiárido
IV	–	Índice de Vegetação
k	–	Kappa
MDA	–	Mean Decrease Accuracy
MDG	–	Mean Decrease Gini
ML	–	Machine Learning
MMA	–	Ministério do Meio Ambiente
MSS	–	Multispectral Scanner System
NDVI	–	NDVI - Normalized Difference Vegetation Index
OLI	–	Operational Land Imager
OOB	–	Out-of-Bag
PP	–	Precisão do Produtor
PU	–	Precisão do Usuário
RF	–	Random Forest
SAA	–	Subarbórea Arbustiva Aberta
SAR	–	Subarbustiva Arbustiva Rala
SAVI	–	Soil Adjusted Vegetation Index
SE	–	Solo Exposto
SR	–	Sensoriamento Remoto
SVM	–	Support Vector Machine
TIR	–	Thermal Infrared Sensor
TM	–	Thematic Mapper (TM)
USGS	–	United States Geological Survey

SUMÁRIO

1. INTRODUÇÃO.....	16
1.1. Objetivo.....	18
1.2. Organização da dissertação	18
2. FUNDAMENTAÇÃO TEÓRICA	19
2.1. Machine Learning	19
2.2. Random Forest	21
2.2.1. Bootstrap.....	21
2.2.2. Bagging.....	21
2.2.3. Conceito do classificador Random Forest	21
2.2.4. Importância de cada variável no Random Forest	22
2.3. Métricas de avaliação de desempenho da classificação	23
2.3.1. Comparação entre classificadores (teste-z)	24
2.3.2. Análise gráfica de um conjunto de dados (Boxplot)	25
2.4. Sensoriamento remoto.....	26
2.4.1. Série Landsat	26
2.4.2. Faixas espectrais	28
2.4.3. Índices de vegetação (IVs)	29
2.5. Abordagens de classificação da cobertura vegetal.....	30
2.5.1. Dados multiespectrais mono-temporais.....	31
2.5.2. Dados multitemporais.....	31
2.6. Florestas tropicais sazonais secas.....	32
2.6.1 Caatinga	32
3. METODOLOGIA.....	35
3.1. Área de estudo.....	35
3.2. Métodos.....	36
3.2.1. Coleta de dados.....	37
3.2.2. Imagens de satélite Landsat	38
3.2.4. Classificação	39
3.2.5 Avaliação da precisão	40
4. RESULTADOS	41

4.1. Precisão das classificações	41
4.2. Análise das variáveis de treinamento no modelo de classificação	42
4.3. Distribuição espacial das classes de cobertura da Caatinga	45
5. DISCUSSÕES	47
6. CONCLUSÃO	52
REFERÊNCIAS BIBLIOGRÁFICAS	54

1. INTRODUÇÃO

A Caatinga é uma das maiores florestas tropicais sazonais secas (FTSS) do planeta (Allen et al., 2017; Queiroz et al., 2017; Pennington et al., 2018), cobrindo uma área de cerca de 844,453 km², correspondendo a 11% do território brasileiro (Brasil-MMA, 2019). No entanto, a Caatinga é um dos biomas menos estudados no Brasil quando comparada às florestas tropicais da Amazônia e das savanas do Brasil central (Santos et al., 2011; Gomes et al., 2020). A conservação da Caatinga tem influência direta em diversos processos ambientais associados à proteção do solo, recursos hídricos, manutenção do clima (Manhães et al., 2016) e atividades econômicas (Brasil-MMA, 2019). A degradação da vegetação da Caatinga é resultante da exploração insustentável, que associada a fatores climáticos, acelera o processo de desertificação da região (Drumond, 2004; Ribeiro et al., 2015). Este ecossistema é uma das florestas secas de maior biodiversidade do planeta, com ~ 1700 espécies, das quais pelo menos 300 são endêmicas (Moro et al., 2014; Castanho et al., 2020). A região carece de informações precisas sobre a cobertura do solo para um monitoramento e desenvolvimento eficientes de estudos ambientais (hidrológicos e ecossistêmicos) e estratégias de gestão (Gomez et al., 2016).

A cobertura vegetal da Caatinga é heterogênea, com árvores, arbustos e campos ocorrendo em mosaicos na floresta seca (Silveira et al., 2018). A precipitação pluvial é o principal regulador fenológico das plantas desta floresta (Moro et al., 2016), concentrada sazonalmente em 2 a 4 meses e altamente variável no espaço e no tempo. Os diferentes padrões de cobertura do solo também são impulsionados por outros aspectos naturais, como solos e topografia, além de fatores antrópicos, atuando em múltiplas escalas espaciais e temporais (Chaves et al., 2008; Moro et al., 2016). Nesses diferentes padrões de cobertura do solo, as estratégias de adaptação ao clima são distintas, resultando em diferentes respostas espaciais e na variação de suas propriedades físicas ao longo do tempo (Meiado et al., 2012; Vico et al., 2015). Padrões de cobertura do solo semelhantes ou iguais (espécies e fitossociologia) podem mostrar respostas distintas às mudanças do clima e do solo e também modificar suas respostas espectrais ao longo do tempo (intra e interanual), dependendo da ocorrência de chuvas. Por outro lado, pelo mesmo motivo, os diferentes padrões de cobertura podem apresentar respostas espectrais semelhantes. Conforme discutido acima, tais particularidades - cobertura heterogênea do solo, complexa interações da vegetação clima-solo e distúrbios antropogênicos - tornam um desafio distinguir os diferentes padrões de cobertura da terra por meio de sensoriamento remoto (Erasmí et al., 2014; Mariano et al., 2018; Cunha et al., 2020).

A extração de informações de cobertura terrestre a partir de imagens de sensoriamento remoto (SR) é o resultado da interação dos alvos na superfície e da radiação eletromagnética nas diferentes bandas espectrais (Jensen, 2009). Os algoritmos para distinguir os diferentes padrões espaciais existentes na paisagem aproveitam essas informações para caracterizar a cobertura do solo. Os algoritmos de *machine learning* (ML), tais como o método *Random Forest* (RF), têm sido empregados por pesquisadores e cientistas de SR para mapeamento da cobertura do solo nos últimos anos (Fassnacht et al., 2016; Thanh Noi e Kappas, 2017; Gyamfi-Ampadu et al., 2020; Jin et al., 2018; Silveira et al., 2018; Nguyen et al., 2020). Esses algoritmos são de natureza não paramétrica e são eficientes em termos de tempo de processo e têm a capacidade de produzir alta precisão (Lantz, 2013; Fassnacht et al., 2016), eles aprendem as características do objeto por meio das amostras de treinamento e as usam para identificar as características dos dados não classificados (James et al., 2013; Belgiu e Drăguț, 2016).

A associação das técnicas de ML e SR está em constante evolução em decorrência da maior disponibilidade de informações da cobertura do solo e aumento da capacidade computacional para o processamento dos dados. A estrutura de dados do Landsat permite realizar análises temporais em uma resolução espacial inferior a 100 metros (Woodcock et al., 2020), pois fornece informações sobre a qualidade da cobertura radiométrica, geométrica e identificação de nuvens e sombras de nuvens (Wulder et al., 2016; Man et al., 2018), facilitando a diferenciação dos padrões de cobertura do solo em alta heterogeneidade espacial. Embora os satélites ofereçam monitoramento praticamente contínuo, a classificação da cobertura do solo comumente usa dados multiespectrais em uma única data de observação (mono-temporal) (Jia et al., 2014; Mahdianpari et al., 2018; Alhassan et al., 2019). No entanto, esta abordagem pode induzir confusão na classificação dos diferentes padrões de cobertura do solo existentes nas FTSS, devido à similaridade da resposta espectral da vegetação em estágios fenológicos específicos (Karnieli, 2002).

O uso de séries temporais pode ser uma alternativa para o mapeamento das FTSS, por permitir o monitoramento dos diferentes estágios fenológicos dos padrões de cobertura do solo (Hüttich et al., 2011; Gomez et al., 2016). Além disso, o uso de índices de vegetação (IVs) permite sintetizar as bandas espectrais, que são mais sensíveis à variação da biomassa e aos processos fotossintéticos, simplificando o número de variáveis de entrada (Tatsumi et al., 2015). No entanto, a maioria dos estudos usando séries temporais de índice de vegetação são realizados em áreas agrícolas (Wardlow e Egbert, 2008; Zheng et al., 2015; Mercier et al., 2020), o que facilita a identificação dos padrões de cobertura fenológica. Nas FTSS as classes

de cobertura do solo e seus padrões fenológicos não estão bem definidos, e as mudanças antrópicas podem prejudicar o mapeamento (Abdi, 2020).

1.1. Objetivo

Esse trabalho tem como objetivo principal avaliar o desempenho de diferentes abordagens de classificação para identificação dos padrões complexos da cobertura do solo na Caatinga, uma floresta tropical sazonal seca brasileira, utilizando o classificador *Random Forest*. A seguir, os objetivos específicos desta dissertação são apresentados:

- Comparar o desempenho de uma abordagem que usa dados multi-temporais do *Normalized Difference Vegetation Index* (NDVI) e outra que usa dados mono-temporal multiespectral, e seus impactos no mapeamento da cobertura do solo.
- Identificar as variáveis que mais contribuem para a identificação das classes de cobertura do solo na Caatinga no modelo de classificação.

1.2. Organização da dissertação

Este trabalho está organizado em seis capítulos. No capítulo 2 são descritos os fundamentos teóricos, com conceitos importantes de *machine learning* (ML), métricas para avaliar o desempenho da classificação, classificador *Random Forest* (RF), sensoriamento remoto e características de uma floresta tropical sazonal seca (FTSS), para auxiliar o entendimento dos demais capítulos. A metodologia para desenvolvimento do trabalho está exposta no capítulo 3, com a descrição da área de estudo, aquisição e processamento de dados, método de classificação e avaliação de desempenho. No capítulo 4 são apresentados os resultados e no capítulo 5 a discussão e avaliação deles. Em seguida, no capítulo 6 são apresentados as conclusões.

2. FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo são apresentados alguns conceitos e tópicos importantes para o desenvolvimento do trabalho como *Machine Learning*, estrutura conceitual do classificador *Random Forest* e métricas de desempenho para avaliar os modelos de classificação. Além disso, são abordados tópicos relativos ao sensoriamento remoto, série Landsat, faixas espectrais, índices de vegetação e abordagens para classificação da cobertura do solo. Por fim, descreve-se as características das florestas tropicais sazonais secas (FTSS) e da Caatinga.

2.1. *Machine Learning*

Machine Learning (ML) é um campo da inteligência artificial usado para identificar padrões e realizar previsões, apresentando implementação mais fácil e de menor custo computacional quando comparado com modelos de base física (Mosavi, 2018). O termo ML se refere a técnicas que estabelecem relação entre dois conjuntos de dados. Por exemplo, a relação entre os elementos do conjunto domínio X a elementos do conjunto imagem Y, quando Y é contínuo a função F (modelo de ML) realiza uma previsão, quando Y é discreto, realiza uma classificação (James et al., 2013) (Figura 1).

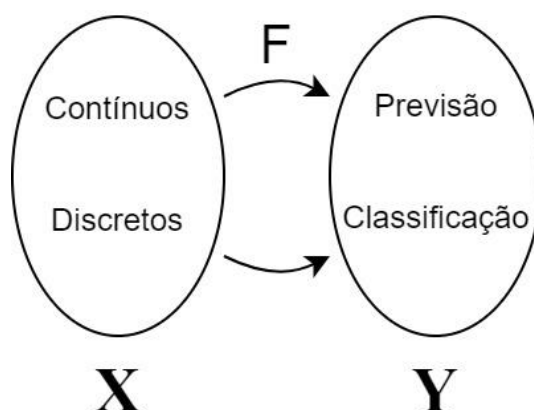


Figura 1: Relação entre dois conjuntos de dados utilizados pelas técnicas de ML com a finalidade de realizar previsão ou classificação.

Este campo é originado em um ambiente onde os dados disponíveis, métodos estatísticos e o poder de computação evoluíram rápida e simultaneamente (Figura 2). O necessário crescimento de dados e poder adicional de computação, estimulou o desenvolvimento de métodos estatísticos para analisar grandes conjuntos de dados, isso criou

um ciclo de avanço permitindo um volume de dados ainda maior e mais interessante a ser coletado (Lantz, 2013).

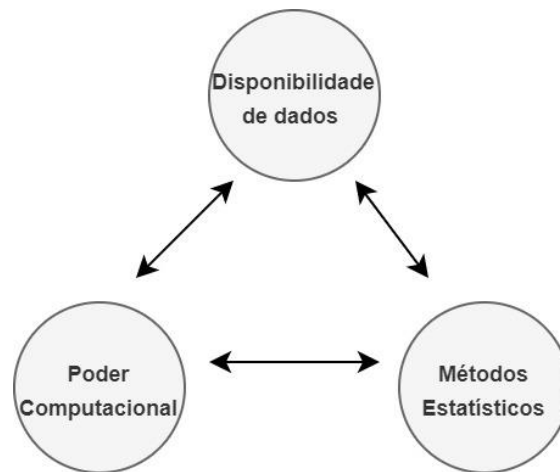


Figura 2: Ciclo de desenvolvimento dos modelos de ML a partir do avanço computacional, métodos estatísticos e disponibilidade de dados. Fonte: Lantz (2013).

A maioria dos problemas de ML recai em uma de duas categorias, supervisionado ou não supervisionado. Por exemplo, na categoria supervisionada existe uma observação da (s) medição (ões) do preditor x_i , $i = 1, \dots, n$, há uma resposta associada y_i , nesse caso a intenção é ajustar um modelo de ML que relacione a resposta aos preditores, com o objetivo de prever com precisão a resposta para novas observações (previsão) ou compreender melhor a relação entre a resposta e os preditores (inferência) (James et al., 2013). No entanto, pode-se buscar entender as relações entre as variáveis ou entre as observações. Uma ferramenta de ML que se pode usar neste cenário é a análise de cluster, ou *clustering* (James et al., 2013). O objetivo da análise de cluster é para determinar, com base em x_1, \dots, x_n , se as observações se aglutinam em grupos relativamente distintos. Por exemplo, em um estudo de segmentação de mercado, pode se utilizar a análise *cluster* para tentar agrupar os clientes a partir dos dados disponíveis que se enquadram em grupos diferentes, como muito gastadores ou pouco gastadores, com o objetivo de identificar grupos distintos de clientes potenciais (James et al., 2013) Existem inúmeros algoritmos de ML utilizados para diferentes finalidades. Por exemplo, para problemas de classificação pode-se utilizar: *Support vector machine (SVM)*, Modelos baseados em árvores (*Random Forest*, *Árvore de decisão*, *XGboost*), Redes Neurais Artificiais. Para problemas de previsão pode-se utilizar também o *Random Forest*, *Regressão Linear* ou *Polinomial*, *Redes Neurais Artificiais*. Já para problemas de clusterização pode-se utilizar *K-means*; *Principal Component Analysis* (James et al., 2013; Kuhn e Johnson, 2013; Lantz, 2013).

2.2. Random Forest

O *Random forest* (RF) possui uma ampla gama de aplicações em vários problemas de classificação e precisão (Zhang et al., 2019). Ele é baseado na construção de árvores aleatória de decisão e utilizada a agregação de *bootstrap* (Bagging) para construção de suas árvores com o objetivo de reduzir a variância do modelo.

2.2.1. Bootstrap

O *bootstrap* é uma ferramenta estatística amplamente aplicável e extremamente poderosa que pode ser usada para quantificar a incerteza associada a um determinado estimador ou método de ML (James et al., 2013). O *bootstrap* permite que amostras sejam selecionadas várias vezes através de um processo de amostragem com reposição (Efron, 1992). Isso significa que de um conjunto de dados com n amostras, o procedimento de *bootstrap* criará um ou mais novos conjuntos de dados que também podem conter até n amostras, alguns dos quais são repetidos (Lantz, 2013).

2.2.2. Bagging

Bagging ou agregação de *bootstrap* gera uma série de conjuntos de dados de treinamento por *bootstrap* de amostragem dos dados de treinamento originais. Esses conjuntos de dados são então usados para gerar um conjunto de modelos usando um único algoritmo de aprendizagem. As previsões dos modelos são combinadas por meio de votação (para classificação) ou média (para previsão numérica) (Lantz, 2013). *Bagging* é um procedimento que tem por objetivo reduzir a variância de um método de ML.

2.2.3. Conceito do classificador *Random Forest*

O RF produz múltiplas árvores de decisões, em que cada árvore vota em uma classe e a decisão de classificação final é a classe que obteve o maior número de votos das árvores, para os problemas de classificação (Belgiu e Drăgut, 2016, Liaw e Wiener, 2002) (Figura 3). O RF usa uma amostra de *bootstrap* de dois terços dos dados de treinamento originais (*In-The-Bag/ITB*) para construir árvores aleatoriamente (n_{tree}), enquanto o terceiro grupo de amostras conhecido como *out-of-bag* (OOB) é usado para obter uma estimativa de erro interno do processo de treinamento (Breiman, 2001; Hüttich et al., 2011; Belgiu e Drăgut, 2016).

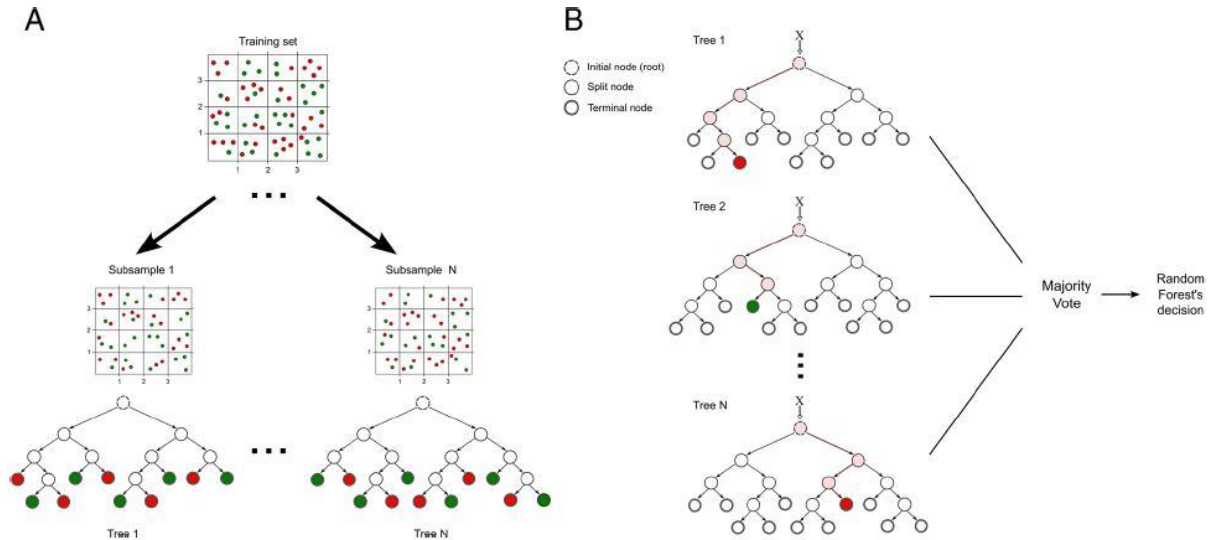


Figura 3: Modelo de classificação no *Random Forest*. A) Cada árvore de decisão no conjunto é construída sobre uma amostra de *bootstrap* aleatória dos dados originais, que contém exemplos positivos (rótulos verdes) e negativos (rótulos vermelhos). B) Classe prevista para novas instâncias usando um modelo de floresta aleatório é baseada em um procedimento de votação por maioria entre todas as árvores individuais. Fonte: Machado et. (2015).

2.2.4. Importância de cada variável no *Random Forest*

No modelo RF, existem dois indicadores que medem a importância das características utilizadas no modelo, *Mean Decrease Accuracy* (MDA) e *Mean Decrease Gini* (MDG) (Corcoran et al., 2013; Nitze et al., 2015). O MDA é baseado no erro interno OOB da árvore (e_t) e no novo erro OOB (e_t^i) é calculado após permutar cada variável preditora X^1 (Equação 1) (Jin et al, 2018). O MDA é a média da diferença dos dois erros de testes OOB. E sua fórmula é a seguinte:

$$V(X^1) = \frac{1}{N} \sum_{t=1}^N (e_t^i - e_t), \text{ Eq.1}$$

O índice Gini é usado para selecionar a melhor divisão em cada nó, é calculado a partir da impureza de cada variável um RF (Breiman, 2001; Jin et al., 2018, 2018; Nitze et al., 2015). A expressão do índice Gini é demonstrada pela Equação 2:

$$Gini(p) = \sum_{k=1}^n p_k (1 - p_k) = 1 - \sum_{k=1}^n p_k^2, \text{ Eq.2}$$

p_k é a probabilidade da k -ésima categoria entre as n categorias. Sua fórmula é a seguinte (Equação 3):

$$J_b(x_j) = Gini(D) - \left(\frac{|D1|}{|D|} Gini(D1) + \frac{|D2|}{|D|} Gini(D2) \right), \text{ Eq.3}$$

Onde D1 e D2 são duas categorias nas quais o conjunto D é dividido de acordo com as características. Nesses dois métodos, quanto maior for o indicador, maior será o grau de importância dos recursos e ambos os métodos de avaliação do grau de importância dos recursos são baseados na qualidade dos resultados da classificação (Jin et al., 2018).

2.3. Métricas de avaliação de desempenho da classificação

A avaliação das classificações supervisionadas pode ser realizada com base em quatro indicadores de desempenho de precisão calculados a partir da matriz de confusão: acurácia geral (AG), coeficiente Kappa (k), precisão do produtor (PP) e precisão do usuário (PU). A matriz de confusão descreve o padrão das classes obtidas pelo modelo de classificação em relação aos dados de referência (Foody, 2002) (Tabela 1).

Tabela 1: Exemplo da Matriz de Confusão.

Classificação	Dados de Referência						Total
	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe m	
Classe 1	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₊
Classe 2	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅	X ₂₆	X ₂₊
Classe 3	X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅	X ₃₆	X ₃₊
Classe 4	X ₄₁	X ₄₂	X ₄₃	X ₄₄	X ₄₅	X ₄₆	X ₄₊
Classe 5	X ₅₁	X ₅₂	X ₅₃	X ₅₄	X ₅₅	X ₅₆	X ₅₊
Classe m	X ₆₁	X ₆₂	X ₆₃	X ₆₄	X ₆₅	X _{mm}	X _{m+}
Total	X ₊₁	X ₊₂	X ₊₃	X ₊₄	X ₊₅	X _{+m}	n

O coeficiente de concordância Kappa (k) utiliza todos os elementos da matriz de confusão em seu cálculo (Equação 4). Valores de k menores que 0 significa nenhuma concordância e próximos de 1 significa concordância perfeita, ele constitui um importante avaliador de precisão na análise de imagens. Onde n é o número total de amostras e m o número total de classes; x_{ii} é o valor da diagonal da matriz de confusão na linha e coluna i ; x_{i+} é a soma da linha i e x_{+i} é a soma da coluna i da matriz:

$$k = \frac{n \sum_{i=1}^m x_{ii} - \sum_{i=1}^m x_{i+} x_{+i}}{n^2 - \sum_{i=1}^m x_{i+} x_{+i}}, \text{ Eq. 4}$$

A acurácia geral (AG) é a divisão do número total de amostras classificadas corretamente (soma dos elementos da diagonal principal da matriz de confusão) pelo número total de amostras de referência (Equação 5).

$$AG = \frac{X_{11} + X_{22} + X_{33} + X_{44} + X_{55} + \dots + X_{mm}}{n}, \text{ Eq. 5}$$

Já PP é a divisão do número total de amostras classificadas corretamente em uma classe pelo número total de amostras de referência para essa classe (Equação 6), enquanto PU é a divisão do número total de amostras que foram classificadas corretamente em uma classe pelo número total de amostras classificados nessa classe (Equação 7) (Congalton, 1991). Esses indicadores de desempenho (AG, PP e PU) têm valores entre 0% e 100% (pior e melhor desempenho, respectivamente).

$$PP = \frac{X_{ii}}{X_{+i}}, \text{ Eq. 6}$$

$$PU = \frac{X_{ii}}{X_{i+}}, \text{ Eq. 7}$$

2.3.1. Comparação entre classificadores (teste-z)

O teste-Z é um teste estatístico baseado na diferença dos valores do índice kappa da matriz de confusão de duas classificações. É um teste amplamente utilizado para comparar a precisão da classificação em sensoriamento remoto (Foody, 2009; Jia et al., 2014) expressa na Equação 8. A diferença nas precisões de duas classificações é estatisticamente significativa se o valor de Z for maior que 1,96 (Congalton e Mead, 1983) ao usar o nível de significância de 5% ($p \leq 0,05$).

$$Z = \frac{\hat{k}_1 - \hat{k}_2}{\sqrt{\hat{\sigma}_{k_1}^2 - \hat{\sigma}_{k_2}^2}}, \text{ Eq. 8}$$

Onde \hat{k}_1 e \hat{k}_2 representam os coeficientes kappa estimados para a classificação derivados de duas classificações diferentes e $\hat{\sigma}_{k_1}^2$ e $\hat{\sigma}_{k_2}^2$ suas variâncias associadas (Congalton e Mead, 1983; Flood, 2009). Onde $\hat{\sigma}_{k_1}^2$ é calculado através da Equação 9:

$$\hat{\theta}_k = \frac{1}{n} \left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4} \right], \text{ Eq. 9, em}$$

que:

$$\theta_1 = \frac{1}{n} \sum_{i=1}^m X_{ii}, \text{ Eq. 10}$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^m X_{i+} X_{+i}, \text{ Eq. 11}$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^m X_{ii} (X_{i+} + X_{+i}), \text{ Eq. 12}$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^m \sum_{j=1}^m X_{ij} (X_{j+} + X_{+j})^2, \text{ Eq. 13}$$

2.3.2. Análise gráfica de um conjunto de dados (*boxplot*)

O gráfico *boxplot* é utilizado para análise e comparação de diferentes conjuntos de dados, representando de forma resumida as estatísticas e distribuição dos dados. Ele permite a visualização do mínimo, quartil inferior, mediana, quartil superior e máximo de qualquer conjunto de dados (Figura 4) (Spitzer et al., 2014). A implementação mais comum do *boxplot* tem uma caixa que representa o (variação interquartil) IQR, com “*wiskers*” (bigodes) que se estendem 1,5 vezes o IQR das bordas da caixa; isto também permite a identificação de potenciais outliers no conjunto de dados (Spitzer et al., 2014).

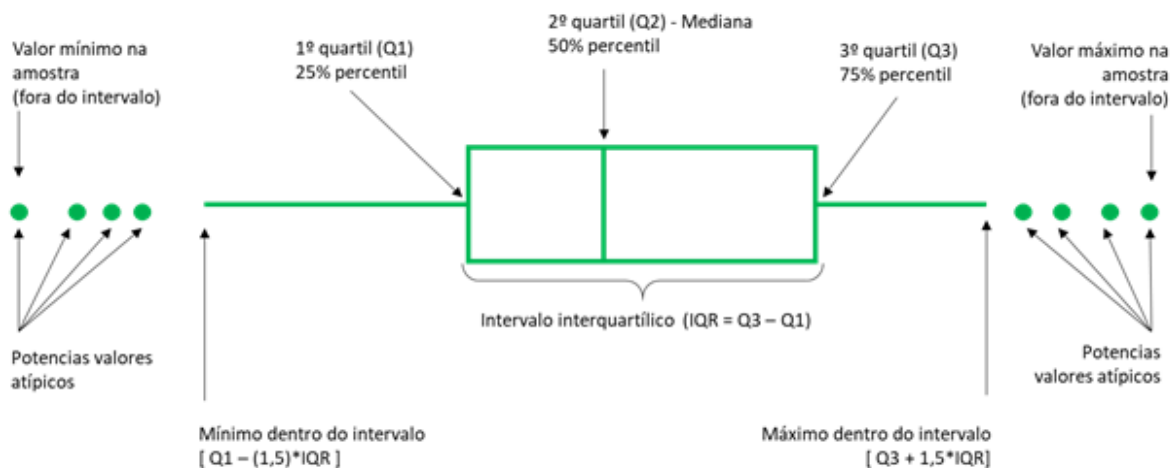


Figura 4: Descrição do gráfico *boxplot* com resumo de seis informações de um conjunto de dados (potências *outlines*, mínimo, quartil inferior, mediana, quartil superior e máximo).

Fonte: Siervi (2019).

2.4. Sensoriamento remoto

O sensoriamento remoto (SR) oferece uma nova perspectiva para o estudo, monitoramento e caracterização de grandes áreas (Almeida et al., 2019). Os sensores a bordo dos satélites permitem a captura de diferentes informações da superfície terrestre, incluindo dados de reflectância, temperatura e albedo de superfície, umidade do solo, índice de área foliar (IAF) e índices de vegetação (IVs). A extração de informações de cobertura terrestre a partir de imagens de sensoriamento remoto é o resultado da interação dos alvos na superfície e da radiação eletromagnética nas diferentes bandas espectrais (Jensen, 2009) (Figura 5).

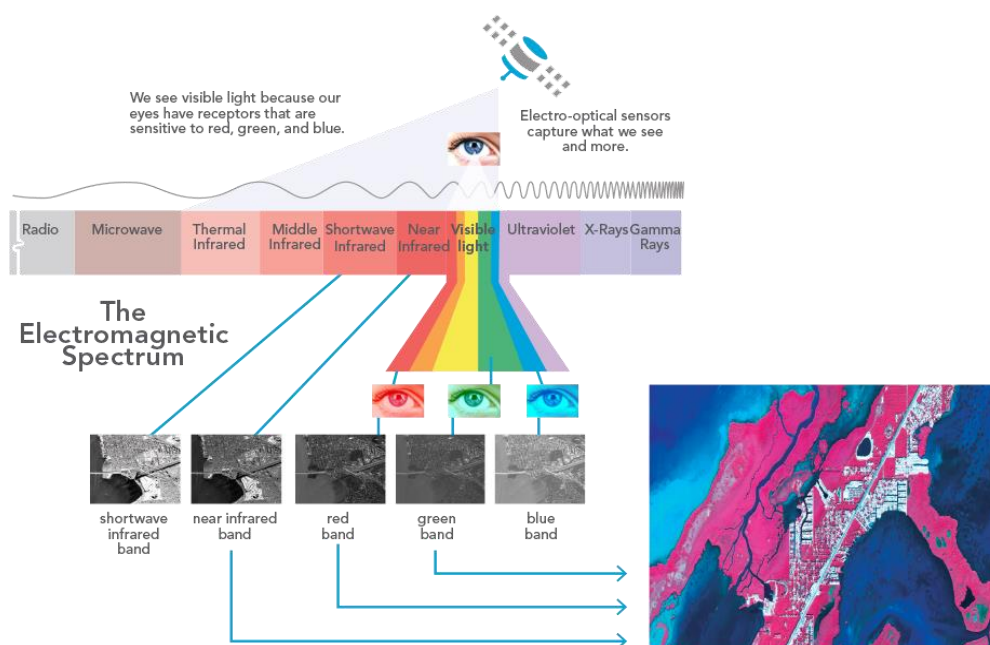


Figura 5: Interação entre superfície e radiação eletromagnética. Fonte: Esri (2020).

2.4.1. Série Landsat

A série Landsat (Figura 6) tem coletado imagens da superfície terrestre desde 1972, fornecendo um registro contínuo sobre como o ecossistema e a cobertura do solo foram modificados ao longo dessas cinco décadas. Desde 2008, todos os dados Landsat novos e arquivados mantidos pelo *United States Geological Survey* (USGS) foram disponibilizados gratuitamente na Internet para qualquer usuário, link: (<https://espa.cr.usgs.gov/>) (Woodcock et al., 2008). A estrutura de dados do Landsat permite realizar análises temporais em uma resolução espacial inferior a 100 metros (Woodcock et al., 2020). Esse conjunto de dados fornece informações sobre a qualidade da cobertura radiométrica, geométrica e identificação

de nuvens e sombras de nuvens (Wulder et al., 2016; Man et al., 2018), facilitando a diferenciação dos padrões de cobertura do solo em alta heterogeneidade espacial.

O Landsat-1 *Multispectral Scanner System* (MSS), lançado em 1972, tinha bandas espectrais que ocupavam comprimentos de onda visíveis e infravermelhos próximos e estavam coletando dados que normalmente eram reamostrados a 60 m em uma quantização de 6 bits - uma configuração de sensor que foi repetida por MSS a bordo dos Landsats -2 e -3 subsequentes em 1975 e 1978. Já o Landsat-4 e -5, lançados respectivamente em 1982 e 1984, o número de bandas espectrais aumentou para sete com o *Thematic Mapper* (TM) com a introdução de dois canais infravermelhos de ondas curtas (SWIR) com uma resolução espacial de 30 m (Wulder et al., 2019; USGS, 2020).

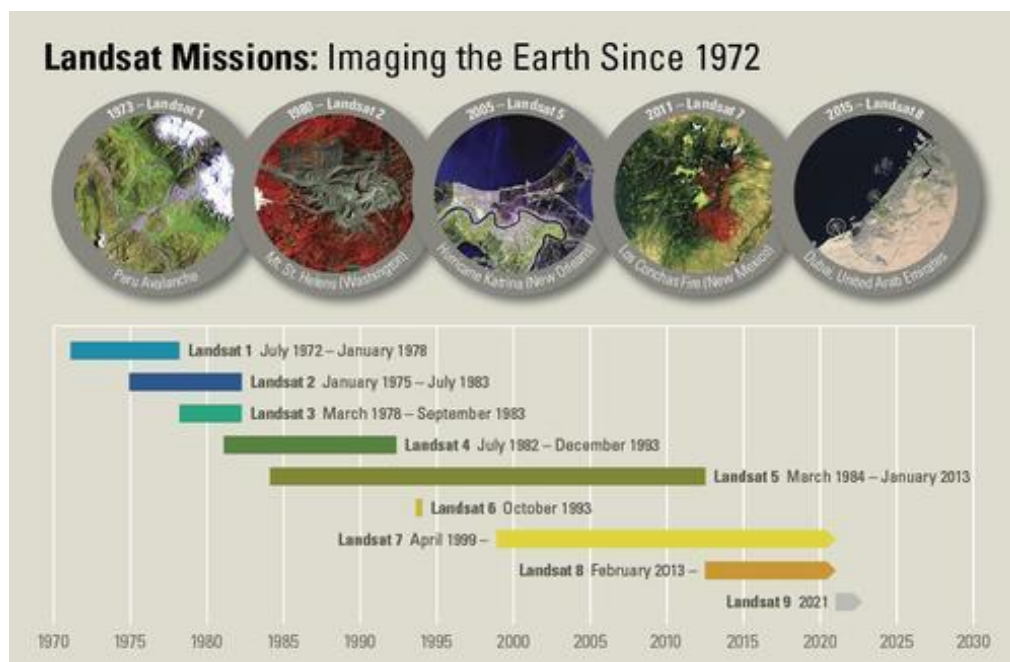


Figura 6: Linha do tempo das missões Landsat, Fonte: USGS (2020).

O Landsat 7 foi lançado em 1999, com o *Enhanced Thematic Mapper Plus* (ETM+) a bordo, tinha características de resolução espectral e espacial muito semelhantes ao TM, com exceção de um canal pancromático de resolução espacial adicional de 15 m e uma banda *Thermal Infrared Sensor* (TIR) de 60 m em vez da resolução espacial de 120 m de TIR e Landsat-4 e -5 (Wulder et al., 2019; USGS, 2020). No Landsat 8, lançado em 2013, estava a bordo o *Operational Land Imager* (OLI) que promoveu a tendência anterior de passagens de banda espectrais refinadas e a adição de novos canais, agora com quantização de 12 bits, contendo também um canal de azul (aerossol costeiro) e cirrus aprimorado para informar a ciência e as aplicações aquáticas e para melhorar a detecção e triagem de nuvens,

respectivamente (Wulder et al., 2019; USGS, 2020; Schott et al., 2016). O Landsat 9 está sendo preparado para ser lançado em meados de 2021 pela NASA, dando continuidade ao registro de dados da série Landsat (Masek et al., 2020). Os instrumentos a bordo Landsat 9 são réplicas aprimoradas dos sensores a bordo do 8. O (OLI-2) é praticamente idêntico ao Landsat 8 OLI e o *Thermal Infrared Sensor -2* (TIRS-2) melhora em relação ao Landsat 8 TIRS, abordando problemas conhecidos, incluindo incursão de luz difusa e um mau funcionamento do espelho de seleção de cena do instrumento, além disso (Masek et al., 2020; USGSb, 2021a).

2.4.2. Faixas espectrais

Os sensores a bordo de cada um dos satélites Landsat foram projetados para adquirir dados em diferentes faixas de frequências ao longo do espectro eletromagnético que podem auxiliar o mapeamento e monitoramento das coberturas do solo. Determinados alvos, como solo exposto, vegetação, umidade do solo respondem melhor a faixas espectrais específicas. A Tabela 2 apresenta as principais contribuições para o mapeamento das bandas espectrais do Landsat 8.

Tabela 2: Principais contribuições das faixas espectrais do Landsat 8 para o mapeamento.

Banda	Intervalo (μm)	Principais aplicações para o mapeamento
Banda 1 - Aerossol costeiro	0,43-0,45	Estudos costeiros e de aerossóis
Banda 2 - Azul	0,45-0,51	Mapeamento batimétrico, distinguindo solo de vegetação e vegetação caducifólia de conífera
Banda 3 - Verde	0,53-0,59	Enfatiza o pico da vegetação, o que é útil para avaliar o vigor da planta
Banda 4 - Vermelho	0,64-0,67	Discrimina encostas de vegetação
Banda 5 - Infravermelho próximo (NIR)	0,85-0,88	Enfatiza o conteúdo de biomassa e linhas costeiras
Banda 6 - Infravermelho de ondas curtas (SWIR) 1	1,57-1,65	Discrimina o conteúdo de umidade do solo e da vegetação; penetra nuvens finas
Banda 7 - Infravermelho de ondas curtas (SWIR) 2	2,11-2,29	Melhor conteúdo de umidade do solo e da vegetação; penetra nuvens finas
Banda 8 - Pancromático	0,50-0,68	Resolução de 15 metros, definição de imagem mais nítida
Banda 9 - Cirrus	1,36-1,38	Deteção aprimorada de contaminação de nuvem cirrus
Banda 10 - TIRS 1	10,60-11,19	Resolução de 100 metros, mapeamento térmico e estimativa de umidade do solo
Banda 11 - TIRS 2	11,50-12,51	Resolução de 100 metros, mapeamento térmico aprimorado e umidade estimada do solo

Fonte: USGS (2021b)

2.4.3. Índices de vegetação (IVs)

Os Índices de vegetação (IVs) têm como objetivo aumentar a sensibilidade dos recursos espectrais relacionados a uma variável biofísica de interesse, principalmente relacionada à vegetação. Por meio de combinações de bandas, um IV é capaz de aumentar a sensibilidade a um parâmetro específico da cultura e suprimir a influência de outros fatores de interferência, minimizando os efeitos de confusão devido ao ângulo do sol, ângulo de visão e composição atmosférica, variação de fundo do dossel da árvore, incluindo topografia, variações do solo e diferenças na vegetação renascida ou lenhosa (Pôças et al., 2020; Dong et al., 2019).

Várias dezenas de IVs foram desenvolvidos com base em bandas espectrais largas, fazendo uso de dados de sensores de satélite de banda larga (multiespectral), combinando principalmente a refletância dos domínios espectrais vermelho e NIR (Pôças et al., 2020). O *Normalized Difference Vegetation Index* (NDVI) é calculado a partir da normalização dessas bandas (Equação 14), ele é um dos índices mais conhecidos e adotados, com aplicações demonstradas na avaliação de parâmetros como biomassa verde, IAF, crescimento vegetal, fenologia e para mapeamento da cobertura da vegetação (Tucker, 1979; Jiang, Z., & Huete, 2010; Dong et al., 2019, Htitiou et al., 2019; Li et al., 2019).

$$NDVI = \frac{Nir-Red}{Nir+Red}, \text{ Eq.14}$$

Além disso, outros IVs, também combinando bandas nos domínios NIR e vermelho, foram desenvolvidos para incorporar ajustes do solo e da atmosfera, com a finalidade de melhorar a sensibilidade a condições de alta biomassa (Poças, 2020). Como por exemplo, o *Soil Adjusted Vegetation Index* – SAVI (Huete, 1988) e o *Enhanced Vegetation Index* - EVI (Huete et al., 2002), que também são amplamente utilizados. O SAVI inclui um fator de calibração do solo (L) para minimizar as influências de fundo do solo (Equação 15):

$$SAVI = \frac{NIR-Red}{NIR+Red+L} \times (1 + L), \text{ Eq. 15}$$

onde L é o índice de condicionamento do solo variando entre 0 e 1, com um valor de L próximo a 1 para vegetação de baixa densidade e próximo a 0,25 para vegetação de densa, com o valor de L de 0,5 representa melhor que NDVI nas várias condições da vegetação (Huete e Liu, 1994).

Já o EVI é um NDVI modificado desenvolvido pelo MODIS *Land Discipline Group* que integra um fator de ajuste do fundo do dossel que aborda a transmissão não linear e diferencial NIR e vermelho através de um dossel (L), bem como dois coeficientes (C1 e C2) que descrevem a correção do espalhamento do aerossol atmosférico usando a informação de uma faixa de domínio azul para corrigir as influências do aerossol na banda vermelha (Huete, 2002) (Equação 16). O EVI foi desenvolvido para otimizar o sinal da vegetação, ele apresenta maior sensibilidade às condições de vegetação de alta biomassa, além disso reduz o impacto do sinal de fundo do solo e a influência atmosférica (Jensen, 2009):

$$EVI = 2.5 \frac{(NIR - Red)}{(NIR + C1Red - C2Blue + L)}, \text{ Eq. 16}$$

2.5. Abordagens de classificação da cobertura vegetal

A classificação da cobertura vegetal pode ser obtida através de levantamento de campo dos agrupamentos vegetais (através de inventário tradicional) ou abordagens de sensoriamento remoto (SR). No entanto, a abordagem tradicional é mais demorada e pode cobrir apenas uma pequena extensão de toda a cobertura vegetal (Gyamfi-Ampadu et al., 2020). Por outro lado, os métodos de classificação da cobertura do solo através de SR, cobrem grandes áreas e fornecem informações da cobertura vegetal com baixo custo e de forma contínua (Pettorelli et al., 2014). As abordagens mais comuns através de SR utilizam dados multiespectrais em uma data específica (mono-temporal), procurando identificar uma imagem de satélite que apresenta baixa influência de nuvem e sombra de nuvem, ou utilizam dados multitemporais. No entanto esse último método exige um maior custo computacional para processamento das imagens, principalmente para eliminar os efeitos de nuvem e sombra de nuvem nas imagens (Schmidt et al., 2015; Vuolo et al., 2018).

2.5.1. Dados multiespectrais mono-temporais

Dados multiespectrais de sensoriamento remoto são utilizados para classificação da cobertura do solo. Os sensores multiespectrais produzem informações sobre os objetos presentes na superfície terrestre devido às bandas espectrais presentes nos sensores (Dalponte et al., 2008). As assinaturas espectrais são utilizadas para realizar a classificação de diversos cultivos na agricultura e florestas, utilizando geralmente uma única data de referência. Por exemplo, Gyamfi-Ampadu et al. (2020) utilizou imagem multiespectral Landsat de única data para mapeamento da cobertura florestal natural em uma reserva florestal de Nkandla, na África do Sul. Já Shaker et al. (2019) usaram imagens multiespectrais LiDAR para a classificação automática da terra-água em diferentes ambientes costeiros e fluviais interiores. Além disso, podem existir faixas espectrais que contribuem mais para diferenciação de determinadas coberturas do solo.

2.5.2. Dados multitemporais

A utilização de séries temporais também é utilizada para identificação e mapeamento de tipos de cobertura do solo. A utilização de dados multitemporais permite capturar o desenvolvimento específico da planta, e a utilização dessa assinatura temporal permite discriminar os tipos de cobertura (Vuolo et al., 2018). Vários estudos recentes utilizam dados ópticos multitemporais para classificação da cobertura do solo, por exemplo Zhong et al. (2019) utilizou séries temporais de EVI para realizar a classificação de um sistema agrícola irrigado no condado de Yolo na Califórnia. Persson et al. (2018) utilizaram dados multitemporais Sentinel-2 para classificar espécies de árvores comuns em uma floresta madura no centro da Suécia. Ghazaryan et al. (2018) utilizaram uma abordagem baseada em métricas fenológicas multi-temporais utilizando o Landsat e Sentinel-1 para mapeamento de safras em Vasilkovsky, na região de Kiev, localizada na porção central da Ucrânia. Além disso, a composição de imagens tem sido utilizada para gerar imagens espacialmente consistentes e livres de nuvens na série temporal para corrigir os efeitos dos diferentes ângulos zenitais e pixels inconsistentes (Hüttich et al., 2011).

2.6. Florestas tropicais sazonais secas

As florestas tropicais sazonais secas (FTSS) estão presentes nas regiões tropicais caracterizadas por até cinco meses de precipitação inferior a 100 mm (Allen et al., 2017) (Figura 7). As FTSS são ecossistemas que apresentam uma vasta biodiversidade e contribuem para a captura de CO₂ da atmosfera (Pennington et al., 2018; Silva,P et al., 2017). Os agrupamentos vegetais nas FTSS são heterogêneos: variam desde florestas altas até arbustos curtos ou cactos, mas são predominantemente dominadas por árvores decíduas, vegetação caracterizada pela perda de mais de 50% de folhas na estação seca (Espírito-santo et al., 2008; Moro et al., 2015). Além disso, a biomassa das FTSS está associada à disponibilidade de água decorrente da precipitação retida no solo que, além de variar intra e inter-anualmente, também depende da variabilidade horizontal e vertical das propriedades do solo (Almeida, 2019).

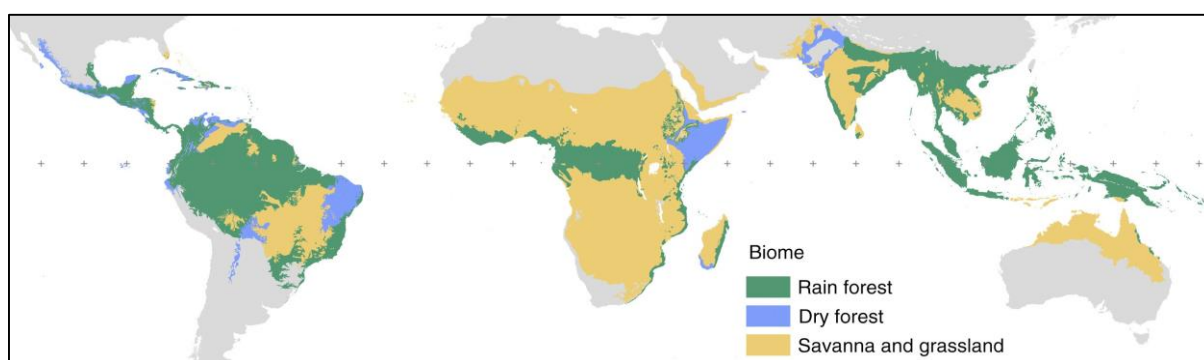


Figura 7: Mapa esquemático global de savanas tropicais e florestas secas.

Fonte: Pennington et al. (2018).

2.6.1 Caatinga

A Caatinga é a maior FTSS contínua do continente Americano (Allen et al., 2017; Queiroz et al., 2017; Pennington et al., 2018), cobrindo uma área de cerca de 844,453 km², correspondendo a 11% da Território brasileiro com cerca de ~ 1700 espécies, das quais pelo menos 300 são endêmicas (Moro et al., 2014; Brasil-MMA, 2019; Castanho et al., 2020). A vegetação da Caatinga apresenta diferentes fisionomias: arbórea, subarbórea, arbustiva, subarbustiva (Chaves et al., 2008) (Figura 8). A distribuição espacial dessas fisionomias está relacionada ao clima e ao grau de antropização da região (Silva et al., 2018; Castro et al., 2019). A Caatinga varia de florestas espinhosas xerófilas - uma combinação de arbustos e pequenas árvores com uma camada herbácea sazonal - a mosaicos de florestas semidecíduas e perenes (Moro et al., 2014; Castanho et al., 2020). As inúmeras combinações dessas fisionomias

produzem diferentes respostas espectrais que mudam com as estações do ano, tornando a classificação das imagens por meio de (SR) mais complicada do que em florestas tropicais temperadas e úmidas (Cunha et al., 2020), o que explica a escassez de mapas de vegetação com detalhes de suas fisionomias mesmo em áreas protegidas (Silveira et al., 2018).

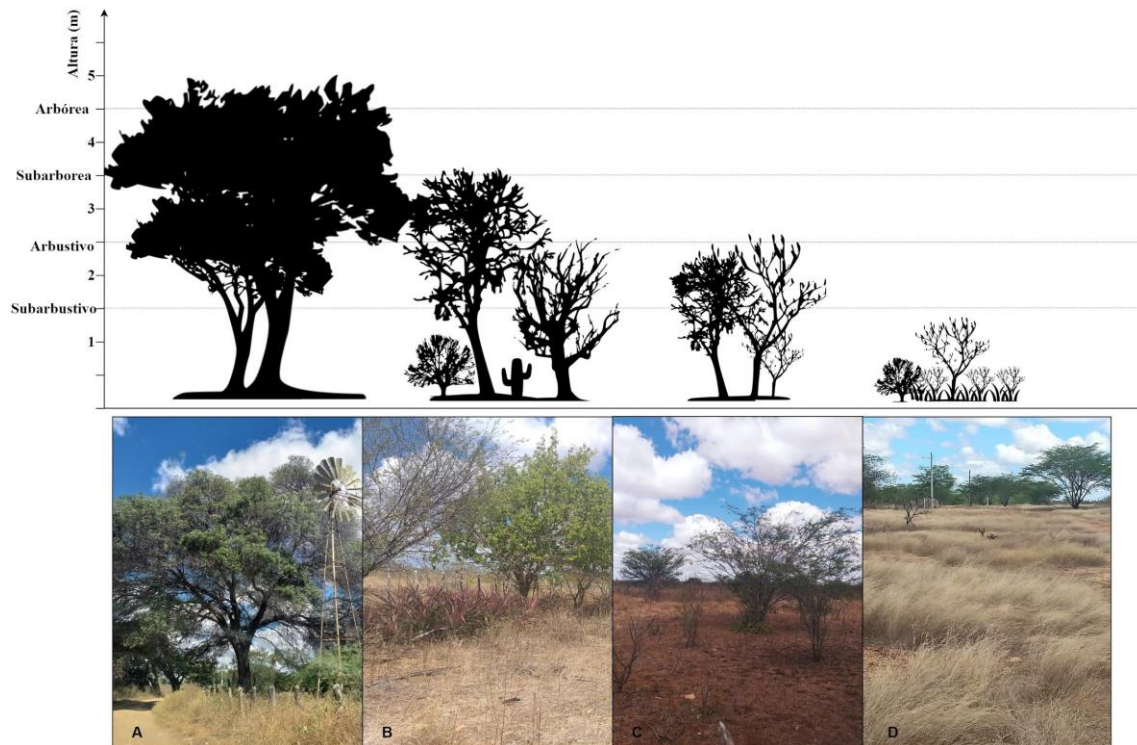


Figura 8: Fisionomias da Caatinga: A) Arbórea, B) Subarbórea, C) Arbustivo, D) Subarbustivo.

A Caatinga recebe na maior parte de sua área uma precipitação inferior a 750 mm/ano (Beuchle et al., 2015). As chuvas são concentradas em três a quatro meses consecutivos, e são extremamente irregulares espacialmente e temporalmente, o que acaba afetando a resposta fenológica da vegetação da Caatinga de forma intra e inter-anual (Moro et al., 2016). Para tolerar a escassez hídrica, as plantas desenvolveram mecanismos de sobrevivência. Suculentas e cactos evitam o estresse de déficit hídrico armazenando água em tecidos modificados (caules ou folhas) durante os períodos de escassez de água e espécies tolerantes à seca são capazes de sobreviver à perda de água sem sofrer danos irreparáveis aos tecidos vegetativos (Moro et al., 2015). As plantas diminuem seu metabolismo para suportar o estresse hídrico, e as taxas de evapotranspiração (ET) podem variar de 0,2 mm/dia na estação seca a 2,6 mm/dia na estação chuvosa (Marques et al., 2020) em uma área preservada da Caatinga.

A dinâmica da vegetação da Caatinga impacta diretamente o balanço energético e de água. Por exemplo, a variação da área foliar da vegetação modifica proporcionalmente a interceptação da precipitação e luz solar, trocas gasosas e o processo de fotossíntese (Almeida et al., 2019); as raízes das plantas afetam a estrutura do solo e a condutividade hidráulica (Gao et al., 2018). Além disso, a rugosidade aerodinâmica e o albedo causado pela sua estrutura vegetal, e a alta sazonalidade fenológica (fases de crescimento e desenvolvimento das plantas) afetam a demanda de água e são determinantes para a evapotranspiração. Nas áreas mais populosas, a *Caatinga* corre risco de desertificação e perda de biodiversidade devido a degradação da vegetação, como à conversão extensiva de terras para agricultura e pastagem, que associada a fatores climáticos, acelera o processo de desertificação da região (Drumond, 2004; Erasmi et al., 2014; Ribeiro et al., 2015). Consequentemente, as mudanças em sua cobertura do solo modificam diretamente a maior parte dos processos ecossistêmicos, especialmente os processos hidrológicos superficiais e podem potencializar o processo de desertificação (Schilling et al, 2008; Noretto et al., 2012).

O SR é utilizado em estudos relacionados a mudança de uso e cobertura do solo na Caatinga e para mapeamento de suas fisionomias. Por exemplo, Silveira et al. (2018) utilizou dados multi-temporais Sentinel-2 e LiDAR para mapeamento das fisionomias na Caatinga. Beuchle et al. (2015) destaca as mudanças na cobertura do solo da Caatinga com base na série Landsat. Oliveira et al. (2017) analisa o desmatamento em áreas protegidas decorrente da expansão agrícola no oeste da Bahia baseado no mapeamento do uso e cobertura do solo geradas através de imagens Landsat 5. A disponibilidade de dados de SR e avanço dos algoritmos de classificação oferecem novas oportunidades para o mapeamento das complexas fisionomias da Caatinga.

3. METODOLOGIA

3.1. Área de estudo

A área de estudo é a bacia hidrográfica do rio Sucuru (Figura 9), com área territorial de 1.682,87 km², localizada entre as coordenadas geográficas 7°28'30" e 7° 49' 30" Sul e 36° 34' 00" e 37° 12' 00" Oeste. Na área de estudo, a degradação da vegetação da Caatinga ocorreu principalmente por atividades antrópicas, como a exploração agropecuária e a extração de madeira (Moreira e Targino, 1997; Alves et al., 2017). O clima é semiárido quente, com duas estações distintas: a estação seca (de junho a janeiro) com menor biomassa da vegetação e a muito chuvosa (de fevereiro a maio) caracterizada pelo ganho de biomassa da vegetação (Cunha et al., 2020), com precipitação média anual de aproximadamente 596 mm e uma temperatura média anual de 24 °C (código de estação # 82792, INMET, 2020) (Figura 10). Os solos são rasos e pedregosos, o que dificulta a retenção de água após os eventos de precipitação (Moro et al., 2015). A área de estudo está localizada no núcleo de desertificação Cariris Velhos. Este núcleo é uma das áreas da região que apresenta alto risco de desertificação (INSA, 2016).

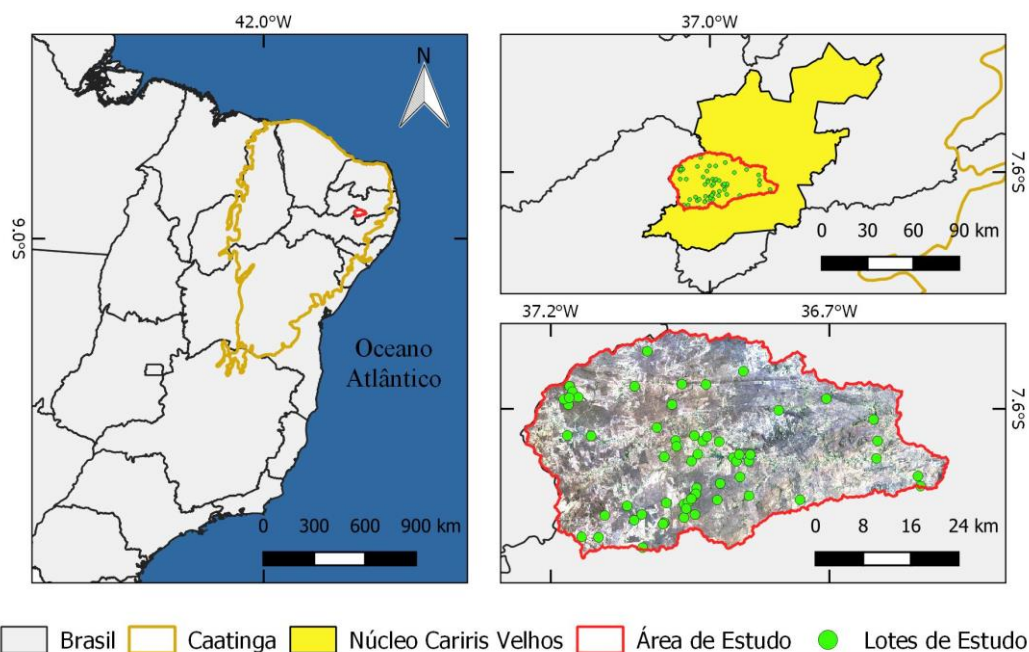


Figura 9: Localização da área de estudo na Caatinga. Os padrões verdes na imagem são as localizações dos dados de referência do solo.

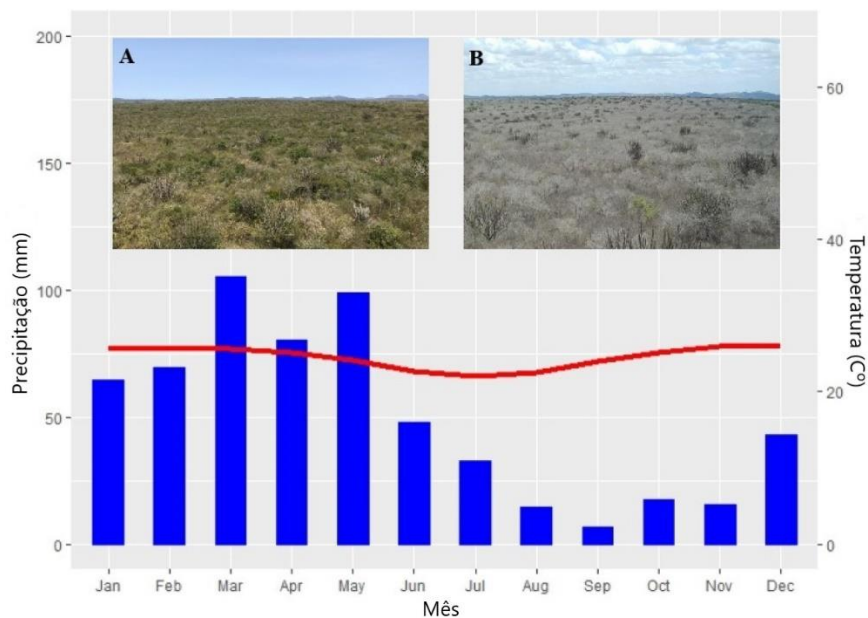


Figura 10: Climograma com base em dados climáticos da estação Monteiro de 1987 a 2016.

Fonte: INMET (2020), código da estação # 82792. A) Vegetação de Caatinga na estação chuvosa, B) Vegetação de Caatinga na estação seca.

3.2. Métodos

A Figura 11 apresenta o fluxo de trabalho da metodologia aplicada neste trabalho para avaliar o desempenho da classificação obtida por dados multiespectrais mono-temporal (data única) e da classificação por dados multi-temporais de NDVI. Primeiro, são coletados dados de campo e seleção das imagens de satélite. Em seguida, a série temporal de NDVI suavizada é reconstruída e os padrões temporais das classes de cobertura vegetal são identificados. Na etapa de processamento, o método *Random Forest* (RF) foi usado tanto para a classificação multiespectral mono-temporal quanto para a classificação multi-temporal NDVI. Por fim, é identificada a precisão e o desempenho dessas classificações.

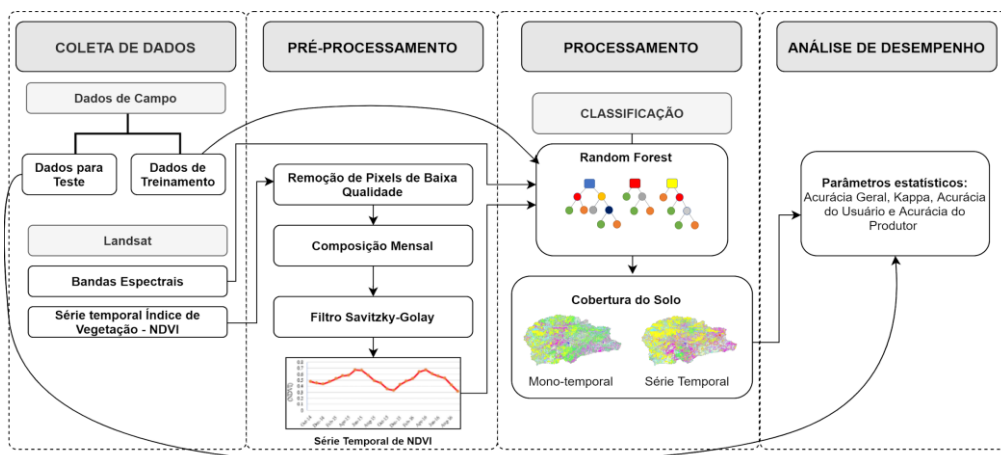


Figura 11: Fluxograma das etapas metodológicas.

3.2.1. Coleta de dados

A classificação da cobertura do solo da Caatinga seguiu a metodologia proposta por Chaves et al. (2008). Esses autores descrevem e avaliam a vegetação da Caatinga em seus diferentes estágios de antropização, com base no tamanho, características morfológicas e graus de cobertura. A Tabela 3 mostra as classes utilizadas de acordo com esta metodologia. Adicionou-se a classe Solo Exposto (SE), quando não há cobertura vegetal, totalizando seis classes de cobertura do solo. Visitas de campo coletaram os dados sobre seis classes de cobertura do solo da Caatinga previamente escolhidos. Os levantamentos ocorreram no início da primavera, quando a vegetação caducifólia ainda não perdeu suas folhas. Neste caso, os levantamentos ocorreram entre 28 de setembro e 7 de outubro de 2016, feito por Silva, J (2017). Utilizaram-se 60 polígonos para identificar os diferentes agrupamentos vegetais (classes) (Figura 9). Os polígonos não apresentam tamanho padrão e sua forma dependem do agrupamento vegetal que foi monitorado. Juntas, as amostras (polígono) com classes de cobertura do solo reuniram um total de 3.000 pixels, que representam as seis classes de cobertura solo da Caatinga. Todo o conjunto de pixels foi agrupado aleatoriamente em conjuntos de dados de treinamento (2.000 pixels) e teste (1.000 pixels) (Tabela 4).

Tabela 3: Classificação da vegetação da Caatinga.

Classes	Altura classe predominante (m)	Altura da segunda classe predominante (m)	Densidade (%)
ASMD	> 4.5 m	3.0 - 4.5 m	>80%
ASD	> 4.5 m	3.0 - 4.5 m	>60 < 80 %
SAA	3.0 - 4.5 m	1.5 - 3.0 m	>40 < 60 %
ATSA	1.5 - 3.0 m	3.0 - 4.5 m	>40 < 60 %
SAR	0 - 1.5 m	1.5 - 3.0 m	>20 < 40 %

Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Fonte: Chaves et al. (2008).

Tabela 4: Conjuntos de dados de treinamento e teste usados no estudo.

Classes	Pixels de treinamento	Pixels de teste
ASMD	320	190
ASD	315	195
SAA	347	163
ATSA	346	164
SAR	344	166
SE	328	122
Total	2000	1000

Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Solo Exposto (SE).

3.2.2. Imagens de satélite Landsat

As imagens de SR utilizadas são dos sensores ETM + e OLI, a bordo dos satélites Landsat 7 e 8, respectivamente. Essas imagens estão disponíveis gratuitamente pelo (USGS) (<https://espa.cr.usgs.gov/>), e há 88 imagens Landsat disponíveis de outubro de 2014 a setembro de 2016 que cobrem a área de estudo (44 de sensor ETM + e 44 do sensor OLI). A combinação das imagens de dois sensores resulta em amostragens para a mesma região em intervalos de oito dias com trinta metros de resolução espacial. Dois algoritmos diferentes geram dados Landsat em reflectância da superfície e dependem do sensor de medição: Os dados Landsat 7 ETM + são obtidos pelo software Landsat *Ecosystem Disturbance Adaptive Processing System* (LEDAPS) (Masek et al., 2006), e os dados Landsat 8 OLI são processados pelo algoritmo *Landsat Surface Reflectance Code* (LaSRC) (Vermote et al., 2016). O NDVI é calculado usando dados de refletância de superfície das regiões espectrais do infravermelho próximo e vermelho (NIR) (Tucker, 1979). O USGS fornece imagens NDVI consistentes em processamento e qualidade de dados para suportar análises de séries temporais e empilhamento de dados.

3.2.3. Pré-processamento

O pré-processamento digital das imagens, para realização da classificação por séries temporais, foi feito por um script R (R Development Core Team, 2018) desenvolvido para remoção dos pixels que apresentem interferências de sombras de nuvens, nuvens e água e redução de ruído. Primeiro, a avaliação da qualidade da refletância da superfície Landsat

(banda pixel_qa) que considera apenas pixels claros (valores 66 e 130 para Landsat 7, ou 322 e 386 para Landsat 8, USGS, 2019a, b) foi usada para reconstruir a série NDVI. Neste estudo, os valores máximos de NDVI são usados para reduzir a série NDVI original para imagens compostas mensais. Holben (1986) mostrou que o valor máximo é uma medida confiável para produzir composições representativas da imagem temporal em escala mensal. Os valores ausentes (removidos) foram preenchidos por interpolação linear. Para redução do ruído, o filtro linear *Savitzky-Golay* é aplicado (Savitzky e Golay, 1964; Cao et al., 2018), com uma janela de cinco meses para suavizar a largura, eliminando os *outliers* e erros introduzidos por interferência atmosférica não detectadas na remoção das interferências na série temporal.

3.2.4. Classificação

A abordagem de classificação mono-temporal multiespectral usa as bandas espectrais azul, verde, vermelho, infravermelho próximo (NIR) e duas bandas de infravermelho curto (SWIR) e (Figura 12) da imagem de 29 de setembro de 2016, Landsat 8 como entrada dados para o método RF. Na abordagem por multi-temporal NDVI, as séries mensais de NDVI de outubro de 2014 a setembro de 2016, totalizando 24 imagens compostas (Figura 11) são os dados de entrada para o procedimento de RF.

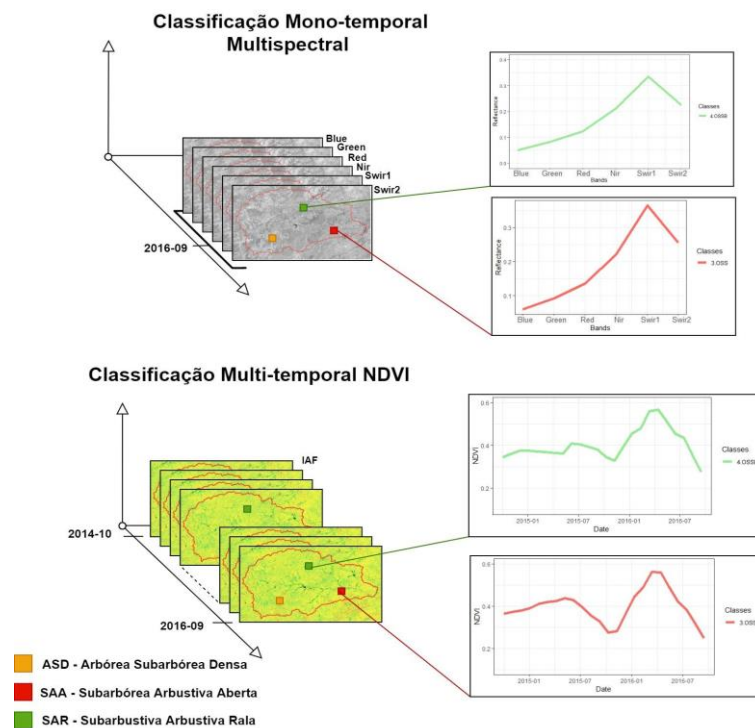


Figura 12: Abordagens de classificação: Mono-temporal multiespectral e multi-temporal NDVI.

A classificação das imagens supervisionadas foi realizada usando o pacote *Random Forest R* (Liaw e Wiener, 2002). O RF é um classificador de conjunto que produz várias árvores de decisão. Este método tem sido indicado para classificar a cobertura do solo devido à sua precisão (Valbuena et al., 2016). Os principais parâmetros dos modelos de RF, definidos pelo usuário, são o número máximo de árvores de decisão a serem geradas na floresta (*ntree*) e o número de variáveis usadas aleatoriamente para dividir cada nó (*mtry*) da árvore (Belgiu e Drăgut, 2016; Htitiou et al., 2019). Neste estudo, foram utilizadas quinhentas árvores de decisão e o *mtry* foi definido como a raiz quadrada do número de variáveis preditoras. Além disso, para avaliar a importância de cada variável para precisão da classificação no RF foi utilizado *Mean Decrease Gini* (MDG) (Corcoran et al., 2013; Nitze et al., 2015). Quanto maior for o valor do indicador, maior será o grau de importância da variável (Zhang et al., 2019).

3.2.5 Avaliação da precisão

A avaliação das classificações supervisionadas foi realizada com base em quatro acurácia geral (AG), coeficiente Kappa (k), precisão do produtor (PP) e precisão do usuário (PU) que são calculados a partir da matriz de confusão. Além disso, como forma de testar a diferença no desempenho das abordagens de classificação, foi realizado o teste Z, visando verificar se há diferença significativa entre as medidas de precisão de diferentes resultados de classificação (Floody, 2009).

A resposta espectral das classes de cobertura da Caatinga em relação às duas abordagens de classificação também foi analisada. Os *boxplots* das classes foram examinados visualmente para avaliar a separabilidade de cada classe. Esta técnica gráfica ilustra como os dados de treinamento das classes de cobertura do solo estão relacionados às entradas usadas nas abordagens de classificação.

4. RESULTADOS

4.1. Precisão das classificações

Os resultados da avaliação de desempenho mostraram que a classificação com base na série temporal mensal NDVI foi mais acurada, com uma AG de 88,8% e um coeficiente kappa de 0,86 (Tabela 5) do que a classificação mono-temporal multiespectral, com uma AG de 81,4% e um coeficiente kappa de 0,78 (Tabela 6). O teste-z produziu um valor 3,13, nesta situação $|z| > 1,96$ indica que há uma diferença estatisticamente significativa na precisão entre as abordagens. A comparação das precisões de cada abordagem de classificação mostra que as classificações obtiveram alta precisão ($> 70\%$) entre as diferentes classes (Figura 13, Tabelas 5 e 7). No entanto, as precisões mais baixas para algumas classes podem ser consideradas críticas para o mapeamento da cobertura do solo, em particular a classificação mono-temporal multiespectral para a vegetação aberta (SAA, ATSA, SAR, SE).

Tabela 5: Matriz de Confusão (%) da classificação multi-temporal de (NDVI).

Classificado	Referência						Total	PU (%)	
	SE	SAR	ATSA	SAA	ASD	ASMD			
SE	94	11	1	1	0	0	107	88	
SAR	20	153	8	5	1	0	187	82	
ATSA	3	2	144	16	2	1	168	86	
SAA	2	0	11	136	10	2	161	84	
ASD	3	0	0	3	182	8	196	93	
ASMD	0	0	0	2	0	179	181	99	
Total	122	166	164	163	195	190	1000		
PP (%)	77	92	88	83	93	94	AG (%)=	88,8	
k	0,86								

PU: precisão do usuário; PP: precisão do produtor; AG: Acurácia geral; k: kappa. Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Solo Exposto (SE).

Tabela 6: Matriz de Confusão (%) para classificação mono-temporal multiespectral.

Classificado	Referência						Total	PU (%)
	SE	SAR	ATSA	SAA	ASD	ASMD		
SE	73	18	4	0	0	0	95	77
SAR	36	139	11	1	0	1	188	74
ATSA	11	9	122	24	3	0	169	72
SAA	2	0	23	134	12	7	178	75
ASD	0	0	4	1	180	16	201	90
ASMD	0	0	0	3	0	166	169	98
Total	122	166	164	163	195	190	1000	
PP (%)	60	84	74	82	92	87	AG (%)=	81,4
k	0,78							

PU: precisão do usuário; PP: precisão do produtor; AG: Acurácia geral; k: kappa. Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Solo Exposto (SE).

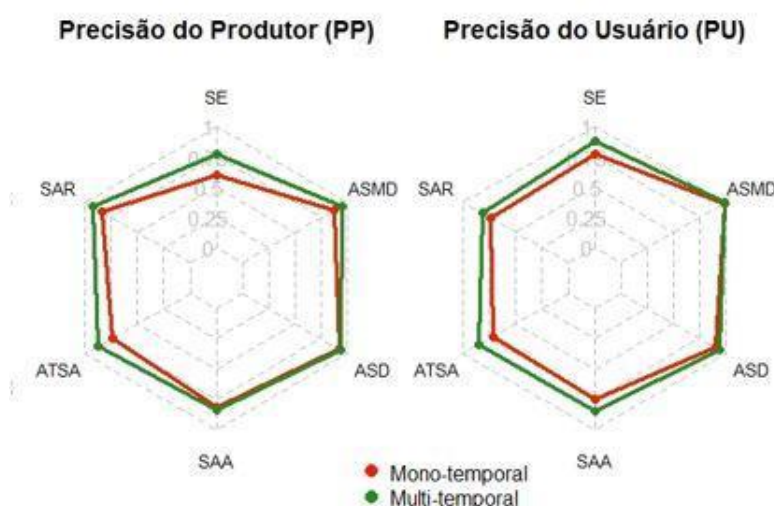


Figura 13: Gráfico de radar que representa as precisões do usuário e do produtor para a classificação NDVI multiespectral e de data única.

4.2. Análise das variáveis de treinamento no modelo de classificação

A Figura 14 mostra as amostras de treinamento das seis classes de cobertura do solo por meio de gráficos de *boxplot* para as duas abordagens. A figura indica que a abordagem multi-temporal NDVI revelou melhor os padrões de distinção para cada classe de cobertura do solo do que a abordagem mono-temporal multiespectral. Na classificação mono-temporal multiespectral, padrões semelhantes foram observados para as classes SAA, ATSA, SAR, SE; essas classes foram justamente aquelas que apresentaram o pior desempenho nesta abordagem de classificação para a precisão do usuário (PU) e precisão do produtor (PP). Na avaliação da

acurácia geral, as classificações ASD e ASMD apresentaram melhor desempenho. O ASMD se destacou dos demais por apresentar maior valor de NIR. Para outras classes, o SWIR 1 apresentou maior refletância. A menor variância para bandas espectrais foi observada para a classe ASD. Nas amostras de treinamento de classificação multi-temporal NDVI, os valores mais baixos de NDVI foram observados na classe SE, porém, seus contornos mostraram semelhanças com as classes mais vegetadas; A classe ASMD apresentou baixa variação fenológica pela classificação multi-temporal NDVI. Os resultados destacam que o pior desempenho pela classificação multi-temporal NDVI é observado na classe SE.

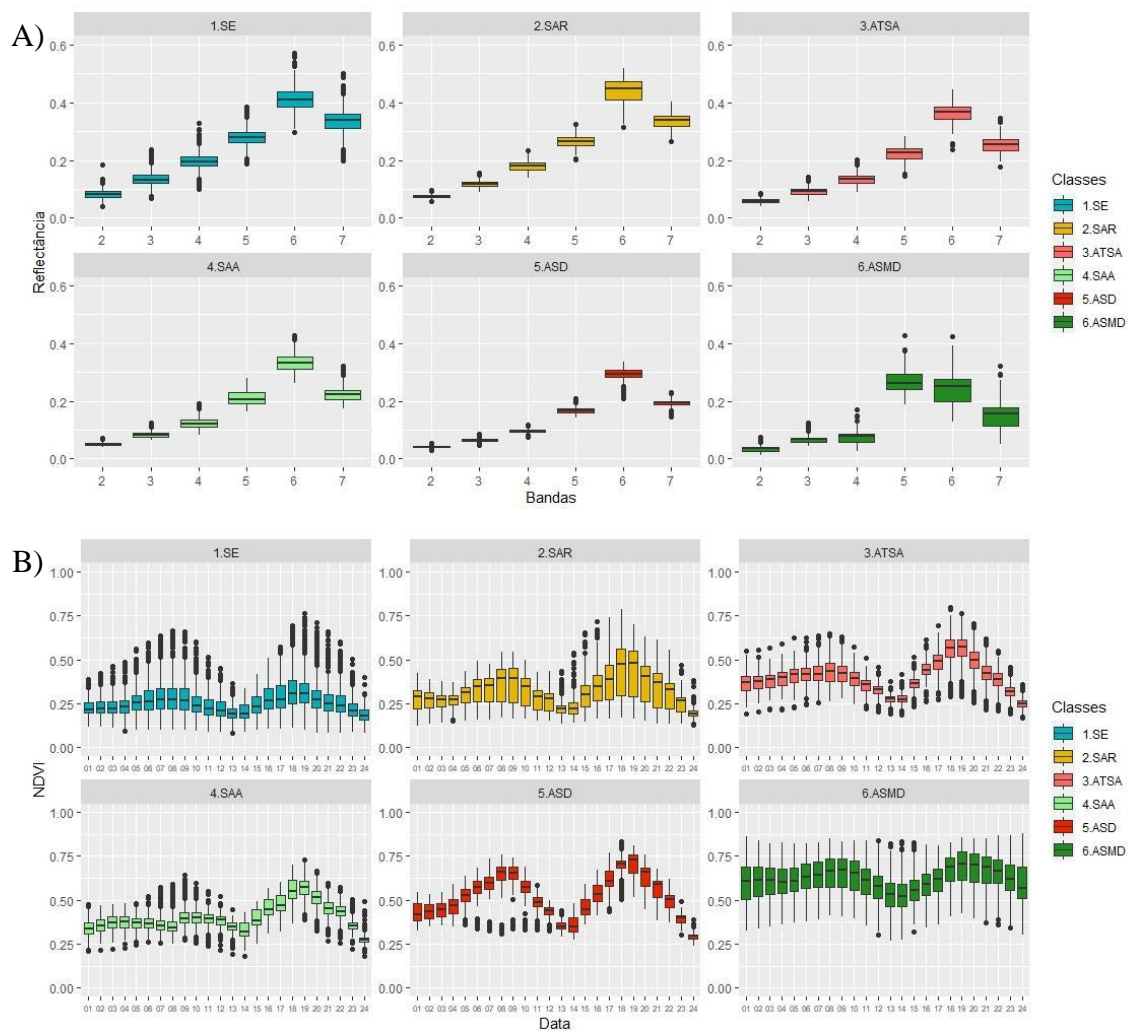


Figura 14: Gráficos de *boxplot* das amostras de treinamento para classes de cobertura da terra A) Bandas espectrais 2 (azul), 3 (verde), 4 (vermelha), 5 (NIR), 6 (SWIR 1), 7 (SWIR 2) da imagem Landsat 8 de 29 de setembro de 2016; B) 24 Composições Mensais de NDVI. Arbórea Subarbórea Muito Densa (ASMD), Arbórea Subarbórea Densa (ASD), Subarbórea Arbustiva Aberta (SAA), Arbustiva Subarbórea Aberta (ATSA), Subarbustiva Arbustiva Rala (SAR), Solo Exposto (SE).

As variáveis essenciais que mais contribuíram para a acurácia geral da classificação foram determinadas e avaliadas com base na *Mean Decrease Gini* (MDG) do RF. Na classificação mono-temporal multiespectral, as bandas vermelho, SWIR 2 e NIR foram determinadas como as variáveis mais importantes (Figura 15A). Na classificação multi-temporal NDVI, os atributos mais críticos para a classificação RF são os meses de outubro de 2015, agosto de 2016, setembro de 2016, enquanto janeiro de 2016 e maio de 2016 são as variáveis menos importantes (Figura 15B).

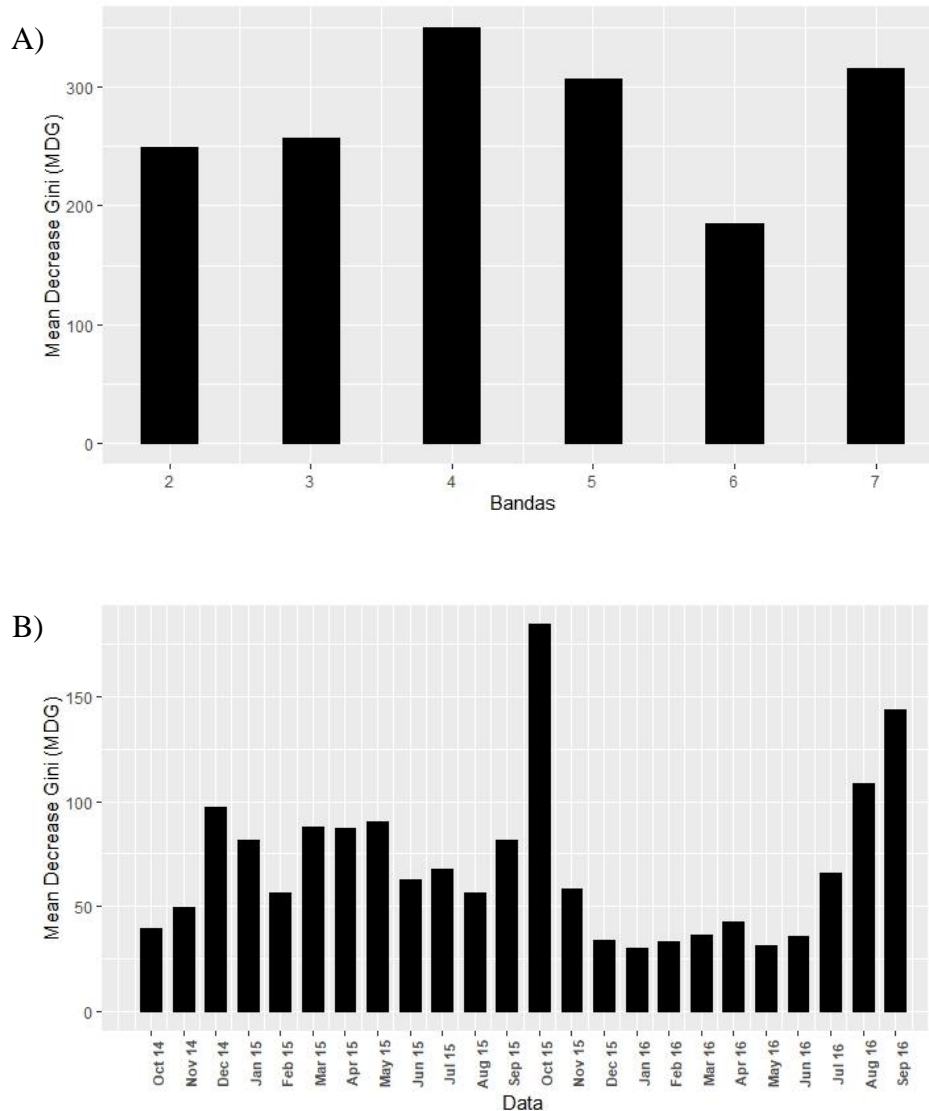


Figura 15: A) Classificação geral de importância variável usando o MDG: A) Classificação multiespectral de dados única, bandas espectrais 2 (azul), 3 (verde), 4 (vermelha), 5 (NIR), 6 (SWIR 1), 7 (SWIR 2) da imagem Landsat 8; B) Classificação Multi-temporal NDVI.

4.3. Distribuição espacial das classes de cobertura da Caatinga

Os valores das áreas de cada classe de cobertura do solo ao longo da bacia estudada (Figura 16), classificados pelas abordagens mono-temporal multiespectral e multi-temporal NDVI, mostram que algumas classes possuem estimativas diferentes para suas áreas de acordo com a abordagem utilizada (Figura 17). A aplicação da primeira abordagem resultou na seguinte composição: SE (4,73%), SAR (20,45%), ATSA (36,52%), SAA (22,82%), ASD (13,36%), ASMD (2,12%). A aplicação do segundo, entretanto, resultou em: SE (4,96%), SAR (22,16%), ATSA (38,63%), SAA (18,49%), ASD (13,25%), ASMD (2,51%). Os mapas da Figura 17 mostram que, embora as áreas apresentem magnitudes bastante semelhantes (Figura 17), existem diferenças espaciais na distribuição da cobertura do solo entre os mapas. Em geral, os mapas de cobertura do solo, de acordo com a classificação mono-temporal multiespectral, apresentam maior fragmentação espacial das classes de cobertura do solo, isso quer dizer que os pixels apresentam maior ruído entre seus vizinhos mais próximos, indicando uma maior dificuldade de distinguir as classes de cobertura de maneira homogênea, quando comparados com os gerados pela classificação multi-temporal NDVI.

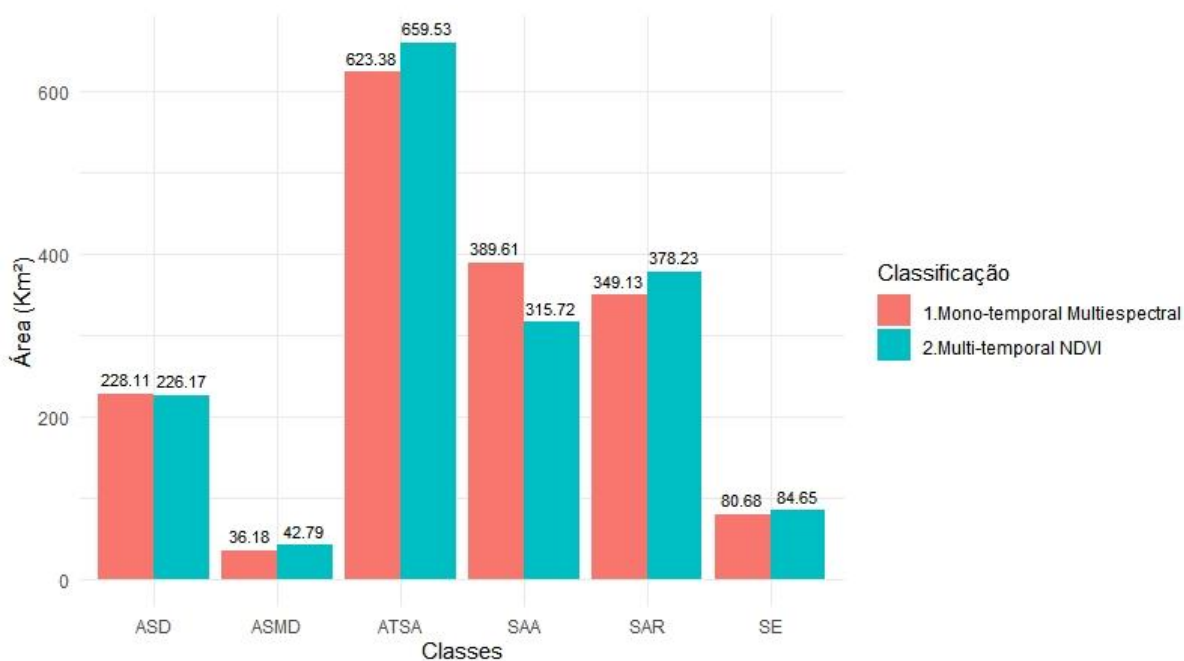


Figura 16: Área (km²) derivada das classificações mono-temporal multiespectral e multi-temporal NDVI.

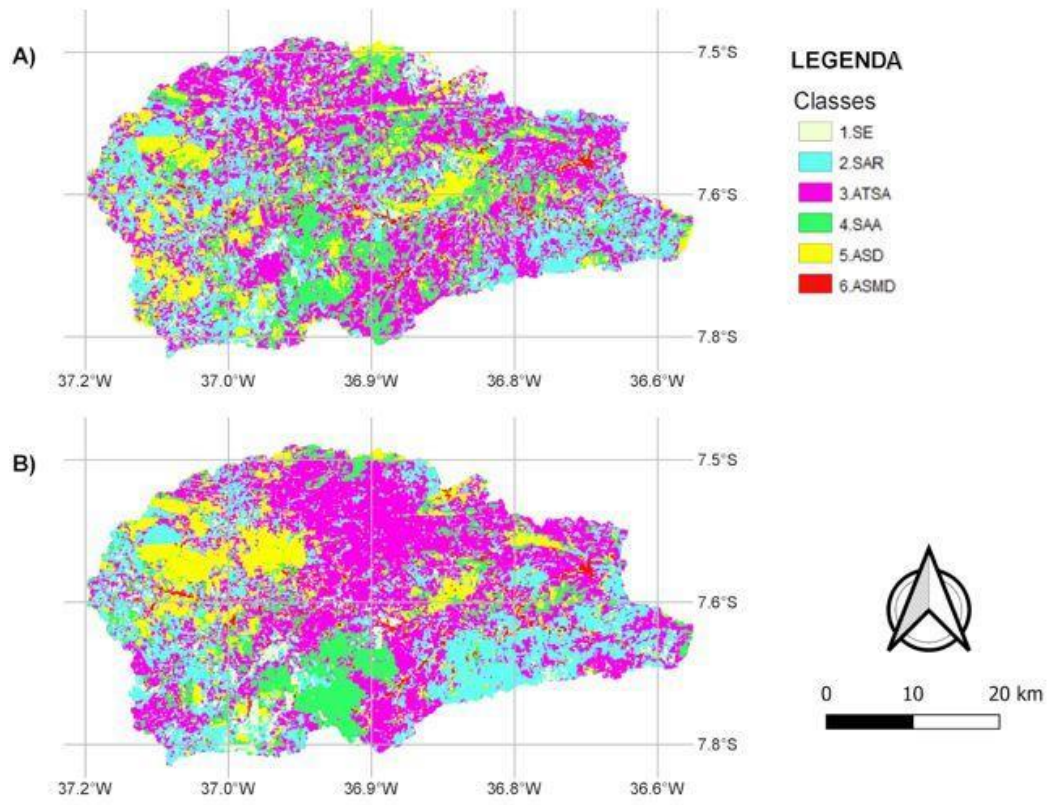


Figura 17: Mapas de cobertura do solo classificados por: A) Mono-temporal multiespectral e B) Multi-temporal NDVI.

5. DISCUSSÕES

O presente trabalho contribui para o ampliar o conhecimento sobre o uso de duas abordagens de classificação (mono-temporal multiespectral e multi-temporal NDVI) e suas capacidades para distinguir as fisionomias da Caatinga. A metodologia proposta busca contribuir para facilitar a compreensão da dinâmica do bioma, bem como cada banda espectral e composições mensais do NDVI contribuíram para o mapeamento das classes de vegetação. Também é possível avaliar a extensão das classes de cobertura do solo em uma região da Caatinga e fornecer informações para seu manejo e conservação, além disso essa informação é essencial para simulação de modelos hidrológicos e ecossistêmicos.

O desempenho inferior apresentado pela classificação mono-temporal multiespectral pode ser explicado pelas semelhanças nas respostas espectrais de algumas classes de cobertura (Figura 14A) (Vuolo et al., 2018). A banda SWIR 1 tem o maior valor de reflectância entre as classes de cobertura do solo, exceto a classe VDA. Isso se explica pela contribuição do solo exposto nessas classes de cobertura (Ciani et al., 2005; Tian e Philpot, 2015), uma vez que a maior parte da vegetação da Caatinga é esparsa. Porém, a classe ASMD possui vegetação muito densa, reduzindo a contribuição do solo exposto em sua resposta espectral, sendo a banda NIR seu maior valor de refletância (Ding et al., 2014). A Figura 18 permite visualizar a semelhança entre as classes SE e SAR e entre as classes ATSA e SAA. Alvos de vegetação caducifólia perdem sua folhagem durante a estação seca no ambiente da Caatinga e podem ser confundidos com áreas sem vegetação (Lima et al., 2012). Essa característica é mais significativa nas áreas de vegetação classificadas como abertas, uma vez que a vegetação do estrato superior, ao perder suas folhas, apresenta grandes porções de solo exposto.

Cada uma das espécies de vegetação da Caatinga responde de forma diferente à precipitação e à quantidade de água armazenada nos solos (Lima e Rodal, 2010; Moro et al., 2015). Durante a estação chuvosa, a maior parte das áreas de cobertura do solo aberto são invadidas por gramíneas de baixa altura, dificultando a distinção entre áreas de diferentes tamanhos de vegetação, devido à elevação da biomassa e alta atividade fotossintética momentânea. Portanto, a banda do infravermelho próximo apresentará alta refletividade ao aumento da biomassa, e as bandas azul e vermelha irão absorver a radiação solar para realizar a fotossíntese (Kumar et al., 2001). A alta heterogeneidade da cobertura do solo na Caatinga torna mais difícil distinguir os seus diferentes padrões por meio da interação entre a radiação eletromagnética e a superfície registrada por sensores a bordo dos satélites Landsat, essa

dificuldade no mapeamento também é uma identificada na vegetação heterogênea (espacialmente e temporalmente) da Savana africana (Hüttich et al., 2011).

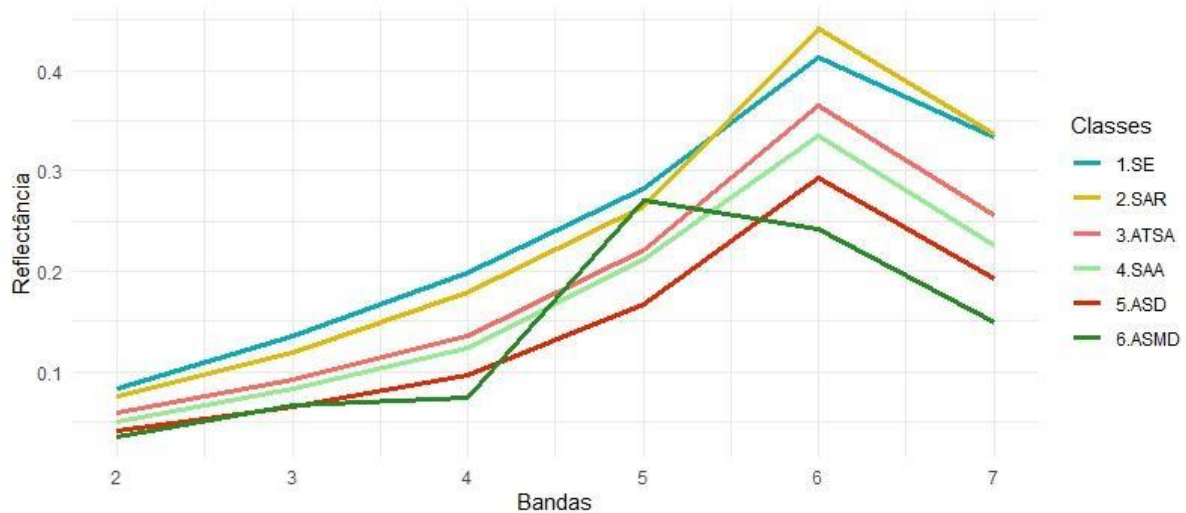


Figura 18: Refletância média da amostra de treinamento para classificação mono-temporal multiespectral.

Os resultados sugerem que o uso da abordagem de série temporal de índice de vegetação tem melhor desempenho para a classificação da cobertura do solo, em comparação com a classificação mono-temporal multiespectral em um ambiente altamente heterogêneo como a Caatinga. Conforme mencionado por Hüttich et al. (2011) e por Silveira et al. (2018), esta abordagem, ao considerar a variabilidade sazonal da atividade da vegetação e do ciclo fenológico no processo de classificação, aumenta o desempenho geral da classificação da cobertura do solo nas FTSS. Além disso, o comportamento fenológico detectado pela série temporal tem sido apontado como um fator chave para o melhor desempenho da classificação da cobertura do solo em ambientes semiáridos (Htitiou et al., 2019). Os *boxplots* na Figura 14B mostram que a classificação multi-temporal NDVI revelou diferentes padrões de cada classe de cobertura, o que ajuda a garantir uma boa discriminação entre as classes na área de estudo. É possível identificar uma assinatura temporal das classes de cobertura monitoradas, revelando um padrão no comportamento de cada tipo de cobertura do solo que facilita a distinção entre os agrupamentos vegetais, facilitando o mapeamento (Figura 19), apesar das semelhança entre as classes SAA vs ATSA e ASD vs ASMD em determinados meses.

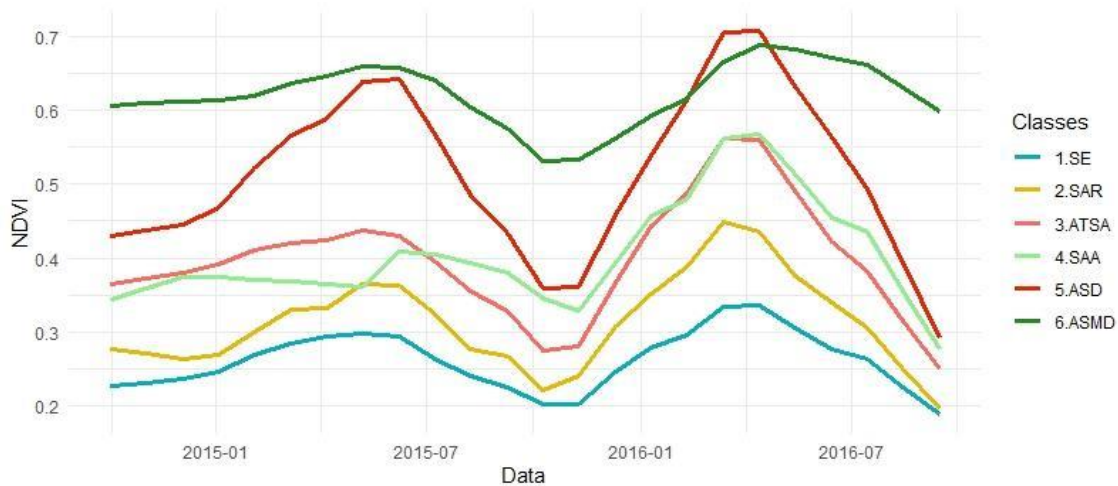


Figura 19: Valores médios de NDVI para cada classe da amostra de treinamento da classificação Multi-temporal NDVI.

As características morfológicas de cada classe podem explicar a distinção das assinaturas temporais, principalmente aquelas relacionadas à vegetação (Arvor et al., 2011). A Caatinga é caracterizada pela variação significativa da biomassa entre as estações seca e chuvosa (Barbosa e Kumar, 2016). Portanto, é necessário avaliar o comportamento sazonal de cada classe (Xia et al., 2017), isso é um fator importante para a distinção dos agrupamentos florísticos da Caatinga. Isso permitirá distinguir as classes de cobertura de forma mais eficaz, pois será possível verificar a resposta da vegetação ao longo dos ciclos de chuvas e secas a que o ecossistema está sujeito (Gomez et al., 2016; Levine e Crews, 2019).

No entanto, as classes com vegetação mais extensa e densa (VDSA, DSA) apresentaram desempenho semelhante nas duas abordagens (Tabela 5 e 4), o que indica que a construção da série temporal trouxe baixo benefício para distinguir essas classes, em comparação com a abordagem mono-temporal multiespectral. A classe ASMD é caracterizada por um comportamento fenológico mais estável (Figura 14B). Isso pode ser explicado pela proximidade do local de ocorrência desta classe aos cursos d'água, permitindo a disponibilidade de água durante os ciclos de chuva e seca. A Figura 14 mostra algumas classes que se destacam os *outliers*, o que pode ser ocasionado pela dificuldade de distingui-las no levantamento de campo, e também pela interferência antrópica em algumas regiões ao longo da série temporal. Atividades antrópicas na cobertura do solo da Caatinga podem prejudicar o desempenho da classificação multi-temporal NDVI (Maldonado et al., 2002; Jianya et al., 2008; Santos et al., 2013). A classe SE apresenta *outliers* com valor de NDVI de regiões com vegetação, o que sugere que alguns pontos podem ter sido usados como área de cultivo em determinados momentos (Figura 14B). Isso explica o pior desempenho entre as classes avaliadas pela

classificação NDVI de várias datas. Porém, ao comparar a classe SE para ambas as abordagens, percebe-se um desempenho inferior para a classificação mono-temporal multiespectral (Tabelas 5 e 6). A vegetação da Caatinga, especialmente para vegetações menores e abertas (SAR e ATSA), apresentam uma contribuição significativa do solo descoberto para a refletância (Figura 14A). Ou seja, a resposta do solo muitas vezes é observada de forma misturada à resposta da vegetação em períodos de baixa massa foliar, o que pode levar a resultados de classificação confusos ou imprecisos.

O MDG permitiu distinguir a relevância individual de cada banda Landsat 8 na classificação mono-temporal multiespectral (Figura 15A), e nas composições mensais do NDVI, como variáveis de entrada para a classificação multi-temporal NDVI (Figura 15B). Na classificação mono-temporal multiespectral, o vermelho foi determinado como a banda essencial no mapeamento. Ela teve a maior ordem de contribuição para a precisão, sendo seguido pelo SWIR 2 e NIR (Figura 15A). Isso pode ser atribuído à sua sensibilidade às variações da clorofila e da vegetação não fotossintética (Roberts et al., 1993; Ding et al., 2014; Delegido et al., 2015), presentes nas fisionomias da Caatinga com vegetação densa e aberta. A alta contribuição das bandas vermelha e NIR no mapeamento de áreas com vegetação é bem conhecida (Fensholt et al., 2012; Li et al., 2016); no entanto, o aspecto interessante desta avaliação é a contribuição da banda SWIR 2, que é sensível à umidade do solo. A participação significativa da banda SWIR no mapeamento da vegetação da Caatinga, provavelmente, se deve à alta exposição do solo à radiação solar. IVs usando as bandas NIR e SWIR mostram melhor capacidade de detectar fenologia do que NDVI e EVI (Jin et al., 2013), devido à sua maior sensibilidade à umidade na vegetação e no solo (Rodríguez-Caballero et al., 2015; Zhao et al., 2018). Na classificação multi-temporal NDVI, os meses de outubro de 2015, setembro de 2016 e agosto de 2016 têm a maior contribuição para a precisão da classificação e os meses de janeiro de 2016 e maio de 2016 têm a pior (Fig 15B). Estes resultados indicam que a separabilidade das classes vegetais é mais complicada na estação chuvosa do que na estação seca, explicada pela atividade fotossintética mais significativa da vegetação atuando em todas as classes de cobertura do solo na estação chuvosa, aumentando o valor de NDVI e tornando é mais difícil identificar diferenças de cobertura do solo nesta temporada (Guan et al., 2015; Walker et al., 2015).

Os resultados indicam o predomínio das classes mais abertas, como a classe ATSA com mais de 36%, em ambos os métodos de classificação. Isso pode ser explicado pela ação antrópica (pecuária, agricultura e extração de madeira, geralmente utilizada para recursos energéticos como carvão), que promove a degradação da vegetação e explica o menor

percentual (Figura 16) para a classe mais densa (ASMD) (Alves et al., 2017). As diferenças notáveis na detecção da cobertura do solo, como aquelas observadas entre as classificações mono-temporal multiespectral e multi-temporal NDVI (Figura 16 e Figura 17), podem interferir em inúmeras aplicações de planejamento ambiental e pesquisa nesta região, às vezes gerando aproximações enganosas da realidade. Por exemplo, essas diferenças nas áreas de cobertura do solo das abordagens podem impactar a aplicação de planos de zoneamento de capacidade agroecológicos, planos das bacias hidrográficas. Além disso, qualquer imprecisão na classificação da cobertura do solo pode direcionar de forma equivocada os investimentos em políticas de conservação da vegetação da Caatinga. Nesse sentido, é relevante avaliar os padrões das séries temporais de diferentes classes de cobertura do solo em florestas sazonalmente secas e, assim, permitir sua caracterização por meio de imagens de satélite. O monitoramento contínuo do solo de diferentes tipos de cobertura do solo é crucial para enfrentar o desafio da classificação da cobertura do solo na Caatinga e em outros ambientes secos e sazonais (Zhao et al., 2016).

6. CONCLUSÃO

Neste trabalho, foi realizada a classificação das complexas classes de cobertura da Caatinga utilizando abordagens baseadas em dados mono-temporal multiespectral e multi-temporal de NDVI com base na técnica de classificação *Random Forest*. O objetivo principal deste estudo era avaliar essas abordagens de classificação e indicar novas perspectivas para o mapeamento desta importante floresta tropical seca do Brasil. Além disso, através do *Random Forest* foi possível identificar as variáveis mais importantes para distinção das fisionomias da Caatinga.

A alta heterogeneidade espacial e variabilidade temporal da vegetação da Caatinga são elementos essenciais a serem considerados no processo de classificação da cobertura do solo. O uso de uma abordagem multi-temporal de NDVI para a caracterização da cobertura do solo neste ambiente tende a ser uma alternativa eficaz, em comparação à abordagem tradicional multiespectral de data única (mono-temporal), que leva em consideração apenas um instante no tempo. O desempenho inferior da classificação mono-temporal multiespectral foi observado nas classes com vegetação aberta e menor porte; para vegetação com maior densidade, o desempenho foi semelhante para ambas as abordagens. A classificação multi-temporal NDVI apresentou aproximadamente o mesmo desempenho para todas as classes de cobertura do solo, exceto para solo exposto. Algumas amostras de treinamento da classe de solo exposto apresentaram valores de NDVI equivalentes à vegetação densa, o que pode ter contribuído para diminuir a precisão da abordagem. A adoção dessa perspectiva permite um melhor reconhecimento e conhecimento aprofundado da dinâmica da cobertura do solo na Caatinga e outras regiões semelhantes, uma vez que é possível identificar uma assinatura temporal de cada classe de vegetação ao longo do tempo, permitindo uma melhor distinção de padrões entre as classes.

Além disso, na classificação mono-temporal multiespectral, as bandas vermelho, NIR e SWIR têm a contribuição mais notável para a precisão da classificação. Isso é atribuído à sensibilidade desses intervalos espectrais às variações da clorofila e da vegetação não fotossintética presentes nas fisionomias da Caatinga com vegetação densa e aberta. Na classificação multi-temporal NDVI, os meses na estação seca foram mais relevantes para a classificação do que os meses na estação chuvosa, explicado pela atividade fotossintética mais significativa da vegetação atuando em todas as classes de cobertura do solo na estação chuvosa, aumentando o valor NDVI e tornando mais difícil identificar as diferenças da cobertura do solo nesta temporada.

As principais contribuições deste trabalho são apresentadas de forma resumida a seguir:

- As imagens de Landsat com resolução espacial de 30 metros identificaram com sucesso as fisionomias complexas do bioma Caatinga tanto na abordagem mono-temporal quanto por série temporal, obtendo uma acurácia geral acima de 80%.
- Quando se compara as duas abordagens utilizadas neste estudo, a classificação realizada a partir de dados multi-temporais de NDVI foi melhor do que mono-temporal multiespectral.
- Em relação ao desempenho individual das classes de cobertura da Caatinga, foi possível observar que as classes que possuem vegetação mais densa e com maior porte apresentaram maior precisão quando se compara com a vegetação aberta em ambas as abordagens realizadas neste estudo.
- Na abordagem mono-temporal as bandas *Red*, NIR e SWIR tiveram uma contribuição mais notável para a precisão, isso se deve à uma maior sensibilidade da vegetação para as bandas do *Red* e NIR e uma maior influência do solo observada pelo SWIR.
- Na abordagem multi-temporal NDVI, os meses da estação seca foram os mais relevantes para a classificação quando comparado com os meses da estação chuvosa.

Além disso, entende-se que alguns trabalhos podem ser derivados deste estudo. Por exemplo, formular uma estratégia para seleção das imagens de satélite em períodos específicos para realizar a classificação multi-temporal, uma vez que alguns meses apresentam maior contribuição do que outros, isso permite obter um menor custo computacional para o processamento das imagens e também otimizar a aquisição delas. Já no que concerne às técnicas de classificação, deve-se investigar outros algoritmos de ML como *Support Vector Machine*, *XGboost*, Rede Neurais e avaliar seu desempenho para distinção das fisionomias da Caatinga. Outra possibilidade também é aplicar essas abordagens para outros satélites com resolução espacial e temporal diferente do Landsat, como MODIS com uma resolução de 250 metros e o Sentinel-2 com 10 metros, ambos os satélites com maior resolução temporal que o Landsat, além disso pode-se também replicar essas metodologias para outras regiões da Caatinga e florestas tropicais secas.

REFERÊNCIAS BIBLIOGRÁFICAS

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1-20. <https://doi.org/10.1080/15481603.2019.1650447>.

Alhassan, V., Henry, C., Ramanna, S., & Storie, C. (2019). A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. *Neural Computing and Applications*, 1-16. <https://doi.org/10.1007/s00521-019-04349-9>.

Allen, K., Dupuy, J. M., Gei, M. G., Hulshof, C., Medvigy, D., Pizano, C., ... Powers, J. S. (2017). Will seasonally dry tropical forests be sensitive or resistant to future changes in rainfall regimes? *Environmental Research Letters*, 12(2), 023001. <https://doi.org/10.1088/1748-9326/aa5968>.

Almeida, C. L., Carvalho, T. R., & Araújo, J. C. (2019). Leaf area index of Caatinga biome and its relationship with hydrological and spectral variables. *Agricultural and Forest Meteorology*, 279, 107705. <https://doi:10.1016/j.agrformet.2019.107705>.

Alves, T. L. B., de Azevedo, P. V., & dos Santos, C. A. C. (2017). Influence of climate variability on land degradation (desertification) in the watershed of the upper Paraíba River. *Theoretical and Applied Climatology*, 127(3-4), 741-751. <https://doi.org/10.1007/s00704-015-1661-1>.

Arvor, D., Jonathan, M., Meirelles, M. S. P., Dubreuil, V., & Durieux, L. (2011). Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, 32(22), 7847-7871. <https://doi.org/10.1080/01431161.2010.531783>.

Barbosa, H. A., & Kumar, T. L. (2016). Influence of rainfall variability on the vegetation dynamics over Northeastern Brazil. *Journal of Arid Environments*, 124, 377-387. <https://doi.org/10.1016/j.jaridenv.2015.08.015>.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>.

Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.

Beuchle, R., Grecchi, R. C., Shimabukuro, Y. E., Seliger, R., Eva, H. D., Sano, E., & Achard, F. (2015). Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Applied Geography*, 58, 116-127. <https://doi.org/10.1016/j.apgeog.2015.01.017>.

BRAZIL, Ministério do Meio Ambiente. Caatinga: exclusivamente brasileira. (Agenda Caatinga). <http://www.mma.gov.br/biomas/caatinga>. Accessed date: 17 March 2019.

Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J., Wang, C., & Yang, W. (2018). A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *Remote Sensing of Environment*, 217, 244-257. <https://doi.org/10.1016/j.rse.2018.08.022>.

Castanho, A. D. A., Coe, M. T., Brando, P., Macedo, M., Baccini, A., Walker, W., & Andrade, E. M. (2020). Potential shifts in the aboveground biomass and physiognomy of a seasonally dry tropical forest in a changing climate. *Environmental Research Letters*, 15(3), 034053. <https://doi.org/10.1088/1748-9326/ab7394>.

Castro, O. G., Francelino, M. R., Arruda, D. M., Fernandes-Filho, E. I., & Schaefer, C. E. G. R. (2019). Climate and soils at the Brazilian semiarid and the forest-Caatinga problem: new insights and implications for conservation. *Environmental Research Letters*, 14(10), 104007. <https://doi.org/10.1088/1748-9326/ab3d7b>.

Ciani, A., Goss, K.-U., & Schwarzenbach, R. P. (2005). Light penetration in soil and particulate minerals. *European Journal of Soil Science*, 56(5), 561–574. <https://doi.org/10.1111/j.1365-2389.2005.00688.x>.

Chaves, I. D. B., Lopes, V. L., Folliott, P. F., & Paes-Silva, A. P. (2008). Uma classificação morfo-estrutural para descrição e avaliação da biomassa da vegetação da caatinga. *Revista Caatinga*, 21(2), 204-213.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35-46. [https://doi.org/10.1016/0034-4257\(91\)90048-b](https://doi.org/10.1016/0034-4257(91)90048-b).

Congalton, R. G., & Mead, R. A. (1983). A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, 49(1), 69-74.

Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press. <https://doi.org/10.1201/9781420055139>.

Corcoran, J., Knight, J., & Gallant, A. (2013). Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota. *Remote Sensing*, 5(7), 3212–3238. <https://doi.org/10.3390/rs5073212>.

Cunha, J., Nóbrega, R. L., Rufino, I., Erasmi, S., Galvão, C., & Valente, F. (2020). Surface albedo as a proxy for land-cover clearing in seasonally dry forests: evidence from the Brazilian Caatinga. *Remote Sensing of Environment*, 238, 111250. <https://doi.org/10.1016/j.rse.2019.111250>.

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5), 1416–1427. <https://doi:10.1109/tgrs.2008.916480>.

Delegido, J., Verrelst, J., Rivera, J. P., Ruiz-Verdú, A., & Moreno, J. (2015). Brown and green LAI mapping through spectral indices. *International Journal of Applied Earth Observation and Geoinformation*, 35, 350–358. <https://doi.org/10.1016/j.jag.2014.10.001>.

Ding, Y., Zhao, K., Zheng, X., & Jiang, T. (2014). Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8

OLI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 30, 139–145. <https://doi.org/10.1016/j.jag.2014.01.009>.

Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., ... Shi, Y. (2019). Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sensing of Environment*, 222, 133–143. <https://doi:10.1016/j.rse.2018.12.032>.

Drumond, M. A., Kiill, L. H. P., Lima, P. C. F., de Oliveira, M. C., de Oliveira, V. R., de Albuquerque, S. G., & Cavalcanti, J. (2004). Estratégias para o uso sustentável da biodiversidade da caatinga.. In: Silva. J.M.C., Tabarelli, Fonseca, M., M.T., Lins,L.V. Lins (Eds.), *Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação*, Ministério do Meio ambiente, Brasília , 329-340.

Efron, B. (1992). Bootstrap Methods: Another Look at the Jackknife. *Breakthroughs in Statistics*, 569–593. https://doi:10.1007/978-1-4612-4380-9_41.

Erasmí, S., Schucknecht, A., Barbosa, M., & Matschullat, J. (2014). Vegetation Greenness in Northeastern Brazil and Its Relation to ENSO Warm Events. *Remote Sensing*, 6(4), 3041–3058. <https://doi.org/10.3390/rs6043041>.

Espírito-Santo, M. M., Fagundes, M., Sevilha, A. C., Scariot, A. O., Sanchez-Azofeifa, G. A., Noronha, S. E., & Fernandes, G. W. (2008). Florestas estacionais decíduas brasileiras: distribuição e estado de conservação. *MG Biota*, 1(2), 5-13.

Esri. Seeing Beyond the Visible. <https://www.esri.com/about/newsroom/insider/seeing-beyond-the-visible/>. Acessado: 22 Dezembro 2020.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., ... & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64-87.

Fensholt, R., Langanke, T., Rasmussen, K., Reenberg, A., Prince, S. D., Tucker, C., ... Wessels, K. (2012). Greenness in semi-arid areas across the globe 1981–2007 — an Earth Observing Satellite based analysis of trends and drivers. *Remote Sensing of Environment*, 121,

144–158. <https://doi.org/10.1016/j.rse.2012.01.017>.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185-201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4).

Foody, G. M. (2009). Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment*, 113(8), 1658–1663. <https://doi.org/10.1016/j.rse.2009.03.014>.

Gao, H., Sabo, J. L., Chen, X., Liu, Z., Yang, Z., Ren, Z., & Liu, M. (2018). Landscape heterogeneity and hydrological processes: a review of landscape-based hydrological models. *Landscape Ecology*, 33(9), 1461–1480. <https://doi.org/10.1007/s10980-018-0690-4>.

Ghazaryan, G., Dubovyk, O., Löw, F., Lavreniuk, M., Kolotii, A., Schellberg, J., & Kussul, N. (2018). A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *European Journal of Remote Sensing*, 51(1), 511–524. <https://doi:10.1080/22797254.2018.1455540>.

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55-72. <https://doi.org/10.1016/j.isprsjprs.2016.03.008>.

Gomes, R. C. F. de S., Teixeira, B. L. B., Gusmão, C. L. S., & Fernandes, A. M. (2020). Humidity effects on avian blood parasites in the Caatinga of Brazil. *Ornithology Research*, 28(2), 98–104. <https://doi.org/10.1007/s43388-020-00009-y>.

Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., ... Lyapustin, A. I. (2015). Photosynthetic seasonality of global tropical forests constrained by hydroclimate. *Nature Geoscience*, 8(4), 284–289. <https://doi.org/10.1038/ngeo2382>.

Gyamfi-Ampadu, E., Gebreslasie, M., & Mendoza-Ponce, A. (2020). Mapping natural forest cover using satellite imagery of Nkandla forest reserve, KwaZulu-Natal, South Africa. *Remote Sensing Applications: Society and Environment*, 18, 100302. <https://doi:10.1016/j.rsase.2020.100302>.

Htitiou, A., Boudhar, A., Lebrini, Y., Hadria, R., Lionboui, H., Elmansouri, L., ... & Benabdelouahab, T. (2019). The Performance of Random Forest Classification Based on

Phenological Metrics Derived from Sentinel-2 and Landsat 8 to Map Crop Cover in an Irrigated Semi-arid Region. *Remote Sensing in Earth Systems Sciences*, 2(4), 208-224. <https://doi.org/10.1007/s41976-019-00023-9>.

Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11), 1417-1434. <https://doi.org/10.1080/01431168608948945>.

Huete, A. . (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi:10.1016/0034-4257\(88\)90106-x](https://doi:10.1016/0034-4257(88)90106-x).

Huete, A. R., & Liu, H. Q. (1994). An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), 897–905. <https://doi:10.1109/36.298018>.

Huete, A., Didan, K., Miura, T., Rodriguez, E. ., Gao, X., & Ferreira, L. . (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2), 195–213. [https://doi:10.1016/s0034-4257\(02\)00096-2](https://doi:10.1016/s0034-4257(02)00096-2).

Hüttich, C., Herold, M., Wegmann, M., Cord, A., Strohbach, B., Schmullius, C., & Dech, S. (2011). Assessing effects of temporal compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring. *Remote Sensing of Environment*, 115(10), 2445-2459. <https://doi.org/10.1016/j.rse.2011.05.005>.

INMET, National Institute of Meteorology of Brazil, 2018. <https://bdmep.inmet.gov.br/#>, Accessed date: October 2020.

INSA (2016). Information and Knowledge Management System of the Brazilian Semi-arid Region (SIGSAB, of the acronym in Portuguese). <http://sigsab.insa.gov.br>. Accessed date: 17 March 2019.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer Texts in Statistics. <https://doi:10.1007/978-1-4614-7138-7>.

Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective 2/e*. Pearson Education India.

Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., & Li, B. (2014). Land cover classification using Landsat 8 Operational Land Imager data in Beijing, China. *Geocarto International*, 29(8), 941–951. <https://doi.org/10.1080/10106049.2014.894586>.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., & Wang, X. (2014). Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. *Remote sensing*, 6(11), 11518-11532.

Jin, C., Xiao, X., Merbold, L., Arneth, A., Veenendaal, E., & Kutsch, W. L. (2013). Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. *Remote Sensing of Environment*, 135, 189–201. <https://doi.org/10.1016/j.rse.2013.03.033>.

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. *International Journal of Remote Sensing*, 39(23), 8703–8723. <https://doi.org/10.1080/01431161.2018.1490976>.

Jiang, Z., & Huete, A. R. (2010). Linearization of NDVI Based on its Relationship with Vegetation Fraction. *Photogrammetric Engineering & Remote Sensing*, 76(8), 965–975. <https://doi:10.14358/pers.76.8.965>.

Jianya, G., Haigang, S., Guorui, M., & Qiming, Z. (2008). A review of multi-temporal remote sensing data change detection algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7), 757-762.

Karnieli, A., Gabai, A., Ichoku, C., Zaady, E., & Shachak, M. (2002). Temporal dynamics of soil and vegetation spectral responses in a semi-arid environment. *International Journal of Remote Sensing*, 23(19), 4073–4087. <https://doi.org/10.1080/01431160110116338>.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. doi:10.1007/978-1-4614-6849-3.

Kumar, L., Schmidt, K.S., Dury, S., Skidmore, A.K. (2001). Review of hyperspectral remote sensing and vegetation science. F. van der Meer (Ed.), *Hyperspectral Remote Sensing*, Kluwer Academic Press, Dordrecht, 111-155.

Lantz, B. (2013). *Machine learning with R: Learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications*. Packt Publishing.

Levine, D., & Crews, K. (2019). Time series harmonic regression analysis reveals seasonal vegetation productivity trends in semi-arid savannas. *International Journal of Applied Earth Observation and Geoinformation*, 80, 94–101. <https://doi.org/10.1016/j.jag.2019.04.007>.

Li, C., Li, H., Li, J., Lei, Y., Li, C., Manevski, K., & Shen, Y. (2019). Using NDVI percentiles to monitor real-time crop growth. *Computers and Electronics in Agriculture*, 162, 357–363. <https://doi.org/10.1016/j.compag.2019.04.026>.

Li, X. B., Li, R. H., Li, G. Q., Wang, H., Li, Z. F., Li, X., & Hou, X. Y. (2016). Human-induced vegetation degradation and response of soil nitrogen storage in typical steppes in Inner Mongolia, China. *Journal of Arid Environments*, 124, 80–90. <https://doi.org/10.1016/j.jaridenv.2015.07.013>.

Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. *R news*, 2(3), 18-22.

Lima, A. L. A., & Rodal, M. J. N. (2010). Phenology and wood density of plants growing in the semi-arid region of northeastern Brazil. *Journal of Arid Environments*, 74(11), 1363–1373. <https://doi.org/10.1016/j.jaridenv.2010.05.009>.

Lima, A. L. A., de Sá Barretto Sampaio, E. V., de Castro, C. C., Rodal, M. J. N., Antonino, A. C. D., & de Melo, A. L. (2012). Do the phenology and functional stem attributes of woody species allow for the identification of functional groups in the semiarid region of Brazil? *Trees*, 26(5), 1605–1616. <https://doi.org/10.1007/s00468-012-0735-2>.

Machado, G., Mendoza, M. R., & Corbellini, L. G. (2015). What variables are important in predicting bovine viral diarrhea virus? A random forest approach. *Veterinary research*, 46(1), 1-15. <https://doi.org/10.1186/s13567-015-0219-7>.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., & Zhang, Y. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing*, 10(7), 1119. <https://doi.org/10.3390/rs10071119>.

Maldonado, F. D., Santos, J. R. D., & De Carvalho, V. C. (2002). Land use dynamics in the semi-arid region of Brazil (Quixaba, PE): Characterization by principal component analysis (PCA). *International Journal of Remote Sensing*, 23(23), 5005–5013. <https://doi.org/10.1080/0143116021000013313>.

Man, C. D., Nguyen, T. T., Bui, H. Q., Lasko, K., & Nguyen, T. N. T. (2018). Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers. *International Journal of Remote Sensing*, 39(4), 1243–1255. <https://doi.org/10.1080/01431161.2017.1399477>.

Manhães, A. P., Mazzochini, G. G., Oliveira-Filho, A. T., Ganade, G., & Carvalho, A. R. (2016). Spatial associations of ecosystem services and biodiversity as a baseline for systematic conservation planning. *Diversity and Distributions*, 22(9), 932–943. <https://doi.org/10.1111/ddi.12459>.

Mariano, D. A., Santos, C. A. C. dos, Wardlow, B. D., Anderson, M. C., Schiltmeyer, A. V., Tadesse, T., & Svoboda, M. D. (2018). Use of remote sensing indicators to assess effects of drought and human-induced land degradation on ecosystem health in Northeastern Brazil. *Remote Sensing of Environment*, 213, 129–143. <https://doi.org/10.1016/j.rse.2018.04.048>.

Marques, T. V., Mendes, K., Mutti, P., Medeiros, S., Silva, L., Perez-Marin, A. M., ... Bezerra, B. (2020). Environmental and biophysical controls of evapotranspiration from Seasonally Dry Tropical Forests (Caatinga) in the Brazilian Semiarid. *Agricultural and Forest Meteorology*, 287, 107957. <https://doi.org/10.1016/j.agrformet.2020.107957>.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., ... Lim, T.-K. (2006). A Landsat Surface Reflectance Dataset for North America, 1990–2000. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 68–72. <https://doi.org/10.1109/lgrs.2005.857030>.

Masek, J. G., Wulder, M. A., Markham, B., McCorkel, J., Crawford, C. J., Storey, J., & Jenstrom, D. T. (2020). Landsat 9: Empowering open science and applications through continuity. *Remote Sensing of Environment*, 248, 111968. <https://doi.org/10.1016/j.rse.2020.111968>.

Meiado, M. V., Silva, F. F. S., Barbosa, D. C. A., & Siqueira Filho, J. A. (2012). Diaspore of the Caatinga: a review. *Flora of the Caatingas of the São Francisco River: Natural History and Conservation*. Rio de Janeiro: Andrea Jakobsson Estúdio Editorial, 306-365.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Hubert-Moy, L. (2020). Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 231–256. <https://doi.org/10.1016/j.isprsjprs.2020.03.009>.

Moreira, E.; Targino, I. (1997). *Capítulos de geografia agrária da Paraíba*. João Pessoa: Editora da UFPB.

Moro, M. F., Silva, I. A., Araújo, F. S. de, Nic Lughadha, E., Meagher, T. R., & Martins, F. R. (2015). The Role of Edaphic Environment and Climate in Structuring Phylogenetic Pattern in Seasonally Dry Tropical Plant Communities. *PLOS ONE*, 10(3), e0119166. <https://doi.org/10.1371/journal.pone.0119166>.

Moro, M. F., Nic Lughadha, E., Filer, D. L., Araújo, F. S. de, & Martins, F. R. (2014). A catalogue of the vascular plants of the Caatinga phytogeographical domain: a synthesis of floristic and phytosociological surveys. *Phytotaxa*, 160-118. <https://doi.org/10.11646/phytotaxa.160.1.1>.

Moro, M. F., Silva, I. A., Araújo, F. S. de, Nic Lughadha, E., Meagher, T. R., & Martins, F. R. (2015). The Role of Edaphic Environment and Climate in Structuring Phylogenetic Pattern in Seasonally Dry Tropical Plant Communities. *PLOS ONE*, 10(3), e0119166. [doi:10.1371/journal.pone.0119166](https://doi.org/10.1371/journal.pone.0119166).

Moro, M. F., Nic Lughadha, E., de Araújo, F. S., & Martins, F. R. (2016). A Phytogeographical Metaanalysis of the Semiarid Caatinga Domain in Brazil. *The Botanical Review*, 82(2), 91–148. <https://doi.org/10.1007/s12229-016-9164-z>.

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536.

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. *International Journal of*

Applied Earth Observation and Geoinformation, 34, 136–146.
<https://doi.org/10.1016/j.jag.2014.08.001>.

Nguyen, L. H., Joshi, D. R., Clay, D. E., & Henebry, G. M. (2020). Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sensing of Environment*, 238, 111017. <https://doi.org/10.1016/j.rse.2018.12.016>.

Nosetto, M. D., Jobbágy, E. G., Brizuela, A. B., & Jackson, R. B. (2012). The hydrologic consequences of land cover change in central Argentina. *Agriculture, Ecosystems & Environment*, 154, 2–11. <https://doi.org/10.1016/j.agee.2011.01.008>.

Oliveira, S. N., de Carvalho Júnior, O. A., Gomes, R. A. T., Guimarães, R. F., & McManus, C. M. (2017). Deforestation analysis in protected areas and scenario simulation for structural corridors in the agricultural frontier of Western Bahia, Brazil. *Land Use Policy*, 61, 40-52. <https://doi.org/10.1016/j.landusepol.2016.10.046>.

Pennington, R. T., Lehmann, C. E. R., & Rowland, L. M. (2018). Tropical savannas and dry forests. *Current Biology*, 28(9), R541–R545. <https://doi.org/10.1016/j.cub.2018.03.014>.

Persson, M., Lindberg, E. & Reese, H., 2018. Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing*, 10(11), p.1794. Available at: <http://dx.doi.org/10.3390/rs10111794>.

Pettorelli, N., Safi, K., & Turner, W. (2014). Satellite remote sensing, biodiversity research and conservation of the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1643), 20130190. <https://doi.org/10.1098/rstb.2013.0190>.

Pôças, I., Calera, A., Campos, I., & Cunha, M. (2020). Remote sensing for estimating and mapping single and basal crop coefficients: A review on spectral vegetation indices approaches. *Agricultural Water Management*, 233, 106081. <https://doi.org/10.1016/j.agwat.2020.106081>.

Queiroz, L. P., Cardoso, D., Fernandes, M. F., & Moro, M. F. (2017). Diversity and evolution of flowering plants of the Caatinga domain. In: Silva J,C; Leal,I; Tabarelli, M.(Eds). *Caatinga: the largest tropical dry forest region in South America*. Switzerland: Springer International Publishing, 23–63. https://doi.org/10.1007/978-3-319-68339-3_2.

R Development Core Team. (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>. Acessado: 20 Fevereiro 2018.

Ribeiro, E. M. S., Arroyo-Rodríguez, V., Santos, B. A., Tabarelli, M., & Leal, I. R. (2015). Chronic anthropogenic disturbance drives the biological impoverishment of the Brazilian Caatinga vegetation. *Journal of Applied Ecology*, 52(3), 611–620. <https://doi.org/10.1111/1365-2664.12420>.

Rodríguez-Caballero, E., Knerr, T., & Weber, B. (2015). Importance of biocrusts in dryland monitoring using spectral indices. *Remote Sensing of Environment*, 170, 32–39. <https://doi.org/10.1016/j.rse.2015.08.034>.

Roberts, D. A., Smith, M. O., & Adams, J. B. (1993). Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, 44(2-3), 255–269. [https://doi.org/10.1016/0034-4257\(93\)90020-x](https://doi.org/10.1016/0034-4257(93)90020-x).

Santos, J. C., Leal, I. R., Almeida-Cortez, J. S., Fernandes, G. W., & Tabarelli, M. (2011). Caatinga: The Scientific Negligence Experienced by a Dry Tropical Forest. *Tropical Conservation Science*, 4(3), 276–286. <https://doi.org/10.1177/194008291100400306>.

Santos, J. M. F. F., Santos, D. M., Lopes, C. G. R., Silva, K. A., Sampaio, E. V. S. B., & Araújo, E. L. (2013). Natural regeneration of the herbaceous community in a semiarid region in Northeastern Brazil. *Environmental Monitoring and Assessment*, 185(10), 8287–8302. <https://doi.org/10.1007/s10661-013-3173-8>.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.

Schmidt, T., Schuster, C., Kleinschmit, B., & Förster, M. (2014). Evaluating an intra-annual time series for grassland classification—how many acquisitions and what seasonal origin are optimal?. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8), 3428–3439. <https://doi.org/10.1109/jstars.2014.2347203>.

Schott, J. R., Gerace, A., Woodcock, C. E., Wang, S., Zhu, Z., Wynne, R. H., & Blinn, C. E. (2016). The impact of improved signal-to-noise ratios on algorithm performance: Case studies

for Landsat class instruments. *Remote Sensing of Environment*, 185, 37–45. <https://doi:10.1016/j.rse.2016.04.015>.

Siervi, C. (2019). Descrevendo uma amostra com o Boxplot. *Medium*. <https://medium.com/@claudio.siervi/interpretando-o-diagrama-de-caixa-boxplot-1876b7c099af>. Acessado: 18 Fevereiro 2020.

Silva, P. F. da, Lima, J. R. de S., Antonino, A. C. D., Souza, R., Souza, E. S. de, Silva, J. R. I., & Alves, E. M. (2017). Seasonal patterns of carbon dioxide, water and energy fluxes over the Caatinga and grassland in the semi-arid region of Brazil. *Journal of Arid Environments*, 147, 71–82. <https://doi:10.1016/j.jaridenv.2017.09.003>.

Silva, J. N. D. C. Análise temporal de índices de vegetação no apoio à classificação de imagens: Cobertura do solo na bacia hidrográfica do rio Sucuru. 107f. Dissertação (Mestrado em Recursos Naturais) - Universidade Federal de Campina Grande, Campina Grande-PB, 2017.

Silva, M., P. L., Oliveira-Filho, A. T., Pennington, R. T., Neves, D. M., Baker, T. R., & Dexter, K. G. (2018). Using tree species inventories to map biomes and assess their climatic overlaps in lowland tropical South America. *Global Ecology and Biogeography*, 27(8), 899–912. <https://doi.org/10.1111/geb.12749>.

Silveira, H. L. F., Galvão, L. S., Sanches, I. D., de Sá, I. B., & Taura, T. A. (2018). Use of MSI/Sentinel-2 and airborne LiDAR data for mapping vegetation and studying the relationships with soil attributes in the Brazilian semi-arid region. *International Journal of Applied Earth Observation and Geoinformation*, 73, 179–190. <https://doi.org/10.1016/j.jag.2018.06.016>.

Spitzer, M., Wildenhain, J., Rappsilber, J., & Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. *Nature methods*, 11(2), 121.

Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115, 171–179. <https://doi.org/10.1016/j.compag.2015.05.001>.

Thanh Noi, P., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18.

Tian, J., & Philpot, W. D. (2015). Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra. *Remote Sensing of Environment*, 169, 280–289. <https://doi.org/10.1016/j.rse.2015.08.007>.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).

U.S. Geological Survey, 2019a. Product Guide: LANDSAT 4–7 SURFACE REFLECTANCE (LEDAPS) PRODUCT. Department of the Interior Version. vol. 8. pp. 12. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1370_L4-7_SurfaceReflectance-LEDAPS_ProductGuide-v2.pdf. Acessado: 03 Janeiro 2019.

U.S. Geological Survey, 2019b. Product Guide: Landsat 8 Surface Reflectance Code (LaSRC) Product. Department of the Interior Version. vol. 4. pp. 11. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1368_L8_SurfaceReflectanceCode-LASRC_ProductGuide-v2.pdf, Acessado: 03 Janeiro 2019.

U.S. Geological Survey, 2020. Landsat Satellite Missions. <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-satellite-missions>. Acessado: 22 Dezembro 2020.

U.S. Geological Survey, 2021a. Landsat Missions - Landsat 9. https://www.usgs.gov/core-science-systems/nli/landsat/landsat-9?qt-science_support_page_related_con=0#qt-science_support_page_related_con. Acessado: 06 de Janeiro, 2021.

U.S. Geological Survey, 2021b. What are the best Landsat spectral bands for use in my research?https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research?qt-news_science_products=0#qt-news_science_products. Acessado: 06 de Janeiro, 2021.

- Valbuena, R., Maltamo, M., & Packalen, P. (2016). Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods. *Revista de Teledetección*, (45), 15. <https://doi.org/10.4995/raet.2016.4029>.
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46–56. <https://doi.org/10.1016/j.rse.2016.04.008>.
- Vico, G., Thompson, S. E., Manzoni, S., Molini, A., Albertson, J. D., Almeida-Cortez, J. S., ... Porporato, A. (2015). Climatic, ecophysiological, and phenological controls on plant ecohydrological strategies in seasonally dry ecosystems. *Ecohydrology*, 8(4), 660–681. <https://doi.org/10.1002/eco.1533>.
- Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., & Ng, W.-T. (2018). How much does multi-temporal Sentinel-2 data improve crop type classification? *International Journal of Applied Earth Observation and Geoinformation*, 72, 122–130. doi:10.1016/j.jag.2018.06.007.
- Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment*, 112(3), 1096–1116. <https://doi.org/10.1016/j.rse.2007.07.019>.
- Walker, J., de Beurs, K., & Wynne, R. (2015). Phenological Response of an Arizona Dryland Forest to Short-Term Climatic Extremes. *Remote Sensing*, 7(8), 10832–10855. <https://doi.org/10.3390/rs70810832>.
- Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., ... Wynne, R. (2008). Free Access to Landsat Imagery. *Science*, 320(5879), 1011a–1011a. doi:10.1126/science.320.5879.1011a.
- Woodcock, C. E., Loveland, T. R., Herold, M., & Bauer, M. E. (2020). Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sensing of Environment*, 238, 111558. <https://doi.org/10.1016/j.rse.2019.111558>.
- Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., ... Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185, 271–283. <https://doi.org/10.1016/j.rse.2015.11.032>.

- Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., ... Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127–147. <https://doi:10.1016/j.rse.2019.02.015>
- Xia, J., Ning, L., Wang, Q., Chen, J., Wan, L., & Hong, S. (2017). Vulnerability of and risk to water resources in arid and semi-arid regions of West China under a scenario of climate change. *Climatic Change*, 144(3), 549–563. <https://doi.org/10.1007/s10584-016-1709-y>.
- Zhang, Z., Yang, Z., Ren, W., & Wen, G. (2019). Random forest-based real-time defect detection of Al alloy in robotic arc welding using optical spectrum. *Journal of Manufacturing Processes*, 42, 51–59. <https://doi.org/10.1016/j.jmapro.2019.04.023>.
- Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., ... Gong, P. (2016). Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. *Remote Sensing of Environment*, 183, 170–185. <https://doi.org/10.1016/j.rse.2016.05.016>.
- Zhao, Y., Wang, X., Novillo, C. J., Arrogante-Funes, P., Vázquez-Jiménez, R., & Maestre, F. T. (2018). Albedo estimated from remote sensing correlates with ecosystem multifunctionality in global drylands. *Journal of Arid Environments*, 157, 116–123.
- Zheng, B., Myint, S. W., Thenkabail, P. S., & Aggarwal, R. M. (2015). A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, 34, 103–112. <https://doi.org/10.1016/j.jag.2014.07.002>.
- Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221, 430–443. <https://doi:10.1016/j.rse.2018.11.032>.