

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CAMPUS DE CASCAVEL
CENTRO DE CIENCIAS EXATAS E TECNOLÓGICAS
PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**ENGENHARIA E EXTRAÇÃO DE RECURSOS EM SÉRIES TEMPORAIS DE ÍNDICE DE
VEGETAÇÃO PARA MAPEAMENTO DE ÁREAS COM CULTURAS ANUAIS**

LUIZ ALMEIDA

CASCAVEL – PARANÁ – BRASIL

LUIZ ALMEIDA

**ENGENHARIA E EXTRAÇÃO DE RECURSOS EM SÉRIES TEMPORAIS DE ÍNDICE DE
VEGETAÇÃO PARA MAPEAMENTO DE ÁREAS COM CULTURAS ANUAIS**

Dissertação apresentada ao programa de Pós-Graduação em Engenharia Agrícola, da Universidade Estadual do Oeste do Paraná, em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração em Engenharia de Sistema Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão.

Orientador: Prof. Dr. Erivelto Mercante
Co-Orientador: Prof. Dr. Jerry Adriani Johann

**CASCADEL – PARANÁ – BRASIL
FEVEREIRO – 2023**

Ficha catalográfica¹

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Almeida, Luiz

Engenharia e extração de recursos em séries temporais de índice de vegetação para mapeamento de áreas com culturas anuais / Luiz Almeida; orientador Erivelto Mercante; coorientador Jerry Adriani Johann. -- Cascavel, 2023.
79 p.

Dissertação (Mestrado Acadêmico Campus de Cascavel) -- Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Engenharia Agrícola, 2023.

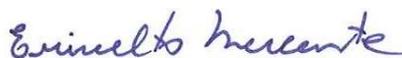
1. aprendizagem de máquina. 2. mineração de dados. 3. sensoriamento remoto. I. Mercante, Erivelto, orient. II. Johann, Jerry Adriani, coorient. III. Título.

¹ Revisor de Português, Inglês e Normas: Prof. Dr. José Carlos da Costa.

LUIZ ALMEIDA

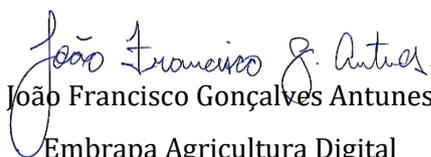
ENGENHARIA E EXTRAÇÃO DE RECURSOS EM SÉRIES TEMPORAIS DE ÍNDICE DE VEGETAÇÃO PARA MAPEAMENTO DE ÁREAS COM CULTURAS ANUAIS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:



Orientador(a) - Erivelto Mercante

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)



João Francisco Gonçalves Antunes

Embrapa Agricultura Digital



Marcio Furlan Maggi

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Cascavel, 17 de fevereiro de 2023.

BIOGRAFIA

Luiz Almeida, nascido em 28 de maio de 1991, natural de Cascavel, Paraná, Brasil, concluiu o curso de Bacharelado em Engenharia Agrícola, na área de concentração de Ciências Exatas e Tecnológicas, na Universidade Estadual do Oeste do Paraná – UNIOESTE, em agosto de 2018. Durante a graduação, participou de projetos de iniciação científica (PIBIC) como voluntário de 2015 a 2017 e, de 2017 a 2018, esteve como bolsista CNPq no Laboratório de Estatística Aplicada (LEA) sob orientação do professor Dr. Jerry Adriani Johann.

Em 2018, recebeu do Conselho Regional de Engenharia e Agronomia do Paraná (CREA-PR) o Prêmio Melhores Trabalhos de Conclusão de Curso (TCC) na área de Agrimensura, pelo trabalho intitulado: “Mapeamento e estimativa de áreas de milho 2º safra e culturas de inverno utilizando aprendizagem de máquina em imagens de satélite”, sob orientação do professor Dr. Jerry Adriani Johann.

Recebeu, no ano de 2019, do Centro de Ciências Exatas e Tecnológicas (CCET), a Menção Honrosa Acadêmica pela 3ª colocação geral entre os formandos 2018 do curso de Bacharelado em Engenharia Agrícola.

No ano de 2020 ingressou no Programa de Pós-Graduação em Engenharia Agrícola – Engenharia de Sistemas Biológicos e Agroindustriais, sob orientação do Prof. Dr. Erivelto Mercante, estudando o tema: Engenharia de recursos em séries temporais de imagens de satélite para mapeamento de culturas anuais.

*Nenhum homem é melhor que uma
máquina. E nenhuma máquina é melhor que um
homem com uma máquina.*

Richard Bookstaber

Almeida, Luiz. **Engenharia e extração de recursos em séries temporais de índice de vegetação para mapeamento de área com culturas anuais**. Orientador: Erivelto Mercante; Coorientador: Jerry Adriani Johann. 2023. 79 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel – Paraná, 2023.

RESUMO

No Brasil, as operações financeiras de crédito, custeio e seguro agrícola são as principais políticas de incentivo ao desenvolvimento da agricultura. De acordo com o Manual de Crédito Rural (MCR), a instituição financeira concedente é responsável pelo monitoramento e pela fiscalização das operações de crédito rural, autorizando a utilização de técnicas de geoprocessamento e sensoriamento remoto para essa finalidade. A aferição e validação da área agricultável do imóvel é uma das informações necessárias ao cumprimento do MCR. Essa demanda pode ser atendida com a utilização de técnicas de mineração de dados em séries temporais de índices de vegetação (STIV) para classificação de culturas anuais. O objetivo deste trabalho foi mapear e estimar áreas com culturas agrícolas anuais, utilizando mineração de dados e técnicas de engenharia e extração de recursos em séries temporais de índice de vegetação. A área de estudo compreende a região oeste do estado da Bahia, Nordeste do Brasil, devido à disponibilidade do Dataset LEM+ com classes de uso e cobertura da terra para a região. A metodologia geral segue o processo descoberta de conhecimento em banco de dados (KDD). O processo de KDD contém 5 etapas, sendo: 1ª etapa - seleção de dados com 48 imagens de índice de vegetação do sensor MODIS dos satélites Terra e Aqua, entre o período de 22/09/2019 e 21/09/2020; 2ª etapa - pré-processamento: divisão dos valores das STIV por 10.000 e empilhamento em cubo temporal de STIV; 3ª etapa - transformação: geração de 2 grupos de imagens, um com suavização da STIV com filtro Savitzky-Golay (SG) e o outro com simplificação da STIV, extraíndo sua componente de tendência (TD) com o algoritmo de Decomposição Sazonal Baseado em Perda (STL), utilizando de técnicas de engenharia e extração de recursos para construção de 25 imagens de atributos derivados (AD) em cada tratamento; 4ª etapa - mineração de dados, foram utilizadas 73 imagens (STIV + AD) de cada tratamento SG e TD e criadas 10 combinações de atributos contendo STIV, AD básicos e AD polares. O algoritmo Boruta foi utilizado para seleção de atributos com maior importância para a tarefa de classificação das culturas anuais e o classificador *Random forest* (RF) otimizado com *Grid search cross validation*, para encontrar o melhor modelo de classificação; 5ª etapa - avaliação e interpretação, foram extraídas estatísticas das 10 combinações de atributos para aferição de acurácia, Kappa e precisão do mapeamento. Os resultados mais promissores foram identificados na combinação 8, que possuía 73 atributos, sendo eles a STIV gerada com o componente de tendência, 15 AD extraídos com métricas básicas e 10 AD gerados com métricas polares, em que, com o Boruta, foram selecionados apenas 60 atributos. A máscara construída com RF mapeou 2,53 milhões de hectares com culturas anuais na região do estudo. Com superestimativa de 12,5%, em relação aos dados oficiais do IBGE a máscara obteve acurácia de 92%, Kappa de 86% e precisão aproximada de 92,2%. Do Cadastro Ambiental Rural (CAR), foi selecionado um imóvel rural e realizada uma simulação de um processo na fiscalização de crédito agrícola. Foi comparada a área de uso consolidado do imóvel, de 852,82 hectares, com casos de mapeamento da área com culturas anuais. Um dos casos obteve área mapeada de 848,145 hectares com acurácia de 99,22%, Kappa de 98,32% e precisão aproximada de 100%. Esse caso de mapeamento foi realizado aplicando-se o modelo RF treinado em imagens de STIV e AD recortadas para o imóvel rural. As técnicas aplicadas no mapeamento de áreas com culturas anuais utilizando mineração de dados, engenharia e extração de atributos em STIV permitem agilidade em processos de gestão territorial e para validação de informações sobre imóveis rurais no crédito agrícola realizado por instituições financeiras.

Palavras-chave: aprendizagem de máquina; mineração de dados; sensoriamento remoto.

ALMEIDA, Luiz. **FEATURE ENGINEERING AND FEATURE EXTRACTION IN TIME SERIES VEGETATION INDEX FOR MAPPING AREAS WITH ANNUAL CROPS**. Advisor: Erivelto Mercante. 2023. 79 p. Dissertation (Master's degree in Agricultural Engineering) – Western Paraná State University, Cascavel - Paraná, 2023.

ABSTRACT

In Brazil, financial credit operations, costing and agricultural insurance are the main policies to encourage the development of agriculture. According to the Manual de Crédito Rural [Rural Credit Manual] (MCR), the granting financial institution is responsible for monitoring and inspecting rural credit operations, authorizing the use of geoprocessing and remote sensing techniques for this purpose. The measurement and validation of the farmable area of the property is one of the information necessary for compliance with the MCR. This demand can be met with the use of data mining techniques in time series vegetation index (STIV) for the classification of annual crops. With this panorama in mind, the objective of this research was to map and estimate areas with annual agricultural crops using data mining and engineering techniques and resource extraction in time series vegetation index. The study area comprises the western region of the state of Bahia, Northeastern Brazil, due to the availability of the LEM+ Dataset with land use and land cover classes for the region. The general methodology is based on the process known as Knowledge Discovery in Databases (KDD). The KDD process contains 5 steps, as follows: 1st stage - data selection with 48 vegetation index images from the MODIS sensor of the Terra and Aqua satellites, between the period of 09/22/2019 and 09/21/2020; 2nd stage - pre-processing: division of STIV values by 10,000 and stacking in STIV temporal cube; 3rd stage - transformation: generation of 2 groups of images, one with STIV smoothing which utilized the Savitzky-Golay filter (SG), and the other with STIV simplification by extracting its trend component (TD) with the Seasonal-Trend Decomposition Procedure Based on Loess algorithm (STL), applying feature engineering and feature extraction techniques in order to build 25 images of derived attributes (AD) in each treatment; 4th stage - data mining: 73 images (STIV + AD) of each SG and TD treatment were employed, and 10 combinations of attributes were created containing STIV, basic ADs and polar ADs. The Boruta algorithm was used to select attributes with greater importance for the annual crops classification task and the Random Forest (RF) classifier was optimized with Grid Search Cross Validation, with the purpose of finding the best classification model; 5th stage - evaluation and interpretation, statistics were extracted from the 10 combinations of attributes to measure Accuracy, Kappa and Mapping Precision. The most promising results were identified in combination 8, which had 73 attributes, namely the STIV generated with the trend component, 15 AD extracted with basic metrics and 10 AD generated with polar metrics, in which, with the Boruta algorithm, only 60 attributes were selected. The mask built with RF mapped 2.53 million hectares with annual crops in the study region. With an overestimation of 12.5% in relation to official Brazilian Institute of Geography and Statistics (IBGE) data, the mask obtained an Accuracy of 92%, Kappa of 86% and Precision of approximately 92.2%. From the Cadastro Ambiental Rural [Rural Environmental Registry] (CAR), a rural property was selected, and a simulation of an agricultural credit inspection process was carried out. The area of consolidated use of the property, which had 852.82 hectares, was compared with cases of mapping the area with annual crops. One of the cases obtained a mapped area of 848.145 hectares with an Accuracy of 99.22%, Kappa of 98.32% and an approximate Precision of 100%. This mapping case was performed by applying the RF model trained on STIV and AD images cropped for said rural property. In conclusion, the techniques applied in mapping areas with annual crops using data mining, feature engineering and feature extraction in STIV allow agility in territorial management processes and for the validation of information on rural properties in agricultural credit carried out by financial institutions.

Keywords: machine learning; data mining; remote sensing.

SUMÁRIO

LISTA DE TABELAS.....	xi
LISTA DE FIGURAS	xii
1 INTRODUÇÃO	1
2 OBJETIVOS.....	3
2.1 Objetivo geral	3
2.2 Objetivos específicos.....	3
3 REVISÃO BIBLIOGRÁFICA	4
3.1 Aspectos da agricultura na mesorregião Oeste da Bahia	4
3.2 O Manual de Crédito Rural	5
3.3 Zoneamento agrícola de risco climático.....	6
3.4 Dados oficiais sobre produção agrícola e cadastro ambiental rural	7
3.5 Sensoriamento remoto na agricultura	7
3.6 Séries temporais	9
3.6.1 Decomposição de séries temporais	10
3.6.2 Dissimilaridade entre séries temporais	11
3.6.3 Séries temporais de índices de vegetação (STIV)	12
3.7 Mineração de dados e descoberta do conhecimento.....	14
3.7.1 Seleção	14
3.7.2 Pré-processamento	15
3.7.3 Transformação	15
3.7.4 Mineração de dados	18
3.7.5 Interpretação e avaliação	19
4 MATERIAL E MÉTODOS.....	21
4.1 Área de estudo.....	21
4.2 Seleção de dados.....	23
4.3 Pré-processamento	24
4.4 Transformação	26
4.5 Mineração de dados	27
4.5.1 Agrupamento de classes amostrais.....	27
4.5.2 Combinação de atributos.....	29
4.5.3 Amostragem, importância e seleção de atributos	29
4.5.4 Treinamento, otimização e seleção do modelo otimizado.....	30
4.6 Interpretação e avaliação	31
4.7 Produção do modelo otimizado	34

4.8	Geração do mapeamento com culturas anuais e extração de área mapeada	35
4.9	Dados oficiais para fins comparativos e estudo de caso.....	35
5	RESULTADOS E DISCUSSÃO	40
6	CONCLUSÕES	60
	REFERÊNCIAS	61

LISTA DE TABELAS

Tabela 1	Métricas derivadas de séries temporais com a biblioteca STMetrics	18
Tabela 2	Datas de referência das estações do ano-safra 2019/2020	24
Tabela 3	Datas das imagens MODIS contemplando o ano-safra 2019/2020, associadas às estações do período	24
Tabela 4	Combinações de atributos utilizados na <i>pipeline</i> de mineração de dados.....	29
Tabela 5	Parâmetros e valores utilizados na pesquisa de hiperparâmetros em grade	30
Tabela 6	Informações documentais constantes no demonstrativo do CAR do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.	37
Tabela 7	Importância acumulada dos atributos mais relevantes até 80%, pelo algoritmo Boruta em cada combinação de atributos.....	48
Tabela 8	Melhores hiperparâmetros identificados com <i>GridSearchCV</i> para a tarefa de classificação com RF.....	50
Tabela 9	Métricas de avaliação geradas a partir da matriz de confusão dos testes de classificação, utilizando RF em cada combinação de atributos de STIV	50
Tabela 10	Área mapeada, em hectares (ha), com distribuição municipal extraída da máscara de classificação em comparativo com dados oficiais do IBGE	54
Tabela 11	Métricas de avaliação geradas a partir da matriz de confusão dos mapeamentos do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A	57
Tabela 12	Área calculada nas máscaras mapeadas com culturas anuais, em comparativo com área de uso consolidado do imóvel.....	57

LISTA DE FIGURAS

Figura 1	Sincronização temporal dinâmica (DTW) entre 2 séries temporais.	12
Figura 2	Visão geral das etapas do processo de KDD.	14
Figura 3	Métricas fenológicas com base no ciclo da cultura.	17
Figura 4	Localização da área de estudo, municípios da mesorregião Oeste da Bahia.	22
Figura 5	Localização do Dataset LEM+ sob a área de estudo, mesorregião Oeste da Bahia.	23
Figura 6	Imagem de NDVI MODIS data 14/09/2019, recortada pela geometria da mesorregião Oeste da Bahia.	25
Figura 7	Representação do cubo temporal de imagens de satélite e da série temporal de índice de vegetação (STIV).	26
Figura 8	Sincronização temporal dinâmica (DTW) entre 2 séries temporais da classe culturas anuais.	27
Figura 9	Boxplot da distribuição dos valores de NDVI (A, C e E); boxplot da distribuição da dissimilaridade das classes amostrais Dataset LEM+ (B, D e F), para o mês de fevereiro de 2020.	28
Figura 10	Subconjuntos de instâncias derivados durante o processo de validação cruzada.	31
Figura 11	Matriz de confusão.	32
Figura 12	Avaliação da máscara de culturas anuais por inspeção visual ponto a ponto. ...	34
Figura 13	Tela de seleção do imóvel no SiCAR em São Desidério – BA.	36
Figura 14	Classes de cobertura do solo SiCAR do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.	37
Figura 15	Recorte de imagem de atributo derivado de STIV-Trend com base no limite do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.	38
Figura 16	Pontos referência de verdadeiro para as classes de culturas anuais (em verde) e outras classes (em vermelho) no imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A, sobre imagens OLI falsa-cor 564.	39
Figura 17	STIV-Raw de culturas anuais (vermelho) e após suavização com filtro SG (azul) STIV-SG.	40
Figura 18	Decomposição da STIV-SG de um <i>pixel</i> contendo culturas de ciclo anual no ano-safra 2019/2020, em componente de tendência (T), sazonal (S) e aleatório (R), utilizando o modelo STL.	41
Figura 19	Decomposição da STIV-SG de um <i>pixel</i> contendo outras classes (eucalipto) no ano-safra 2019/2020, em componente de tendência (T), sazonal (S) e aleatório (R), utilizando o modelo STL.	42

Figura 20	Representação da STIV-Trend na forma polar de <i>pixels</i> com as classes de culturas anuais (A – soja e C – algodão) e outras classes (B – eucalipto e D – pastagem).	43
Figura 21	Imagens dos atributos derivados de STIV-Trend com métricas básicas.	44
Figura 22	Imagens dos atributos derivados de STIV-Trend com métricas polares.....	45
Figura 23	Boxplot de distribuição do valor de NDVI e de distribuição da dissimilaridade das classes amostrais agrupadas em culturas anuais e outras classes.	46
Figura 24	Quantidade de atributos selecionados e descartados em cada combinação de atributos testada.	47
Figura 25	Atributos descartados pelo algoritmo Boruta em cada combinação de atributos testada na <i>pipeline</i> de mineração de dados.....	49
Figura 26	Comparativo classificação de culturas anuais (A), máscara de culturas anuais (B), após aplicação do filtro majoritário e simplificação geométrica.	52
Figura 27	Mapa com culturas anuais para o ano-safra 2019/2020 da mesorregião Oeste da Bahia.	53
Figura 28	Máscaras de classificação de área com culturas anuais no imóvel, nos 3 casos de mapeamento realizados: A – caso M1, B – caso M2; C – caso M3.....	55

1 INTRODUÇÃO

O Brasil é um dos principais produtores e exportadores de *commodities* agrícolas do mundo, principalmente soja, milho e algodão, demandando mais áreas para semeadura e incrementos na produtividade das culturas agrícolas. Isso implica um aumento da procura por parte dos produtores de condições financeiras para custear a produção, acompanhando os incrementos de área e a produtividade das culturas anuais (ARTUZO et al., 2018).

Uma das principais formas de incentivo à agricultura é o crédito agrícola. Entre outras adversidades do setor, a geração de receitas possui defasagem temporal em relação ao investimento na produção de culturas anuais (BELIK, 2014; SOUZA et al., 2015). O agricultor requer crédito para custeio da produção e a receita é gerada somente após a safra, de quatro a seis meses depois do custeio (SOUZA et al., 2015). Em 2015, o Banco Central do Brasil (BACEN) autorizou as instituições financeiras a aplicarem técnicas de sensoriamento remoto (SR) e geoprocessamento para monitoramento e fiscalização de propriedades rurais que utilizem o crédito rural para custeio agrícola, criando a Resolução CMN nº 4.427 atualizada pela Resolução CMN Nº 4.895 (BRASIL, 2021), como complemento do Manual de Crédito Rural (MCR) (BACEN, 2021).

Em uma operação financeira para crédito de custeio ou seguro agrícola, as instituições financeiras demandam de técnicas e recursos de análise de dados. Separando os dados básicos para uma operação financeira de crédito agrícola, tem-se três categorias: dados do produtor; dados internos ou de relacionamento com o cliente; dados da propriedade agrícola e da produção. Os dados do produtor, em geral, são informações documentais tanto pessoais, como Registro Geral (RG), Cadastro de Pessoa Física (CPF) e/ou Nacional de Pessoa Jurídica (CNPJ), Cadastro de Produtor Rural (CadPro), documentos comprobatórios de posse, propriedade ou arrendamento de imóvel para a finalidade financiada. As informações internas ou de relacionamento podem ser diversas e dependem do nível de detalhamento e segurança informacional de cada instituição financeira e da estratégia de gestão de relacionamento com o cliente (*Customer relationship management* – CRM). Sobre a propriedade agrícola as informações, geralmente, são o Cadastro Ambiental Rural (CAR), geolocalização dos talhões de produção; a cultura financiada; a variedade da cultura; datas de semeadura e colheita; área total a ser cultivada com os recursos e o nível tecnológico empregado na produção (BACEN, 2021).

Nesse sentido, Séries Temporais de Índice de Vegetação (STIV), obtidas de imagens de sensores a bordo de satélites, podem ser aplicadas no monitoramento de culturas agrícolas (ROUSE et al., 1974). É possível detectar mudanças no uso e cobertura da terra analisando grandes STIV, permitindo escala (quantidade) e espacialidade das informações obtidas

(JÖNSSON; EKLUNDH, 2004). Metodologias de geração de atributos (Engenharia de recursos) e extração de métricas em STIV, que auxiliam na descrição da variabilidade e fenologia da vegetação, são empregadas visando potencializar análises de STIV em processos de mineração de dados e descoberta do conhecimento (JÖNSSON; EKLUNDH, 2004; DUARTE et al., 2018; BELDA et al., 2020; FONSECA et al., 2021).

Hokama e Miranda (2019) observaram que não é vasto o ambiente literário que aborda metodologicamente as aplicações e os estudos de SR em operações financeiras, como a de crédito rural. Nesse contexto, a aplicação de técnicas de análise de dados em STIV pode auxiliar instituições financeiras na obtenção de informações de áreas agrícolas com culturas anuais, bem como em estudos e acompanhamento do uso e cobertura da terra.

A necessidade de informações ágeis e em escala com confiabilidade sobre culturas agrícolas é uma ferramenta importante no seguimento financeiro, assim como em outras áreas de estudo. Sendo assim, o objetivo deste trabalho foi mapear e quantificar área com culturas anuais, aplicando técnicas de mineração de dados e descoberta do conhecimento em séries temporais de imagens de sensores a bordo de satélites.

2 OBJETIVOS

2.1 Objetivo geral

Mapear e estimar áreas com culturas agrícolas anuais, aplicando mineração de dados e técnicas de engenharia e extração de recursos em séries temporais de índices de vegetação.

2.2 Objetivos específicos

- Aplicar técnicas de engenharia e extração de recursos em séries temporais de índices de vegetação para auxiliar na detecção de áreas com culturas anuais.
- Mapear áreas com culturas anuais na região Oeste do Estado da Bahia.
- Utilizar dados oficiais para comparativo com o mapeamento realizado.
- Comparar diferentes técnicas para mapear a área com culturas anuais em imóvel rural, simulando um processo na etapa de pré-concessão de crédito agrícola.

3 REVISÃO BIBLIOGRÁFICA

3.1 Aspectos da agricultura na mesorregião Oeste da Bahia

Com cobertura pedológica majoritária de Latossolos de textura arenosa e média, a região do oeste baiano é de alta relevância para produção agrícola (FREITAS et al., 2014). Devido às limitações de nutrientes, os solos da região demandam manejo cultural e suplementação de fertilidade para obtenção de produtividade favorável ao uso agrícola comercial, além da necessidade de provimento de água para as plantas (FREITAS et al., 2014).

O oeste da Bahia faz parte da região agrícola do MATOPIBA, formada pelos estados do Maranhão (MA), Tocantins (TO), Piauí (PI) e Bahia (BA) sob o bioma Cerrado, com topografia e clima favoráveis para o cultivo agrícola (EGER, 2021). A região tem suportado forte avanço da fronteira agrícola nacional nas últimas décadas, atendendo grande parte da produção brasileira de grãos, fibras e frutas (OLIVEIRA; PACHECO; 2020; EGER et al., 2021). Em detrimento, nas áreas de expansão agrícola a vegetação nativa foi sendo suprimida da região, ao longo do tempo (OLIVEIRA; PACHECO; 2020).

A Associação de Agricultores e Irrigantes do Oeste da Bahia (AIBA) é uma instituição não governamental com ações direcionadas ao meio ambiente, responsabilidade social, infraestrutura e fitossanidade, sendo um importante representante dos agricultores desde 1990 (OLIVEIRA et al., 2017). A AIBA realiza e disponibiliza, de maneira pública, levantamentos de safras, boletins de safra e estudos sobre as principais culturas, reunindo dados e pesquisas sobre agricultura no oeste baiano (AIBA, 2019). Devido a transformações sistêmicas a matriz produtiva do Estado da Bahia é caracterizada por grande variedade de *commodities* possibilitada pelas condições edafoclimáticas, logística e do mercado regional (AIBA, 2019).

O estado da Bahia é o maior produtor de grãos do Nordeste brasileiro, responsável por, aproximadamente, 50% de toda a produção da região (BAHIA-SEAGRI, 2021). As últimas safras de soja no oeste baiano têm expressado valores relevantes de produtividade, tendo em vista o incremento em nível tecnológico, além do manejo e da fertilidade do solo, com produção acima de seis milhões de toneladas e área colhida 1,62 milhões de hectares em 2020. Com 74 usinas de beneficiamento e área colhida com mais de 300 mil hectares, e alta qualidade do algodão produzido, a região é possuidora do selo de origem *Pure Brazil Cotton*, uma iniciativa entre produtores e a indústria têxtil (AIBA, 2019; IBGE, 2022).

A cultura do milho, além de produtiva e de valor econômico considerável, é utilizada como um agregador agrônomo na rotação de culturas. Em 2020, a área colhida foi de 204 mil hectares e a produtividade alcançou 130 sacas por hectare.

As culturas de arroz e feijão correspondem a cerca de 66 mil hectares plantados no Oeste, porém, devido ao baixo retorno econômico, a área plantada com essas culturas vem sendo reduzida. A cultura de arroz vem sendo utilizada apenas em alguns casos para aberturas de novas áreas e, o feijão, ainda mantém relevância, colocando o estado entre um dos principais produtores do Brasil, com produção acima de 83 mil toneladas (AIBA, 2019; IBGE, 2022).

3.2 O Manual de Crédito Rural

O Manual de Crédito Rural (MCR) apresenta as normas relativas ao crédito rural, aprovadas pelo Conselho Monetário Nacional (CMN) e anunciadas pelo Banco Central do Brasil (BACEN) (BACEN, 2021). O MCR 2-7 trata do monitoramento e fiscalização das operações de crédito rural, sendo a instituição financeira, cedente do crédito, responsável pelos procedimentos, conforme Resolução CMN nº 4.895, de 26 de fevereiro de 2021 e Atualização MCR nº 695, de 7 de julho de 2021 (BRASIL, 2021).

O BACEN (MCR 2-7-3-“a”) permite que a instituição financeira utilize técnicas de sensoriamento remoto na fiscalização, entre elas, por sistemas ópticos que captem o fluxo de radiação eletromagnética refletida por objetos distantes, sob a forma de imagens (BACEN, 2021). Nesse caso, a utilização de Sensoriamento Remoto Orbital (SRO) com imagens de sensores a bordo de satélite, atende às determinações do MCR.

A aplicação de técnicas de mineração de dados e descoberta do conhecimento em imagens de satélite podem auxiliar na extração de informações relevantes sobre a condução dos empreendimentos agrícolas, de maneira remota e em grande escala. A seleção de imagens de satélite para monitoramento na propriedade ou no talhão agrícola deve considerar as resoluções espacial e temporal adequadas para confirmação da cultura semeada (CAI et al., 2018).

No monitoramento devem ser extraídas informações que apresentem a aplicação dos recursos financeiros em área plantada, também quanto à confirmação da cultura semeada e/ou plantada pelo mutuário (beneficiário do crédito) constante na contratação do crédito (MCR 2-7-4-“a”-II) (BACEN, 2021). Além disso, as instituições financeiras vêm demandando informações do cadastro ambiental rural dos imóveis rurais para que seus proprietários, possuidores ou responsáveis, acessem linhas de crédito agrícola e, assim, possibilite às

concedentes do crédito que façam as devidas validações, análises e fiscalização dos empreendimentos (BRASIL, 2022).

3.3 Zoneamento agrícola de risco climático

O Brasil é um experiente formulador de políticas que tratam da variabilidade climática no setor agrícola desde 1996, com o desenvolvimento do Zoneamento Agrícola de Risco Climático (ZARC) (MERCURE et al., 2019). Essas políticas são utilizadas por diversas instituições na construção de procedimentos administrativos e financeiros, desde bancos até seguradoras e cooperativas de crédito agrícola, devido às normas de crédito rural e seguro agrícola (EMBRAPA, 1995; MERCURE et al., 2019).

O ZARC é um calendário de ano-safra utilizado em todo o país para indicação de períodos de plantio e semeadura por cultura agrícola com distribuição municipal, levando em consideração as características de clima, ciclo vegetativo de cada cultivar e tipo de solo (EMBRAPA, 1995; MERCURE et al., 2019). O objetivo do ZARC é minimizar perdas na produção agrícola, reduzindo a exposição das culturas em momentos de alta sensibilidade às adversidades climáticas, indicando o nível de risco no período (20, 30 e 40%) (EMBRAPA, 1995).

Com caracterização da aptidão agrícola municipal, o ZARC é implementado anualmente seguindo e aplicando metodologias da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), com a coordenação do Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2019), sendo publicado no Diário Oficial da União (D.O.U) em portarias. Além disso, o ZARC pode ser acessado e consultado, de maneira simplificada, nos Painéis de Indicadores Estatísticos – MAPA – no *site* oficial do Governo Federal (EMBRAPA, 1995).

Os métodos alternativos para a obtenção de informações sobre a área de cultivo, como o levantamento de campo e as estimativas com base em históricos de informações, são processos demorados e onerosos, enquanto as ferramentas de sensoriamento remoto, como as imagens de satélite, são mais objetivas e menos dispendiosas, oportunizando a repetibilidade, bem como maior chance de acerto. Assim, a utilização de imagens de satélite tem sido inestimável nos processos de tomada de decisão de governos (alívio de desastres), negócios e gestores de recursos naturais (extensão e impacto de inundações ou granizo) e produtores (agricultura de precisão) em diferentes escalas (POTGIETER et al., 2010).

3.4 Dados oficiais sobre produção agrícola e cadastro ambiental rural

Com objetivo de fornecer informações conjunturais mensais o Levantamento Sistemático da Produção Agrícola (LSPA) é realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE) desde 1972, fazendo parte do planejamento estratégico do instituto (IBGE, 2018). O LSPA visa o acompanhamento da movimentação econômica de curto prazo do Brasil, aplicando conceitos, metodologias e processos internacionalmente reconhecidos para levantamento e fornecimento de informações sobre área, produção e rendimento médio de produtos agrícolas (IBGE, 2018). O acompanhamento mensal abrange, aproximadamente, 97% da área agrícola do país, sendo realizado para 18 lavouras temporárias (culturas anuais de curta duração com ciclo inferior 1 ano civil), lavouras de longa duração (semiperenes) e as lavouras permanentes (culturas perenes). As informações são compiladas para a base de Produção Agrícola Municipal (PAM) (IBGE, 2022).

Criado pela Lei nº 12.651, de 25 de maio de 2012, o Cadastro Ambiental Rural (CAR) é um registro público eletrônico e obrigatório para todos os imóveis rurais, possuindo o objetivo de agregar as informações ambientais das posses e propriedades rurais acerca de áreas de preservação permanente (APP), de uso restrito, reserva legal (RL), remanescentes de vegetação nativa (RVN), áreas de uso consolidado, entre outras informações para acompanhamento e planejamento ambiental (SICAR, 2022). As informações para realização do CAR são autodeclaradas pelos proprietários, possuidores ou responsáveis pelos imóveis rurais, reunindo informações documentais que comprovem a propriedade e ou posse, assim como as informações georreferenciadas dos limites do imóvel, que permitam a identificação e localização correta do imóvel (BRASIL, 2022; SICAR, 2022).

O Sistema Nacional de Cadastro Ambiental Rural (SICAR) criado pelo Decreto nº 7.830, de 17 de outubro de 2012, é responsável por emitir e disponibilizar o demonstrativo do imóvel e o recibo de Inscrição do Imóvel Rural no CAR, confirmando o cadastramento documental (SICAR, 2022). As documentações disponibilizadas pelo SICAR são exigidas pelas instituições financeiras, desde 2018, para concessão de crédito agrícola em todas as modalidades com intuito de localizar o imóvel, às áreas de uso consolidado e de plantio e áreas de RL dentro do imóvel (BRASIL, 2022; SICAR, 2022).

3.5 Sensoriamento remoto na agricultura

O Sensoriamento Remoto (SR) é caracterizado pela aquisição de informações sobre um determinado (alvo) objeto ou fenômeno, sem que ocorra o contato físico ou íntimo do

dispositivo de registro. As aplicações de SR são fundamentadas na interação da Radiação Eletromagnética (REM), emitida por uma fonte de energia (normalmente sol) com um alvo de interesse (água, área urbanas, mata nativa, reflorestamento, solo etc.) (MULLA, 2013).

O monitoramento agrícola via SR é uma ferramenta que permite capturar o comportamento sazonal das culturas (FONSECA et al., 2021). Para o setor agrícola, as características de maior relevância são a espacialidade e temporalidade das informações (espaço-temporal) (WEISS; JACOB; DUVEILLER, 2020).

Os dados do sensor MODIS, a bordo dos satélites Terra e Aqua, aumentaram significativamente a disponibilidade de dados de índices de vegetação (IVs). A resolução temporal dos produtos MODIS é de 16 dias, em cada satélite, com intervalo de oito dias entre as plataformas Terra e Aqua (DUARTE et al., 2018). As vantagens de uso de imagens MODIS são a grande cobertura espacial da terra, alta resolução temporal e aquisição gratuita dos produtos, fornecendo, assim, base de dados para estudos da superfície da terra (DUARTE et al., 2018). A Embrapa Agricultura Digital disponibiliza recortes estaduais de imagens (*Raster*) MODIS em formato GeoTiff, previamente processados. Os valores reais dos índices são multiplicados por 10.000, com intervalo de dados válidos de -2.000 a 10.000, enquanto, nos dados faltantes e corpos hídricos o valor fixo é -3.000.

Produtos de SR, provenientes da matemática de bandas de imagens de satélite, foram desenvolvidos e estudados para monitorar o comportamento espectral com realce da vegetação (ROUSE et al., 1974; HUETE; JUSTICE; LIU, 1994; JUSTICE et al., 1998; DEFEO; ESQUERDO, 2016). Esses produtos são conhecidos como índices de vegetação que facilitam a identificação do padrão espectral de interação da vegetação com a radiação eletromagnética solar (HUETE; JUSTICE; LIU, 1994).

Dentre os IVs, destaca-se o índice de vegetação por diferença normalizada (*Normalized Difference Vegetation Index* – NDVI) (ROUSE et al., 1974). NDVI é matematicamente a diferença entre as bandas do infravermelho próximo (NIR) e do vermelho visível (RED) normalizado pela soma do NIR com RED, variando entre -1 e 1 (Equação 1), permitindo detectar a presença da vegetação (ROUSE et al., 1974; HUETE et al., 2002; DEFEO et al., 2016).

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \quad (1)$$

em que: ρ_{NIR} é a banda do infravermelho próximo e ρ_{Red} a banda do vermelho visível.

O NDVI é amplamente empregado em estudos e estimativas agrícolas, devido ao realce da vegetação em relação a outras classes de uso e cobertura da terra (ROUSE et al., 1974; HUETE et al., 2002; OLDONI et al., 2019). Normalmente, as culturas saudáveis têm maior refletância na banda NIR do que na banda visível. À medida que as safras se desenvolvem o valor de NDVI aumenta, atingindo valores máximos do índice próximo ao vigor pleno da cultura, o valor de NDVI diminui (AJADI et al., 2021).

Estudos recentes com aplicação SR e IVs ao monitoramento agrícola visando relacionar dados de áreas agrícolas e operações financeiras. Hokama e Miranda (2019), perceberam a eficiência do uso do EVI no monitoramento do desenvolvimento de lavouras, baseado no MCR, com intuito da mitigação de riscos em operações de crédito rural utilizando SR. Enquanto Merlo (2016), aplicando geotecnologias com uso de sensoriamento remoto na gestão de títulos e garantias no crédito agrícola, obteve êxito nas aplicações de NDVI para monitoramento de matrículas agrícolas (propriedades rurais).

Culturas com cultivo anual, como milho e soja são facilmente identificadas quando se observa o comportamento cíclico do NDVI (MERLO, 2016), ou seja, observações feitas em séries temporais de NDVI e suas relações com o desenvolvimento fenológico com culturas anuais aplicando geotecnologias e SR para análise e gestão de informações. Essas relações estabelecidas podem ser aplicadas diretamente em outras áreas a fim de determinar classes de uso e cobertura da terra, reduzindo consideravelmente a demanda por inspeções a campo (*in loco*), por mão de obra e recursos financeiros (LIEN; DOLGORSUREN, 2021).

3.6 Séries temporais

Uma série temporal é uma sequência de observações indexadas pelo tempo de forma contínua ou discreta (SILVA et al., 2018), podendo ser descrita matematicamente por dois tipos de modelos: o modelo aditivo e o multiplicativo, formados pelos componentes de tendência (T), sazonalidade (S) e variações aleatórias ou irregulares (R) (HYNDMAN; ATHANASOPOULOS, 2018; SILVA et al., 2018; BERVEGLIERI et al., 2021).

O modelo aditivo (Equação 2) é aplicável quando a magnitude das flutuações sazonais ou as variações em torno do ciclo de tendência não ocorrem com a mesma intensidade na série temporal (HYNDMAN; ATHANASOPOULOS, 2018). Por outro lado, o modelo multiplicativo (Equação 3) é mais apropriado quando a variação no ciclo de tendência ou no padrão sazonal tem maior proporcionalidade em relação ao nível de base (média) da série temporal (HYNDMAN; ATHANASOPOULOS, 2018).

Os componentes de séries temporais podem ser estudados individualmente para detectar singularidades ou comportamentos em um intervalo de tempo, permitindo também a previsão de ocorrências e variáveis futuras com base na análise histórica (SILVA et al., 2018). A análise do componente de tendência permite a compreensão dos efeitos de longo prazo, permitindo identificar eventos de perturbação nas séries (BERVEGLIERI et al., 2021; SILVA et al., 2018), podendo caracterizar o formato e direção dos eventos observados com base em

sua evolução temporal. O componente de sazonalidade descreve padrões cíclicos que variam próximos à linha de base da série temporal e se completam ao longo do tempo.

Em geral, nos estudos de séries temporais é importante identificar e realizar o isolamento de efeitos sazonais (SILVA et al., 2018). Variações irregulares são mudanças não previstas em relação à linha de tendência, sendo comportamento característico do componente aleatório da série temporais, decorrentes de eventos extremos ou inesperados (SILVA et al., 2018), como incêndio, desmatamento ou geadas no caso de séries temporais de índices de vegetação, comumente referido como ruído.

$$Y_t = T_t + S_t + R_t \quad (2)$$

$$Y_t = T_t * S_t * R_t \quad (3)$$

em que: Y_t : é o dado observado no tempo t ; T_t : é o componente de tendência; S_t : é o componente sazonal; R_t : é o componente aleatório.

3.6.1 Decomposição de séries temporais

O estudo individual dos componentes derivados de séries temporais é conhecido como decomposição de séries temporais (CLEVELAND et al., 1990; MILLARD et al., 2022). A decomposição é uma técnica estatística que pode extrair cada componente de uma série temporal (MILLARD et al., 2022). Uma das técnicas utilizadas é decomposição de tendência sazonal, que permite a identificação de padrões e extração de informações sobre os eventos descritos pelas séries temporais (HE; ZHANG; CHEW, 2022; MILLARD et al., 2022).

Outro procedimento para decomposição de tendência sazonal é a técnica não paramétrica baseada em perda (*Seasonal-trend decomposition procedure based on Loess – STL*) consiste em uma cadeia de operações de suavização desenvolvido por Cleveland et al. (1990). STL suaviza a série temporal utilizando técnicas de regressão localmente ponderadas ou regressões de perda (CLEVELAND et al., 1990).

A suavização por regressão de perda (*Loess*) é um método para estimativa de relações não lineares. A *Loess* é realizada em todos os pontos do eixo da variável dependente (Y) com base no eixo da variável independente (X), ou seja, a estimativa é realizada em todos os pontos da série temporal. Essa característica permite que o STL atue em dados ausentes, além de reduzir diretamente o componente de sazonalidade (CLEVELAND et al., 1990). O STL pode ser aplicado em séries temporais com sazonalidade variável, pois o método permite o ajuste da taxa de ocorrência sazonal (LU et al., 2003). O procedimento é robusto para dados discrepantes, fazendo com que as ocorrências não afetem as estimativas dos componentes

sazonais e do ciclo de tendência, nesse caso a discrepância será atrelada ao componente aleatório (CLEVELAND et al., 1990; HYNDMAN; ATHANASOPOULOS, 2018).

A biblioteca *Python Statsmodel* disponibiliza o modelo de decomposição STL (PERKTOLD; SEABOLD; TAYLOR, 2019). O modelo permite a “quebra” de uma série temporal complexa e irregular em suas três componentes T, S e R (CLEVELAND et al., 1990 e LU et al., 2003). Além de eficiência computacional as decomposições resultantes do modelo permitem a geração de visualizações gráficas simplificadas, facilitando a associação com o comportamento da vegetação (LU et al., 2003).

3.6.2 Dissimilaridade entre séries temporais

A sincronização temporal dinâmica (*Dynamic Time Warping* - DTW) é uma técnica de comparação de séries temporais, comprimindo ou estendendo localmente pontos de uma série temporal em relação a outra, possibilitando encontrar séries com menor distância entre si (GIORGINO, 2009). O algoritmo DTW otimiza o sincronismo ideal entre séries temporais aplicando cálculos de menor distância entre pontos, realizando a soma cumulativa das distâncias, sendo reconhecida como distância DTW ou medida de dissimilaridade (GIORGINO, 2009). A dissimilaridade entre séries temporais pode ser interpretada como uma medida de heterogeneidade das séries, ou seja, de menor semelhança entre elas (TORMENE et al., 2008).

Tormene et al. (2008) relatam sobre a popularidade e aplicações do algoritmo DTW, desde 1970, em aplicações de mineração de dados para agrupamento e classificação em diversas áreas de domínio. Devido a sua capacidade de processar distorções temporais, tornou-se uma ferramenta importante na comparação de séries temporais (ZHANG et al., 2021). A biblioteca para linguagem, *Python DTAIDistance* versão 2.3.X, implementa o algoritmo DTW de maneira otimizada para comparação entre duas séries temporais do tipo vetor (em duas dimensões) ou organizadas de modo matricial (MEERT et al., 2022). Na Figura 1 é possível verificar a sincronização entre duas séries temporais calculadas com o algoritmo DTW, realizada na biblioteca Python (MEERT et al., 2022).

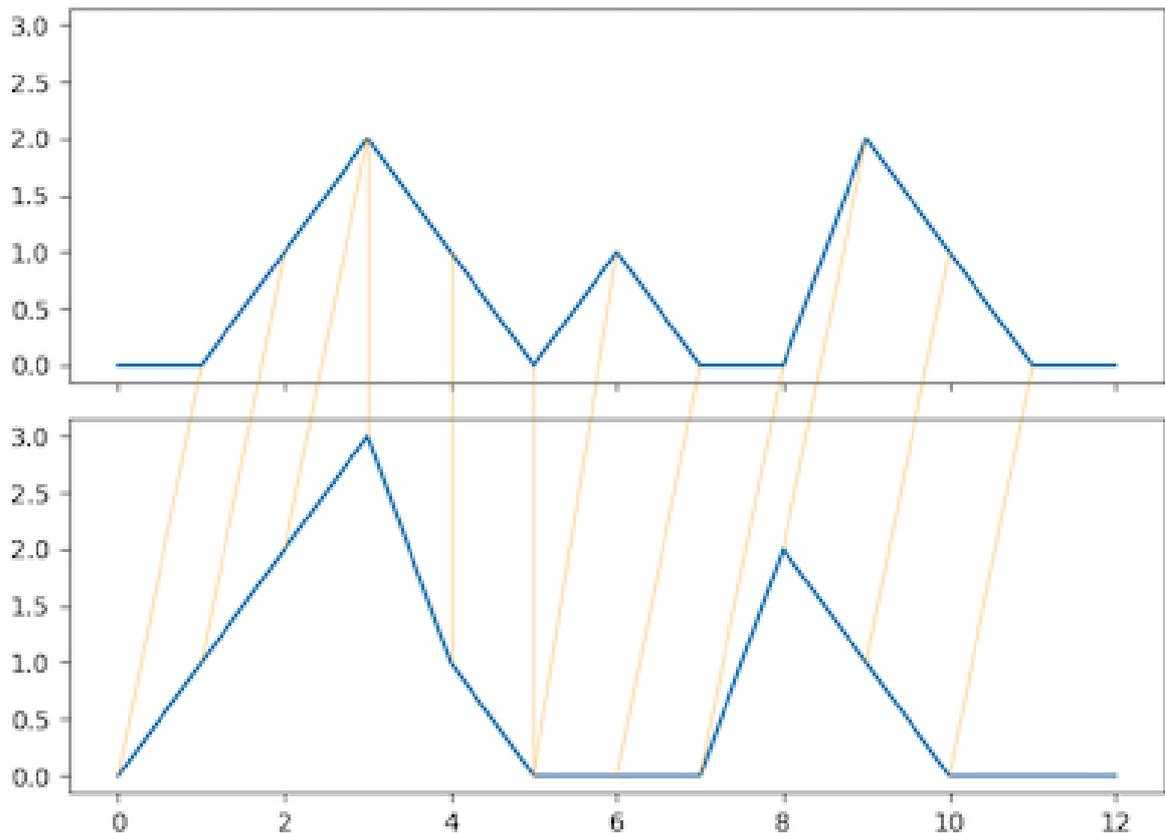


Figura 1 Sincronização temporal dinâmica (DTW) entre 2 séries temporais.

Fonte: Meert et al. (2022).

3.6.3 Séries temporais de índices de vegetação (STIV)

A utilização de STIV para monitoramento de culturas anuais é uma estratégia que reúne informações espectro-temporais do mesmo alvo geográfico, com repetição temporal do valor espectral ou do IV da classe, aumentando a precisão das predições, isso deve-se as diferenças de sazonalidade no plantio/semeadura e colheita das culturas (CAI et al., 2018). Quando as mudanças de UCT ocorrem em várias escalas de tempo, as análises em STIV demandam de métodos que descrevam as variações em escala sazonal, detectando mudanças em qualquer tendência de longo prazo (WATTS; LAFFAN, 2014; BERVEGLIERI et al., 2021).

Enquanto, o uso de métodos tradicionais de mapeamento utilizando imagens de data única podem apresentar erros elevados e baixa precisão nas classificações, devido similaridade de resposta espectral em diferentes classes de uso e cobertura da terra (LIEN; DOLGORSUREN, 2021; ESQUERDO et al., 2020). Um exemplo de utilização de STIV em estudos de análises de UCT incluem Cai et al. (2018), que verificaram mais diferenças entre soja e milho no início (fase de crescimento) e fim (senescência) do ciclo vegetativo das

culturas, notando o potencial de contribuição das STIV na classificação de culturas anuais (soja e milho), durante a classificação de alto desempenho utilizando séries temporais e aprendizagem de máquina.

O Sistema de Análise Temporal da Vegetação (SATVeg), uma plataforma web que disponibiliza STIV para a análises de UCT na América do Sul, permitindo ao usuário a visualização e aplicação de procedimentos de filtragem em STIV (ESQUERDO et al., 2020). Ajadi et al. (2021), no mapeamento de área cultivada e tipo de cultura em larga escala no Brasil, aplicaram de STIV de NDVI para extração de dados e recursos (métricas estatísticas), determinação dos estágios de crescimento e simplificação da dinâmica de crescimento das culturas anuais.

Ocorrência de nebulosidade, dispersantes atmosféricos, a calibração de equipamento e outros agentes podem causar ruídos em STIV, ocorrendo distorção entre valores reais e coletados. Os métodos de filtragem apresentam potencial para suavizar ruídos e reconstruir dados com maior qualidade em séries temporais com intervalo de tempo irregular (WU et al., 2021; XUN et al., 2021).

O uso de métricas extraídas de STIV pode ser uma técnica estratégica nos métodos de classificação para mapeamento agrícola, com potencial para mapear sistemas de cultivos múltiplos em regiões de alta variabilidade de culturas, em regiões agrícolas altamente dinâmicas (BENDINI et al., 2019). Além disso, as métricas de STIV tem potencial para integrar características de múltiplas safras, que podem ser relevantes para aprimorar os detalhes temáticos no mapeamento agrícola (RUFIN et al., 2015).

As métricas espectro-temporais são ferramentas para captura de informações fenológicas interessantes podendo ser analisada desde sazonais a plurianuais, permitindo ampla aplicação para caracterização de sistemas de uso e cobertura da terra no espaço e tempo (MÜLLER et al., 2015). O conhecimento dos componentes de uma série temporal possibilita a identificação e estudo dos padrões e processos evolutivos de um bioma (SILVA et al.; 2018).

Bendini et al. (2009) utilizaram-se de imagens serie temporais densas baseado em métricas estatísticas para descrever padrões fenológicos de ambiente agrícola no Cerrado brasileiro, nesse contexto, é importante ressaltar a complexidade e a dinâmica deste tipo de ambiente, no entanto, foram capazes de mapear a distribuição de culturas nessas áreas com acurácias acima de 90% (BENDINI et al., 2019).

Em geral, a análise de STIV requer o manuseio de um grande volume de dados, demandando de alta capacidade de processamento computacional. Linguagens de programação e outras ferramentas computacionais modeláveis para automatização e processamento são normalmente empregadas para essa tratativa (ESQUERDO et al., 2020).

3.7 Mineração de dados e descoberta do conhecimento

Neste estudo, a mineração de dados e descoberta do conhecimento (MDDC) trata da utilização de algoritmos de aprendizagem de máquina (AM) para ajuste de modelos de determinação de padrões em STIV com a tarefa de classificar culturas anuais (FAYYAD; STOLORZ, 1997). A mineração de dados é uma importante etapa do processo de descoberta do conhecimento em banco de dados (*Knowledge Discovery in Databases – KDD*) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a). O processo de KDD consiste na transformação ágil e automatizada de grandes bancos de dados brutos em conhecimento útil, podendo ser descrito em 5 etapas distintas e iteráveis: 1 – seleção; 2 – pré-processamento; 3 – transformação; 4 – mineração de dados; 5 – interpretação e avaliação (Figura 2).

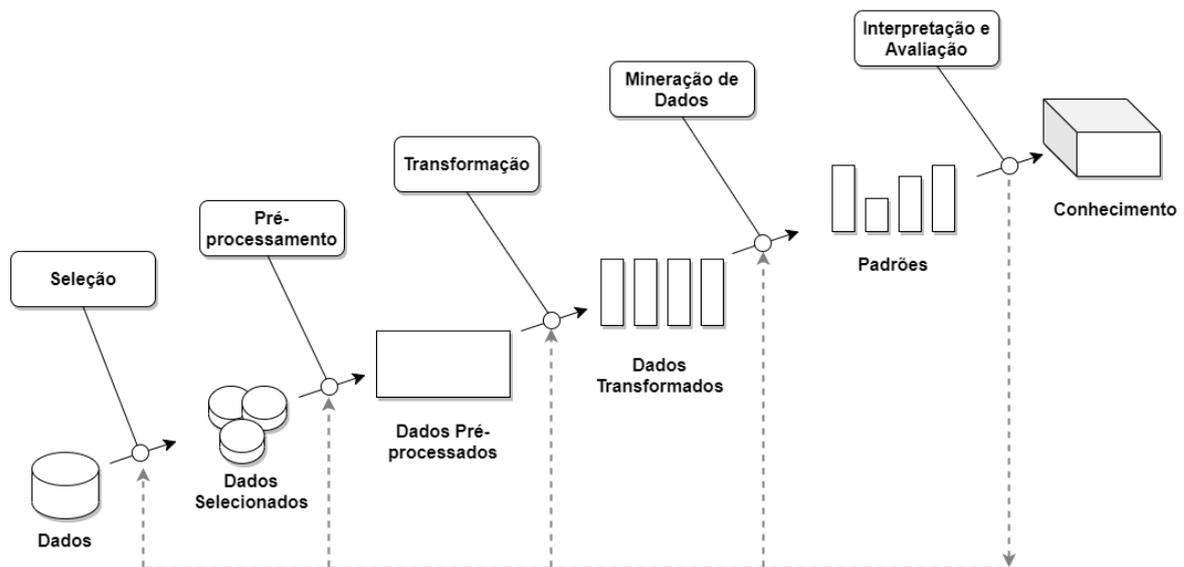


Figura 2 Visão geral das etapas do processo de KDD.

Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996a).

3.7.1 Seleção

Etapa para o entendimento dos objetivos da aplicação e definição de hipóteses iniciais, seleção do conjunto de dados que representam o problema e as variáveis/atributos que podem auxiliar no processo de MDDC (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b). Também é importante definir as métricas de performance do modelo obtido e da performance de negócio, otimizando as operações dentro do processo de MDDC.

A seleção de atributos é necessária para a redução de dimensionalidade do problema a ser resolvido, tendo em vista que, a maioria dos algoritmos de mineração de dados são

desenvolvidos para aplicações em “colunas”, características, variáveis ou atributos (LAKSHMIPADMAJA; VISHNUVARDHAN, 2018). À medida que o conjunto de dados aumenta, os custos computacionais de processamento também aumentam, assim como as etapas do *pipeline* e a complexidade dos modelos resultantes do processo, necessitando de operações para melhoria de *performance* (LAKSHMIPADMAJA; VISHNUVARDHAN, 2018).

O método *wrapper* Boruta foi desenvolvido com base no classificador RF (KURSA; RUDNICK, 2010). Esse método utiliza árvores de decisão fracas, imparciais e independentes para classificar diferentes subamostras de treinamento por atributo, avaliando a precisão dos resultados. A redução de precisão da classificação é utilizada como medida de importância de um atributo, devido à aleatoriedade da permutação de valores nos atributos em cada ocorrência (KURSA; RUDNICK, 2010).

3.7.2 Pré-processamento

Consiste na estratégia de preparação ou adequação do conjunto de dados para a tarefa de mineração a ser realizada como o preenchimento de dados faltantes, remoção de ruídos e tratamento de *outliers*, uniformização, codificação de caracteres e integração entre diferentes fontes de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a; LEPRINCE; MILLER; ZEILER, 2021). A biblioteca Python Scikit-Learn realiza a codificação de maneira nativa e automática em tarefas de classificação de dados categóricos (PEDREGOSA et al., 2011).

3.7.3 Transformação

São técnicas matemático-estatísticas aplicadas em conjuntos de dados com intuito de aumentar a capacidade dos algoritmos de mineração de dados em reconhecer padrões. As técnicas de transformação de dados dependem da tarefa de mineração de dados a ser realizada (LEPRINCE; MILLER; ZEILER, 2021). Em geral, técnicas de transformação de dados visam à construção de um conjunto de dados que podem ser utilizados complementando o conjunto original (aumentando a dimensionalidade) ou analisado isoladamente (reduzindo a dimensionalidade de atributos).

A discretização, normalização, agregação de atributos e suavização são métodos que podem ser aplicados para transformação de dados (LEPRINCE; MILLER; ZEILER, 2021). É possível que sejam empregados outros algoritmos mineração de dados nessa etapa, com

tarefas específicas, objetivando à geração de recursos de aprendizagem (*Learning features*) (BUDNIK et al., 2017).

3.7.3.1 Extração e engenharia de recursos

As técnicas de extração de recursos (*Feature extraction*) não demandam de um especialista de domínio da aplicação, tendo se popularizado pela simplicidade da técnica (BUDNIK et al., 2017; HOPPE et al., 2019). Enquanto, técnicas de engenharia de recursos (*Feature engineering*) em dados geralmente é utilizada por especialistas de domínio da aplicação ou do negócio (HOPPE et al., 2019). Técnicas de extração de recursos podem ser utilizadas com o objetivo de reduzir a dimensionalidade, construindo um novo conjunto de recursos partindo do conjunto de dados originais (LAKSHMIPADMAJA; VISHNUVARDHAN, 2018).

Engenharia de recursos é a técnica de gerar novos atributos para um conjunto de dados, utilizando regras baseadas em leis físicas e experimentação (HOPPE et al., 2019). A engenharia de recursos pode ser inserida na etapa de transformação do processo de KDD. Em mineração de dados para aplicações e estudos do uso e cobertura da terra as regras para engenharia de recursos podem ser geradas a partir de observações da variação temporal de IVs. No mapeamento anual de áreas agrícolas, utilizando mineração de dados e imagens OLI Landsat-8, Oldoni et al. (2019) criaram um atributo capaz de diferenciar alvos permanentes, como áreas urbanas, reflorestamento e floresta de culturas anuais. O atributo desenvolvido por Oldoni et al. (2019) é o somatório das diferenças de NDVI em uma ST, com relação direta, para alvos permanentes o NDVI apresenta pouca variação e em culturas anuais variação significativa ao longo da ST.

O *software* TIMESAT, desenvolvido para analisar séries temporais de imagens de sensores a bordo de satélites, realiza a engenharia de recursos para geração de atributos com base na sazonalidade (safra e/ou ciclo da cultura) (JÖNSSON; EKLUNDH, 2004). O TIMESAT aplica estratégias de suavização dos dados (filtragem) para redução de ruídos, devido às características de origem dos dados. O *software* realiza o processamento das STIV utilizando um método de ajuste de mínimos quadráticos (*Least-Squares Fitting*) (JÖNSSON; EKLUNDH, 2004), permitindo a extração de 11 recursos da STIV (métricas fenológicas). Na Figura 3, podem ser verificadas as métricas fenológicas extraídas pelo TIMESAT; as marcações (a, b, c, d, e, f, g, h e i) indicam diretamente 9 recursos; outros 2 recursos são a derivada de (a)(c), taxa de crescimento, e a derivada de (d)(b), taxa de senescência.

A biblioteca para linguagem Python STMetrics permite a extração de recursos básicos (métricas básicas) e a engenharia de recursos da forma polar de séries temporais (métricas polares) (SOARES et al., 2020). As métricas básicas são calculadas utilizando abordagens de estatística descritiva comum em série temporal, permitindo extrair 15 métricas, enquanto

as 10 métricas polares são calculadas a partir da projeção polar da série temporal (KÖRTING; FONSECA; CÂMARA, 2013; SOARES et al., 2020).

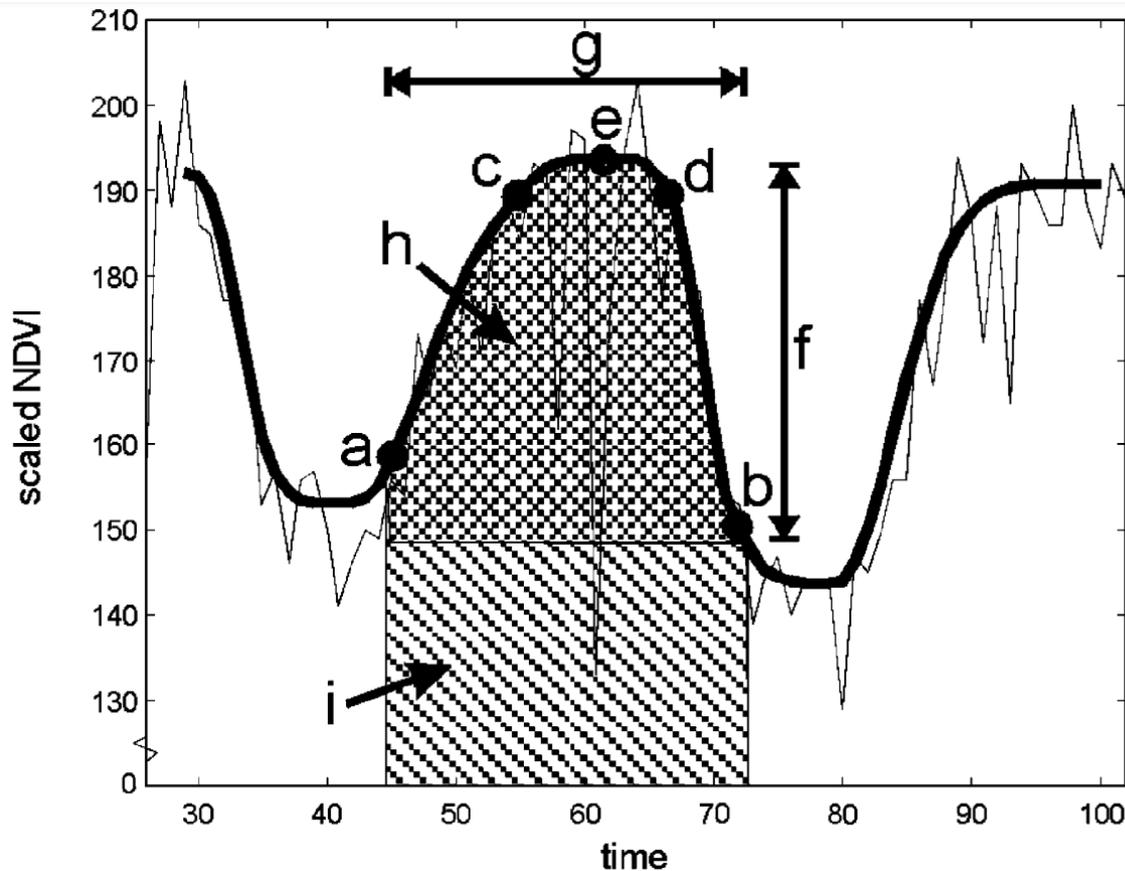


Figura 3 Métricas fenológicas com base no ciclo da cultura.

Nota: (a) início do ciclo da cultura; (b) final do ciclo da cultura; (c) e (d) representa a distribuição de dados a 90%, à esquerda e à direita, respectivamente; (e) valor máximo do ciclo; (f) amplitude; (g) duração do ciclo da cultura (dias); (h) produtividade total – integral maior e (i) produtividade do ciclo –integral menor.

Fonte: Adaptado de Jönsson e Eklundh (2004).

A engenharia de recursos realizada para as métricas polares é baseada em Körting, Fonseca e Câmara (2013), em que a série temporal de 1 ciclo é projetada de maneira fechada em um plano cartesiano, num intervalo de 0 a 2π . A representação de ciclo pode ser associada à recorrência de um evento, no caso, o período de 1 ano, permitindo a comparação de observações constantes no tempo a circunferência e seus quadrantes (KÖRTING; FONSECA; CÂMARA, 2013). Na Tabela 1 é possível verificar a descrição das métricas básicas e polares disponíveis na biblioteca STMetrics.

Tabela 1 Métricas derivadas de séries temporais com a biblioteca STMetrics

Tipo da métrica	Métrica	Descrição
Básica	Abs Sum	Soma absoluta dos valores ao longo da série temporal
	AMD	Derivada média absoluta
	Amplitude	A diferença entre os valores máximo e mínimo da série temporal
	First Slope	Valor máximo da primeira inclinação do ciclo
	FQR	Primeiro quartil da série temporal
	IQR	Intervalo interquartil da série temporal
	Max	Valor máximo da série temporal
	Mean	Valor médio da série temporal
	Min	Valor mínimo da série temporal
	MSE	Energia Espectral Média
	Skew	Mede a assimetria da série temporal
	Sum	Soma dos valores ao longo do ciclo
	SQR	Segundo quartil da série temporal
	Std	Desvio padrão da série temporal
	TQR	Terceiro quartil da série temporal
Polar	Angle	Ângulo principal da ST fechada na forma polar
	Area Q1	Área da forma polar sobre o primeiro quadrante
	Area Q2	Área da forma polar sobre o segundo quadrante
	Area Q3	Área da forma polar sobre o terceiro quadrante
	Area Q4	Área da forma polar sobre o quarto quadrante
	Area TS	Área total da forma polar da ST
	CSI	Índice de forma celular - medida quantitativa adimensional de morfologia, desvio padrão de um objeto em relação a um círculo
	ECC	Excentricidade, valores próximos a 0 maior a semelhança ao círculo, próximos a 1 menor semelhança
	Gyration Rad	Distância média dos pontos ao centroide da forma
	Polar Balance	Equilíbrio polar, desvio padrão das áreas por estação, considerando 4 estações

3.7.4 Mineração de dados

A mineração de dados é a etapa na qual são construídos os modelos de geração de conhecimento, ajustando algoritmos de reconhecimento de padrões e aprendizagem de máquina (*Machine learning*). Um modelo de mineração de dados é gerado a partir da parametrização de um algoritmo aplicado em um conjunto de dados extraindo características significativas (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b). Dentre outras tarefas do processo de KDD, a classificação é uma tarefa preditiva que treina algoritmos para atribuir a classe de uma ocorrência (objeto ou instância) em um conjunto de classes pré-estabelecidas (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b).

A MDDC em imagens de satélite permite a análise de grande volume de dados, contemplando faixas multiespectrais, extensão territorial e/ou temporalidade, sendo uma importante ferramenta nas aplicações e estudos sobre o uso e cobertura da terra (FAYYAD; STOLORZ, 1997; LIEN; DOLGORSUREN, 2021). A MD em STIV pode fornecer uma

compreensão ágil com melhor precisão na distinção de classes de uso e cobertura da terra, possibilitando a criação de processos e modelos replicáveis.

O desenvolvimento do fluxo de processos para mineração de dados (*Pipeline*) deve ser planejado e melhorado continuamente, permitindo que os algoritmos de mineração tenham alto desempenho e escalabilidade (LEPRINCE; MILLER; ZEILER, 2021). Um *pipeline* é o empacotamento ou concentração de uma sequência de procedimentos que auxiliam na definição de parâmetros, extração de recursos, redução de dimensionalidade, treino, teste e replicação de operações unitárias em um processo (BUITINCK et al., 2013).

Os algoritmos mineradores devem ter capacidade de escalar de acordo com o crescimento do banco de dados (FAYYAD; STOLORZ, 1997). Ainda, Fayyad e Stolorz (1997) indicam que devem ser utilizadas estratégias eficientes de amostragem para que possam ser criados algoritmos de baixa complexidade (linear ou sublinear), melhorando a compreensão e escala.

Random Forests (RF) é um estimador baseado em árvores de decisão que realiza amostragens aleatórias com substituição no conjunto de dados de treinamento, criando diversos estimadores com reposição (Método *bootstrap*) (PEDREGOSA et al., 2011, WU et al., 2021). Proposto por (BREIMAN, 2001) o RF realiza diversas subamostragens no conjunto de dados, criando várias árvores de decisão para classificação, ajustando a precisão das predições e controlando ajustes excessivos do modelo (decorar). O RF tem menor erro de generalização, sendo robusto a ruídos e pequenas discrepâncias, devido à convergência do modelo para evitar problemas de sobreajuste (*Overfitting*) (BREIMAN, 2001; PEDREGOSA et al., 2011). Durante a criação das árvores o estimador encontra os valores que melhor as dividem.

Na tarefa de classificação supervisionada, onde existe o processo de treinamento prévio do modelo, as árvores do RF realizam um processo de decisão em cada classe de treinamento, em seguida a classe com maior pontuação é rotulada ao dado amostral (BREIMAN, 2001; PEDREGOSA et al., 2011; BARGIEL, 2013). De acordo com Pedregosa et al. (2011), a biblioteca *Scikit-Learn* para linguagem Python possui uma implementação do algoritmo de classificação RF que combina estimadores pela média das previsões probabilísticas, em contraste com a descrição de Breiman (2001), na qual cada estimador elenca uma classe.

3.7.5 Interpretação e avaliação

A interpretação é a abstração dos resultados dos modelos de mineração de dados para a compreensão do especialista, etapa em que se deve examinar os padrões extraídos, o que é um conhecimento descoberto. A avaliação é o processo que documenta as

potencialidades da MDDC e informações não triviais obtidas. Nessa etapa, também é importante que o especialista entenda a necessidade de retornar a alguma etapa anterior para iterar novamente, buscando melhores resultados, compreensão ou simplicidade do modelo final (FAYYAD; STOLORZ, 1997). Na documentação é importante conter as definições iniciais, hipóteses, métricas ou indicadores de desempenho, para comparação com os resultados alcançados.

4 MATERIAL E MÉTODOS

A metodologia geral do trabalho foi conduzida de acordo com o processo de Descoberta do Conhecimento em Banco de Dados (*Knowledge Discovery in Databases - KDD*) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a), que apresenta as seguintes etapas: 1 - seleção de dados, 2 - pré-processamento, 3 - transformação, 4 - mineração de dados e 5 – interpretação e avaliação.

Utilizou-se um processador Intel Core i5–10210U (1.60GHz) com 8GB de RAM 64 bits. Os *softwares* utilizados foram o sistema operacional *Windows* 10, QGIS 3.16.5 para produção de mapas temáticos e a linguagem Python 3.8.5 para o processamento de dados. As bibliotecas Python adicionais utilizadas foram Boruta (KURSA; RUDNICK, 2010), GDAL, GeoPandas, Matplotlib, Numpy, Scikit-Learn, SciPy, Spyder e STMetrics (SOARES et al., 2020). Adequações no código-fonte da biblioteca STMetrics foram necessárias para melhoria de performance e integração de dados, devido aos cálculos matriciais e multitemporais aplicados, conforme Licença MIT - Instituto Nacional de Pesquisas Espaciais - INPE 2019©.

4.1 Área de estudo

A área de estudo possui 117.395.800 hectares (ha), compreendendo a mesorregião Oeste da Bahia, Nordeste do Brasil, com 24 municípios, entre as latitudes 10°00'S e 15°30'S e longitudes 43°30'W e 46°30'W (Figura 4). A mesorregião de estudo é caracterizada por dois tipos de biomas, sendo ocupada em 17,1% pela Caatinga e 82,9% de predominância do bioma Cerrado (IBGE, 2021).

O conjunto de dados (*Dataset*) LEM+: para aplicações de sensoriamento remoto agrícola (OLDONI et al., 2020) (Figura 5) possui informações mensais sobre 16 classes de UCT, entre outubro de 2019 e setembro de 2020 (um ano-safra) com 1854 instâncias de campos georreferenciados (OLDONI et al., 2020).

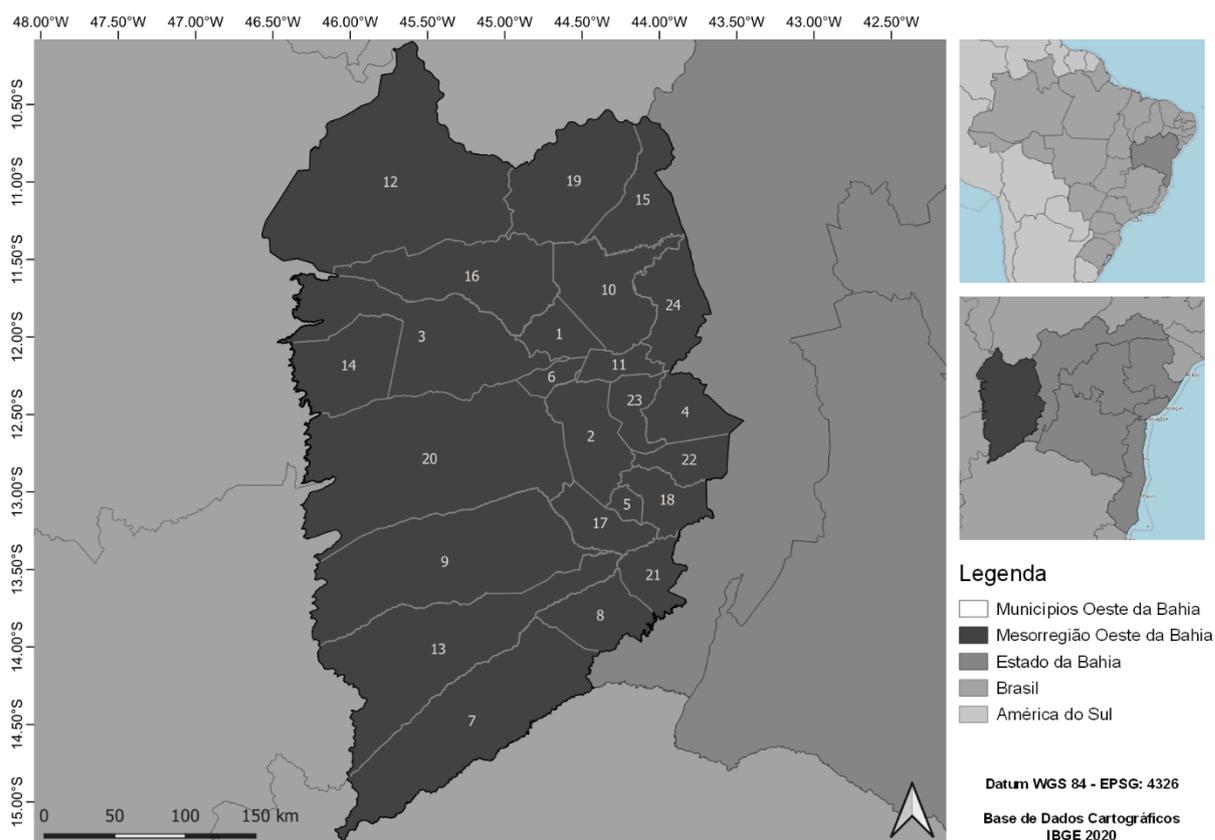


Figura 4 Localização da área de estudo, municípios da mesorregião Oeste da Bahia.

Nota: 1 – Angical, 2 – Baianópolis, 3 – Barreiras, 4 – Brejolândia, 5 – Canápolis, 6 – Catolândia, 7 – Cocos, 8 – Coribe, 9 – Correntina, 10 – Cotegipe, 11 – Cristópolis, 12 – Formosa do Rio Preto, 13 – Jaborandi, 14 – Luís Eduardo Magalhães, 15 – Mansidão, 16 – Riachão das Neves, 17 – Santa Maria da Vitória, 18 – Santana, 19 – Santa Rita de Cássia, 20 – São Desidério, 21 – São Félix do Coribe, 22 – Serra Dourada, 23 – Tabocas do Brejo Velho, 24 – Wanderley.

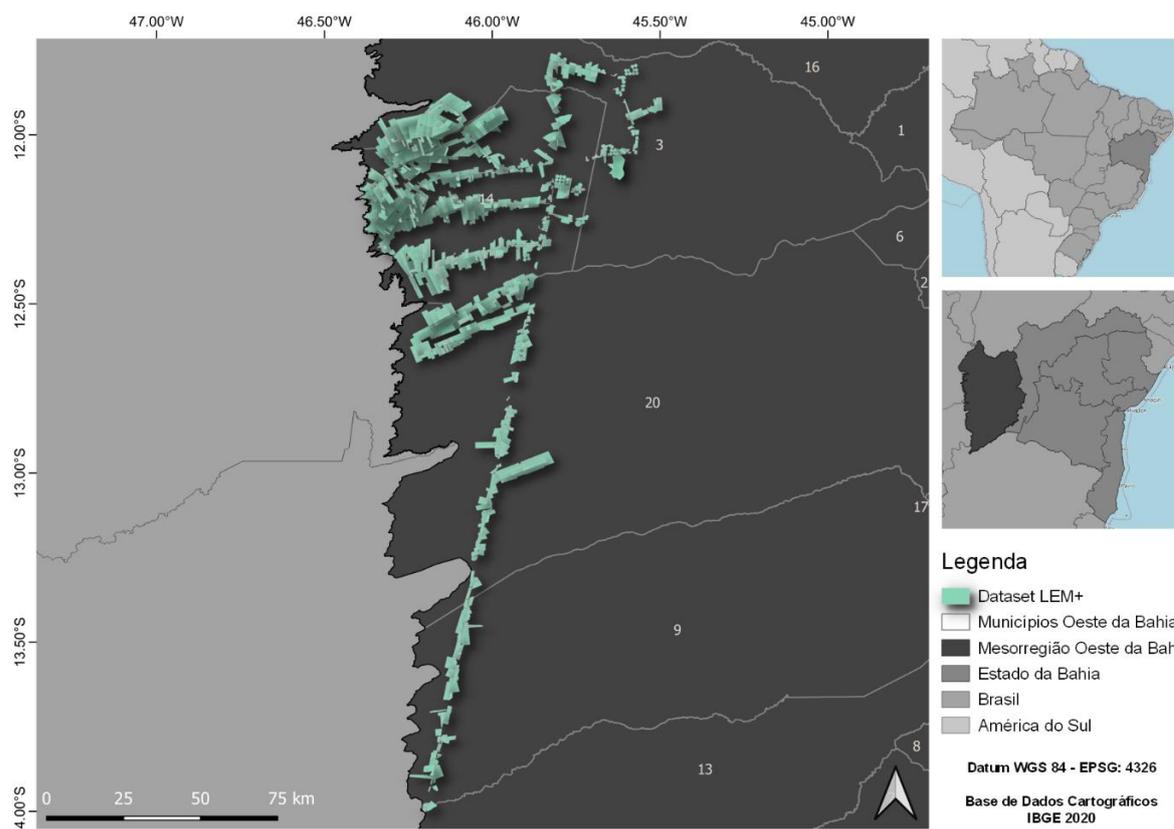


Figura 5 Localização do Dataset LEM+ sob a área de estudo, mesorregião Oeste da Bahia.

Fonte: Adaptado de Oldoni et al. (2020).

4.2 Seleção de dados

Para a verificação do período de campo das culturas utilizou-se o ZARC, selecionando no Dataset LEM+ o atributo “Feb_2020”, que representa as características de uso e cobertura da terra (UCT) do mês de fevereiro de 2020, para geração das amostras de treinamento e validação do estudo. A seleção do atributo apresentou 14 classes únicas de UCT.

Para padronização do intervalo da série temporal e seleção de imagens de NDVI, utilizou-se uma abordagem baseada na referência das estações do ano-safra 2019/2020 (Tabela 2). A estratégia padronizada para seleção de datas de imagens derivou de Körting, Fonseca e Câmara (2013), que sugerem a utilização dos ciclos das culturas anuais em representação polar da ST associada às estações do ano. Logo relacionou-se ao intervalo de seleção das imagens do estudo.

Foram adquiridas 48 imagens de NDVI do sensor MODIS dos satélites Terra e Aqua, com resolução temporal de 16 dias, com diferença de 8 dias entre cada produto, possibilitando uma resolução temporal de 8 dias. As imagens foram selecionadas utilizando as datas de referência das estações do ano. Para padronizar em igual quantidade de imagens por estação

do ano-safra, a seleção de datas foi ajustada para no intervalo entre 14/09/2019 e 29/09/2020, compreendendo o período do ano-safra 2019/2020, contendo 12 imagens por estação do ano. As datas das imagens MODIS contemplando o ano-safra 2019/2020 podem ser visualizadas Tabela 3.

Tabela 2 Datas de referência das estações do ano-safra 2019/2020

Estação	Primavera	Verão	Outono	Inverno
Início	2019-09-23	2019-12-22	2020-03-20	2020-06-20
Término	2019-12-21	2020-03-19	2020-06-19	2020-09-21

Fonte: Adaptado de IAG (2021).

Tabela 3 Datas das imagens MODIS contemplando o ano-safra 2019/2020, associadas às estações do período

Primavera	Verão	Outono	Inverno
2019-09-14	2019-12-19	2020-03-21	2020-06-25
2019-09-22	2019-12-27	2020-03-29	2020-07-03
2019-09-30	2020-01-01	2020-04-06	2020-07-11
2019-10-08	2020-01-09	2020-04-14	2020-07-19
2019-10-16	2020-01-17	2020-04-22	2020-07-27
2019-10-24	2020-01-25	2020-04-30	2020-08-04
2019-11-01	2020-02-02	2020-05-08	2020-08-12
2019-11-09	2020-02-10	2020-05-16	2020-08-28
2019-11-17	2020-02-18	2020-05-24	2020-09-05
2019-11-25	2020-02-26	2020-06-01	2020-09-13
2019-12-03	2020-03-05	2020-06-09	2020-09-21
2019-12-11	2020-03-13	2020-06-17	2020-09-29

A partir do *United States Geological Survey* (USGS) foram obtidas imagens do sensor *Operational Land Imager* (OLI) a bordo do satélite Landsat-8 (imagens OLI), também disponíveis no formato GeoTiff, para auxiliar na inspeção visual da classificação. As imagens OLI possuem 30 metros de resolução espacial para as bandas espectrais (4, 5 e 6) utilizadas, sendo selecionadas com nível de cobertura de nuvens inferior a 20%. As bandas das cenas das imagens OLI foram mescladas aplicando-se técnicas de geração de mosaicos para cobertura total da região de estudo, na composição falsa-cor RGB-564 para facilitar a análise do comportamento espectral das culturas anuais.

4.3 Pré-processamento

Para abstração de dados foi utilizada a biblioteca Python GDAL. Os valores de NDVI das imagens MODIS foram divididos por 10.000 para manter os valores reais, mudando o intervalo dos dados válidos para -0,2 a 1,0. Todas as imagens MODIS foram recortadas utilizando-se a geometria de delimitação da mesorregião do Oeste da Bahia (Figura 6).

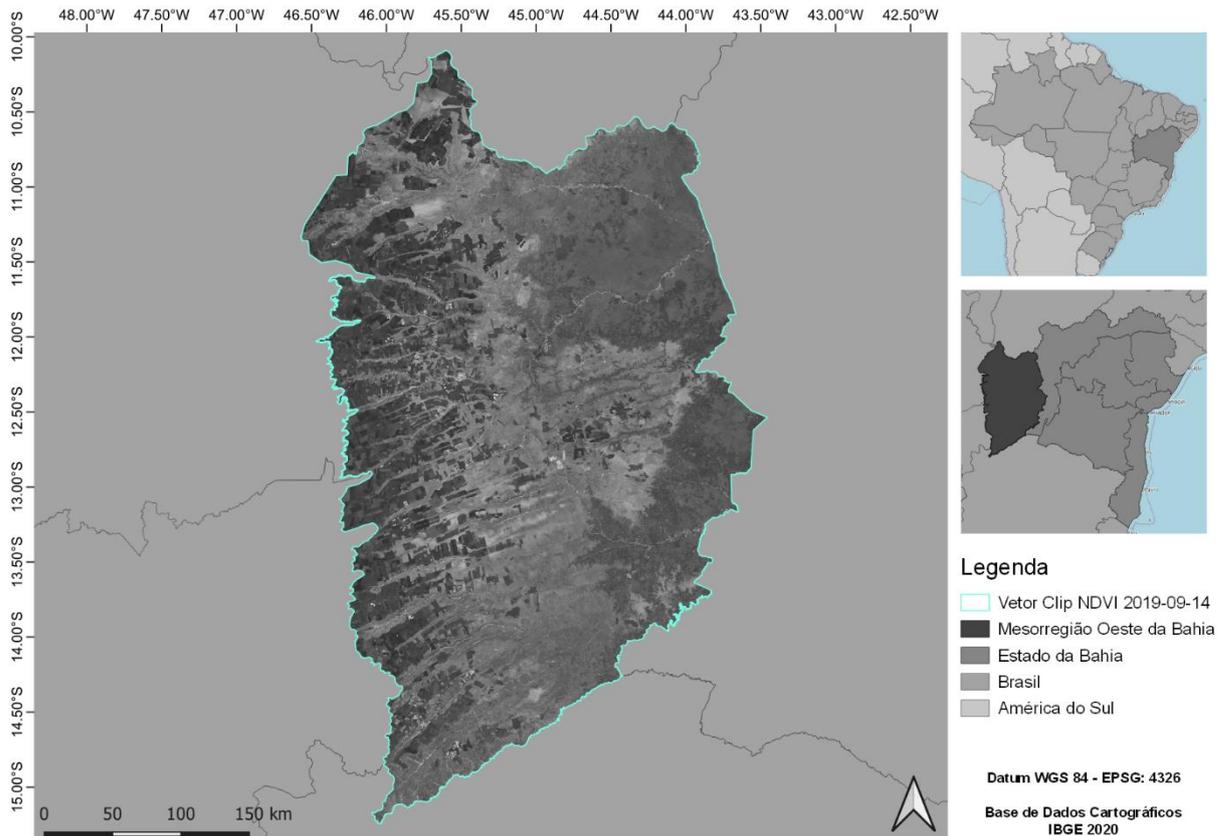


Figura 6 Imagem de NDVI MODIS data 14/09/2019, recortada pela geometria da mesorregião Oeste da Bahia.

Com objetivo de realizar análises temporais de cada *pixel* das imagens, foi construído um empilhamento das imagens em cubo temporal. Foram utilizadas 48 imagens para construção do cubo temporal de imagens com satélite de NDVI, utilizando as datas como parâmetro de ordenação das imagens no cubo (Figura 7). Cada *pixel* do empilhamento temporal apresenta uma série temporal de índice de vegetação com 48 pontos (imagens) de NDVI, ou seja, um *pixel* ao longo do tempo observado.

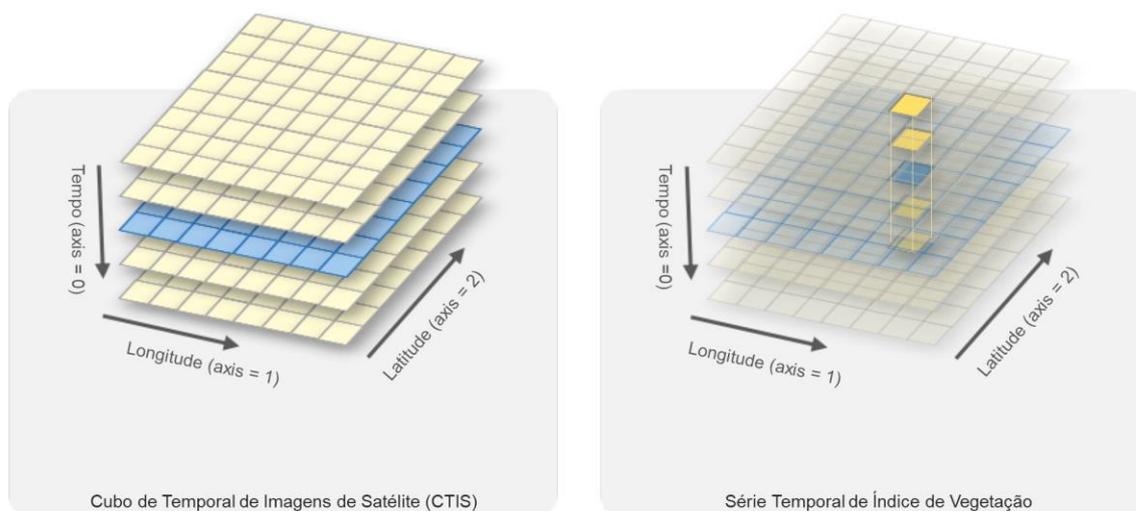


Figura 7 Representação do cubo temporal de imagens de satélite e da série temporal de índice de vegetação (STIV).

Fonte: Adaptado de Simões et al. (2021).

4.4 Transformação

O método de suavização Savitzky-Golay (SG) (SAVITZKY; GOLAY, 1964) foi utilizado para transformação da STIV. O algoritmo SG foi aplicado no eixo do tempo da STIV com suporte da biblioteca Scipy, os parâmetros aplicados no algoritmo foram *window_length* = 5 e *polyorder* = 3, resultando em uma STIV suavizada STIV-SG.

Para reduzir a complexidade do modelo de classificação na etapa de mineração de dados, realizou-se a extração e análise dos componentes da STIV de cada *pixel*. A simplificação da STIV-SG foi realizada extraíndo-se os componentes da série T, S e R, considerando o período de análise de tendência a cada 4 datas (*period* = 4, argumento do modelo), utilizando a biblioteca *Python Statsmodel*.

Foram ajustadas estruturas no código-fonte da biblioteca STMetrics para permitir maior integração de dados com outras bibliotecas. Com intuito de potencializar o desempenho da etapa de mineração de dados em STIV, realizou-se a extração de 15 métricas básicas das STIV-SG e STIV-Trend com auxílio do pacote STMetrics. Essas métricas básicas provêm de abordagens estatísticas descritivas comuns.

Com técnicas de engenharia de recursos e suporte do pacote STMetrics foram calculadas 10 métricas polares a partir da projeção circular (forma polar) das STIV-SG e STIV-Trend. A forma polar da série temporal é a transformação dos valores da STIV para ângulos no intervalo $[0, 2\pi]$ aplicado sobre o plano cartesiano, formando quatro quadrantes.

4.5 Mineração de dados

4.5.1 Agrupamento de classes amostrais

Realizou-se a extração e estudo descritivo de STIV em cada classe de UCT selecionada no Dataset LEM+ para verificar a variabilidade entre instâncias. Para análise da dissimilaridade entre as STIV foi utilizado o algoritmo DTW disponível na biblioteca Python *DTAIDistance* (MEERT et al., 2022). Analisou-se a sincronização temporal dinâmica entre séries temporais da classe de culturas anuais, visualizadas na Figura 8.

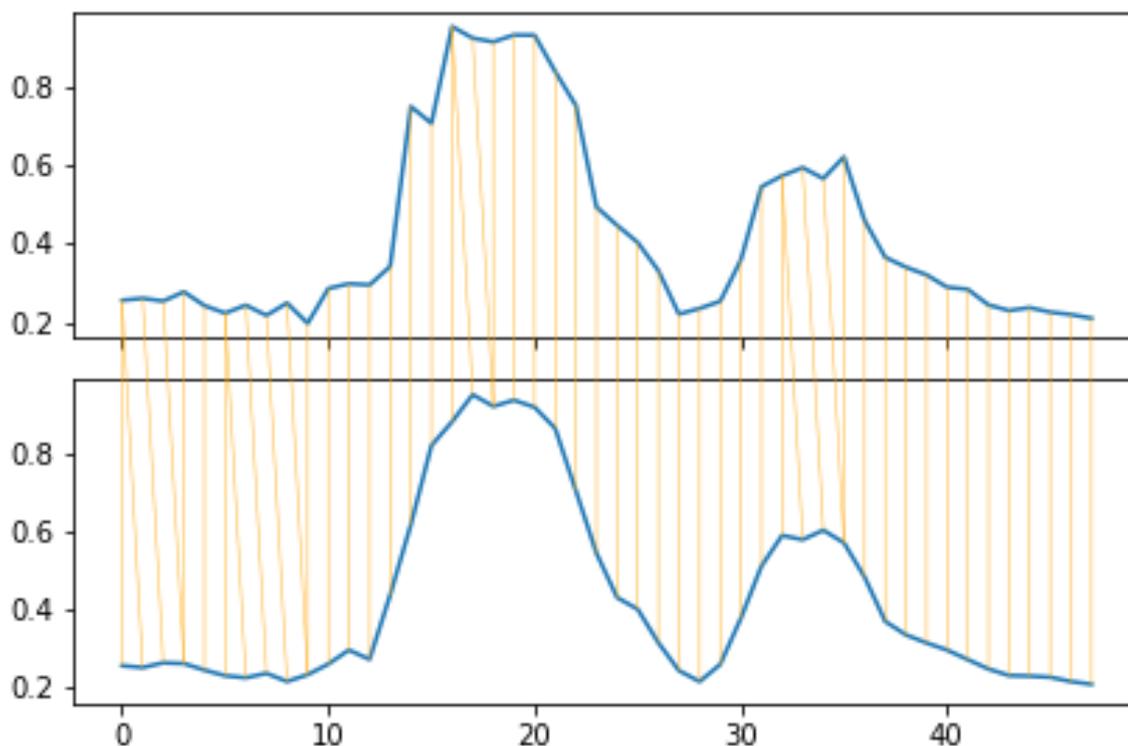


Figura 8 Sincronização temporal dinâmica (DTW) entre 2 séries temporais da classe culturas anuais.

Fonte: Adaptado de Meert et al. (2022).

Na Figura 9 estão dispostos os gráficos tipo boxplot para os valores de índice de vegetação e de dissimilaridade das classes, para as STIV brutas (*Raw*), pré-processadas com SG e STIV transformadas com TD. As 14 classes únicas de UCT foram organizadas e reclassificadas em 2 grupos, classes para amostragem e treinamento do modelo, culturas anuais: algodão, feijão, milho, soja, sorgo e outros: área de conversão com vegetação nativa

- cerrado, braquiária, café, eucalipto, feno, painço, pastagem, sem cultivo. O agrupamento de classes foi realizado com objetivo de aumentar o espaço de amostragem com culturas anuais, mas que contemplassem a nova categorização.

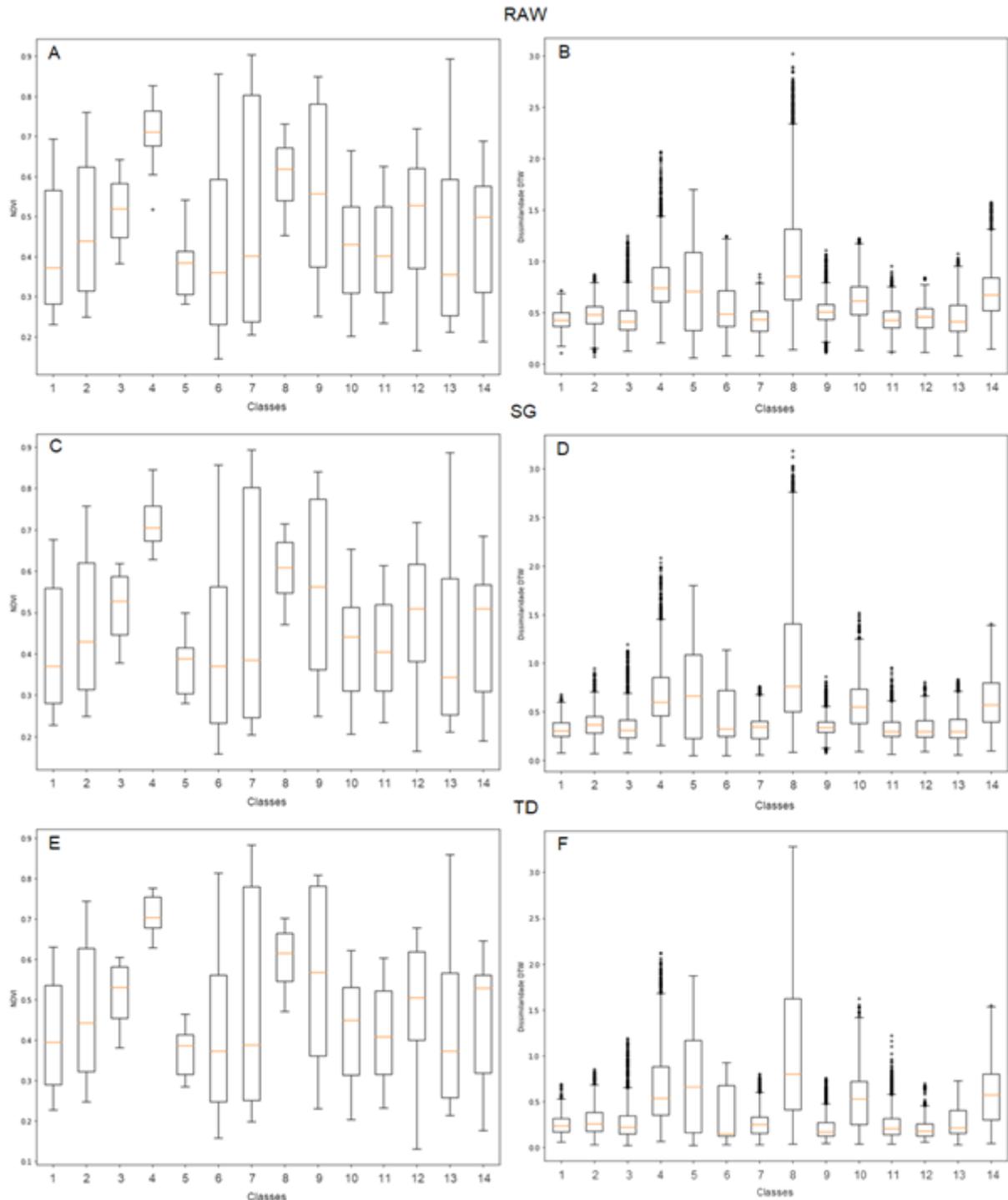


Figura 9 Boxplot da distribuição dos valores de NDVI (A, C e E); boxplot da distribuição da dissimilaridade das classes amostrais Dataset LEM+ (B, D e F), para o mês de fevereiro de 2020.

Nota: 1 – Feijão, 2 – Braquiária, 3 – Cerrado, 4 – Café, 5 – Área de conversão, 6 – Milho, 7 – Algodão, 8 – Eucalipto, 9 – Feno, 10 – Painço. 11 – Pastagem, 12 – Sorgo, 13 – Soja, 14 – Sem cultivo.

4.5.2 Combinação de atributos

Foi desenvolvido um fluxo de processos automatizados (*pipeline*) para MD, com a tarefa de classificação supervisionada em STIV e seus atributos derivados (básicos e polares). O *pipeline* foi composto por 7 operações unitárias para MD: amostragem estratificada, importância e seleção de atributos, treinamento do classificador, pré-avaliação, otimização do classificador, seleção do modelo otimizado e produção do modelo otimizado.

Os tratamentos de séries temporais STIV-SG e STIV-Trend (TD) foram organizados em 10 combinações de atributos (Tabela 4). As combinações foram utilizadas em testes de mineração de dados, buscando melhores performance na tarefa de MD. O processo de MD foi realizado e analisado em cada uma das combinações de atributos, buscando compreender a variação de performance da classificação com base nas métricas de avaliação dos resultados.

Tabela 4 Combinações de atributos utilizados na *pipeline* de mineração de dados

Combinação	Tratamento	Atributos	Quantidade de atributos
I	SG	TS	48
II		TS BS	63
III		TS BS PL	73
IV		TS PL	58
V		BS PL	25
VI	TD	TS	48
VII		TS BS	63
VIII		TS BS PL	73
IX		TS PL	58
X		BS PL	25

Nota: SG: STIV-SG; TD: STIV-Trend; TS: Série temporal; BS: Atributos extraídos; PL: Atributos calculados na forma polar da STIV.

4.5.3 Amostragem, importância e seleção de atributos

Para redução de possíveis ruídos amostrais ou amostragem de instâncias sequenciais, *pixels* provenientes do mesmo campo ou geometria amostrada no *Dataset* LEM+, foi aplicado o utilitário *train_test_split* do módulo de seleção de modelos da biblioteca Scikit-Learn, com o parâmetro *stratify* direcionado para os rótulos das classes. Durante o processo, foram selecionadas aleatoriamente 750 instâncias por classe de amostragem (culturas anuais e outros). Com o utilitário, as instâncias foram subdivididas em 2 subconjuntos. Para o subconjunto de treinamento do classificador foram direcionadas 2/3 das instâncias em cada

classe, enquanto outros 1/3 das instâncias foram reservadas para um subconjunto de teste da MD após aplicação do modelo.

O algoritmo RF com parâmetros padrões (*Default*) foi aplicado para auxiliar no ranqueamento da importância relativa de cada atributo durante o processo de treinamento na tarefa de classificação. Utilizou-se o algoritmo Boruta para detectar e selecionar atributos com maior relevância para a classificação.

A métrica de suporte do Boruta foi utilizada para criar a listagem de atributos recomendados para utilização no processo de MD, assim como a lista de atributos para exclusão em cada combinação de classificação. Posteriormente realizou-se a comparação com o *ranking* de importância relativa dos atributos. Após a avaliação da importância relativa dos atributos e da utilização do Boruta para seleção de atributos, foram removidas as variáveis com menor contribuição na performance da tarefa de classificação.

4.5.4 Treinamento, otimização e seleção do modelo otimizado

Após a avaliação da importância e seleção de atributos foi implementado o algoritmo RF com a tarefa de classificação de dados, o algoritmo está disponível na biblioteca *Scikit-Learn* e foi utilizado com parâmetros padrões. Nessa etapa o classificador foi treinado com um utilitário de validação cruzada (*Cross validation – CV*) para pré-avaliação da acurácia da modelagem, realizando o processo para cada uma das 10 combinações de atributos.

O utilitário de pesquisa de hiperparâmetros em grade com validação cruzada *GridSearchCV* foi adicionado ao *pipeline* com o classificador RF para encontrar o melhor modelo com base na acurácia de cada combinação, possibilitando testar 40 combinações de hiperparâmetros (Tabela 5).

Tabela 5 Parâmetros e valores utilizados na pesquisa de hiperparâmetros em grade

Parâmetro	Valores testados
bootstrap	True, False
criterion	gini, entropy
n_estimators	500, 1000, 1500, 2000, 2500
max_features	sqrt, log2

Nota: *Bootstrap*: uso de amostras para construção de cada árvores, com ou sem reamostragem (true/false); *criterion*: critério de separação de amostra nas árvores, *entropy* – separa pelo ganho de informação e *gini* – impureza da amostra; *n_estimators*: quantidade de árvores aleatórias utilizadas; *max_features*: número de recursos a serem utilizado em uma divisão, sqrt – raiz quadrada, log2 – logaritmo base 2.

Fonte: Adaptado de Pedregosa et al. (2011).

O *GridSearchCV* utiliza como base um algoritmo simplificado, com parâmetros padrão, de MD e uma sequência de parâmetros nomeados com objetivo de otimizar, realizando exaustivamente o treino e avaliação do algoritmo de MD (BUITINCK et al., 2013).

A validação cruzada embutida no utilitário permitiu subdividir o conjunto de treinamento em subconjunto de treino e subconjunto de teste do classificador (*k-Fold*). Foram realizadas 10 repetições automatizadas de treino e teste do classificador variando o particionamento das instâncias utilizadas no processo (*k-split*) (Figura 10). Ao todo, em cada combinação foram testados 400 classificadores, em seguida avaliados e obtendo-se as estatísticas. Com as estatísticas de cada teste foram detectados os melhores hiperparâmetros e construído o melhor modelo de MD para classificação de culturas anuais em cada combinação de atributos foi extraído.

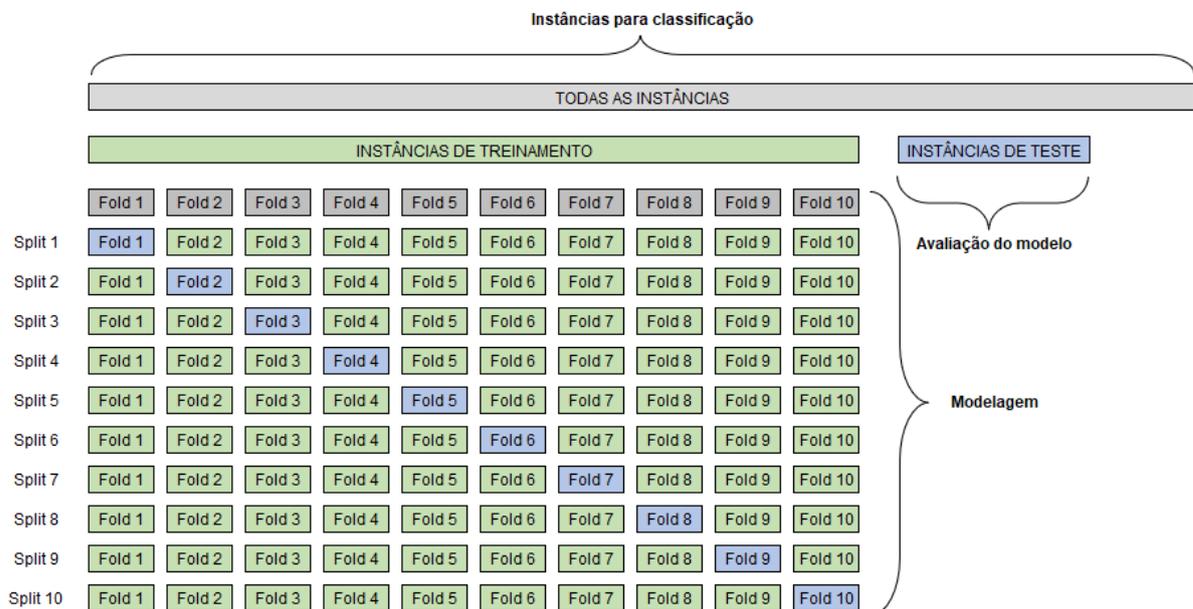


Figura 10 Subconjuntos de instâncias derivados durante o processo de validação cruzada.

Fonte: Adaptado de Pedregosa et al. (2011).

4.6 Interpretação e avaliação

A avaliação do desempenho de cada classificação foi realizada com auxílio da matriz de confusão (MC) (Figura 11). Em geral as métricas de avaliação calculadas estão no intervalo de 0 a 1, em que valores mais próximos de 1 indicam melhores resultados (BEM et al., 2021), exceto os erros de omissão e comissão. As métricas de avaliação da MC são geradas a partir de cálculos dos possíveis resultados de verdadeiro-positivo (TP), verdadeiro-negativo (TN), falso-positivo (FP) e falso-negativo (FN), comparando as instâncias reservadas para o

subconjunto de teste da MD, após aplicação do modelo com os rótulos de classes gerados após o processo de classificação com RF.

		AMOSTRA	
		+	-
PREDIÇÃO	+	TP	FP
	-	FN	TN

Figura 11 Matriz de confusão.

O coeficiente de concordância Kappa (Kappa) (Equação 4) indica a força da concordância entre as amostras reais e valores preditores na classificação (PEDREGOSA et al., 2011; SINGH; TYAGI, 2021). Classificações com valores de Kappa próximos a 1 apresentam concordância ideal entre amostras reais e preditas, enquanto valores próximos a 0 (zero) indicam pouca concordância na classificação (SINGH; TYAGI, 2021).

$$\text{Kappa} = \frac{(P_o - P_e)}{(1 - P_e)} \quad (4)$$

em que: P_o é a probabilidade empírica de concordância no rótulo atribuído a qualquer amostra (a razão de concordância observada) e P_e é a concordância esperada quando em ambas as observações são atribuídos rótulos aleatoriamente. P_e é estimado usando uma priorização empírica por observação sobre os rótulos de classe.

A precisão do produtor é a probabilidade de que a classe dada seja classificada corretamente, sendo calculado pelo índice de recuperação (*Recall*) em termos de TP e FN (Equação 5), medindo a sensibilidade da classificação.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

A precisão do usuário (Precisão) escrita em termos de TP e FP (Equação 6), apresenta a probabilidade de que a classe foi prevista corretamente ou não.

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (6)$$

A previsão de um valor correto para outra classe sendo medida como um falso-negativo é um erro de omissão, indicando a acurácia da classificação, podendo ser observado na primeira coluna da CM e escrito em valor de FN e TP (Equação 7).

$$\text{Erro de Omissão} = \frac{FN}{FN+TP} \quad (7)$$

A previsão de um valor errado na classe correta é um falso-positivo, escrito em termos de FP e TP (Equação 8), sendo reconhecido como erro de comissão, apresentando a confiabilidade da classe predita estar correta.

$$\text{Erro de Comissão} = \frac{FP}{FP+TP} \quad (8)$$

A média harmônica entre a precisão e a recuperação (F1-Score), informando o nível de precisão do modelo de classificação pela taxa de acertos (Equação 9), assim como sua robustez na identificação da classe em instâncias mais complexas de distinguir. F1-Score busca um equilíbrio estatístico entre a precisão e o *Recall* de um modelo classificador.

$$\text{F1 - Score} = \frac{2*TP}{2*TP+FP+FN} \quad (9)$$

em que:

- TP é o número de observações verdadeiras positivas;
- TN é o número de observações verdadeiras negativas;
- FP é o número de observações falsas positivas;
- FN é o número de observações falsas negativas.

A acurácia calcula a precisão do subconjunto, em que um vetor conhecido os rótulos de uma classe devem coincidir com a rotulação prevista, calculada a partir da Equação 10.

$$\text{Acurácia} = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i) \quad (10)$$

em que: Se \hat{y}_i é o valor predito da i-ésima amostra e y_i é o valor verdadeiro correspondente, então a fração das predições corretas n_{samples} é definida como onde está a função do indicador.

A avaliação do mapeamento de culturas anuais foi executada ponto a ponto por meio de inspeção visual, sobrepondo o mapa de classificação sobre a composição de imagens OLI falsa-cor RGB-564 (Figura 12). Foram sorteados aleatoriamente 100 pontos sobre o alvo culturas anuais e outros 100 pontos sobre os demais alvos fora da classe alvo. A verificação ocorreu analisando se o ponto inspecionado pertencia (ponto amarelo) ou não (ponto rosa) à classe objetivo. Posteriormente os valores foram armazenados em variáveis e com auxílio da linguagem Python calcularam-se as métricas de avaliação Kappa, precisão e acurácia para o mapeamento.



Figura 12 Avaliação da máscara de culturas anuais por inspeção visual ponto a ponto.

4.7 Produção do modelo otimizado

Após os processos de otimização, seleção e avaliação do modelo otimizado de classificação, com objetivo de reutilizar o modelo fez-se necessária a realização do processo de perpetuidade do classificador, construindo um arquivo que armazenasse todas as informações e características de construção do modelo gerado ou simplesmente o processo de produção do modelo. A biblioteca Python Joblib possibilita a produção de modelos com eficiência e transparência devido ao processo de persistência em disco para objetos com grande volume de dados utilizando um mecanismo computacional de cache (VAROQUAUX, 2021), processo crucial para que aplicações de MD possam operar de forma assíncrona, permitindo repetição e escalabilidade.

4.8 Geração do mapeamento com culturas anuais e extração de área mapeada

Foi gerado o mapeamento com a classe de UCT culturas anuais para toda a região de estudo, permitindo a construção de mapas de classificação. Para esse processo foi importado o modelo de classificação RF e aplicado sobre a combinação de atributos selecionados gerando arquivo *raster* (matricial) no formato TIFF.

Para reduzir efeitos de *pixels* isolados nos limites das classes, aplicou-se um filtro majoritário na máscara *raster*. O filtro permite a substituição do valor de um *pixel* com base no valor da maioria (moda) dos *pixels* de sua vizinhança espacial (ao seu redor). O filtro majoritário foi construído com auxílio das bibliotecas Python GDAL, Numpy e Scipy. A filtragem considerou a frequência de valores ao redor do *pixel* analisado uma janela móvel de 5 *pixels*.

O mapeamento da classificação no formato *raster* foi poligonizada. Esse processo, em geral, é custoso computacionalmente, tendo em vista que os *pixels* de mesma classe são agrupados e convertidos em uma forma geométrica vetorial. Após a poligonização do mapeamento, as geometrias foram agrupadas e exportadas no formato geoJSON, contendo apenas a classificação de culturas anuais.

Para aumentar a qualidade de contornos das geometrias do mapeamento e reduzir o efeito “*pixelado*” foi aplicado um método de simplificação geométrica. O processo consiste em ajustar a quantidade de pontos (latitude e longitude) que compõe a forma baseando-se em uma distância de tolerância, mantendo as características topológicas da geometria.

Na sequência foi realizada a extração das geometrias mapeadas por município da região Oeste da Bahia utilizando a máscara simplificada com culturas anuais e o arquivo contendo com os limites dos municípios. Posteriormente, foram realizados cálculos das áreas mapeadas em cada município, em hectares (ha), gerando uma planilha eletrônica com as referidas informações para análises comparativas.

4.9 Dados oficiais para fins comparativos e estudo de caso

Dados de área com culturas anuais do LSPA foram utilizados como referência comparativa do mapeamento realizado. A área com culturas anuais do LSPA foi obtida a partir do somatório da mediana dos últimos 3 anos de dados disponíveis na pesquisa PAM do IBGE

Da Tabela 1612 as quantidades de áreas com as culturas de algodão, feijão, soja, sorgo e trigo; e os dados da cultura de milho 1º safra (de verão) adquirido da Tabela 839, anos-safra 2019 e 2020, 2021 (IBGE, 2022).

Utilizando a plataforma do SiCAR foi escolhido aleatoriamente um imóvel sob a região mapeada para realização de estudo de caso. O imóvel rural escolhido com cadastro ativo e número de registro BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A, com centroide em latitude 12°28'21,62"S e longitude 45°35'27,13"W, é localizado no município de São Desidério – BA, visualizado na Figura 13. Realizou-se a coleta de informações vetoriais (*Shapefile*) e de informações documentais constantes no demonstrativo do CAR, verificando a condição do cadastro que estava em aguardo de análise, na data consultada. As datas de registro e de última retificação do imóvel no SiCAR são 27/10/2014 e 05/12/2014, respectivamente.

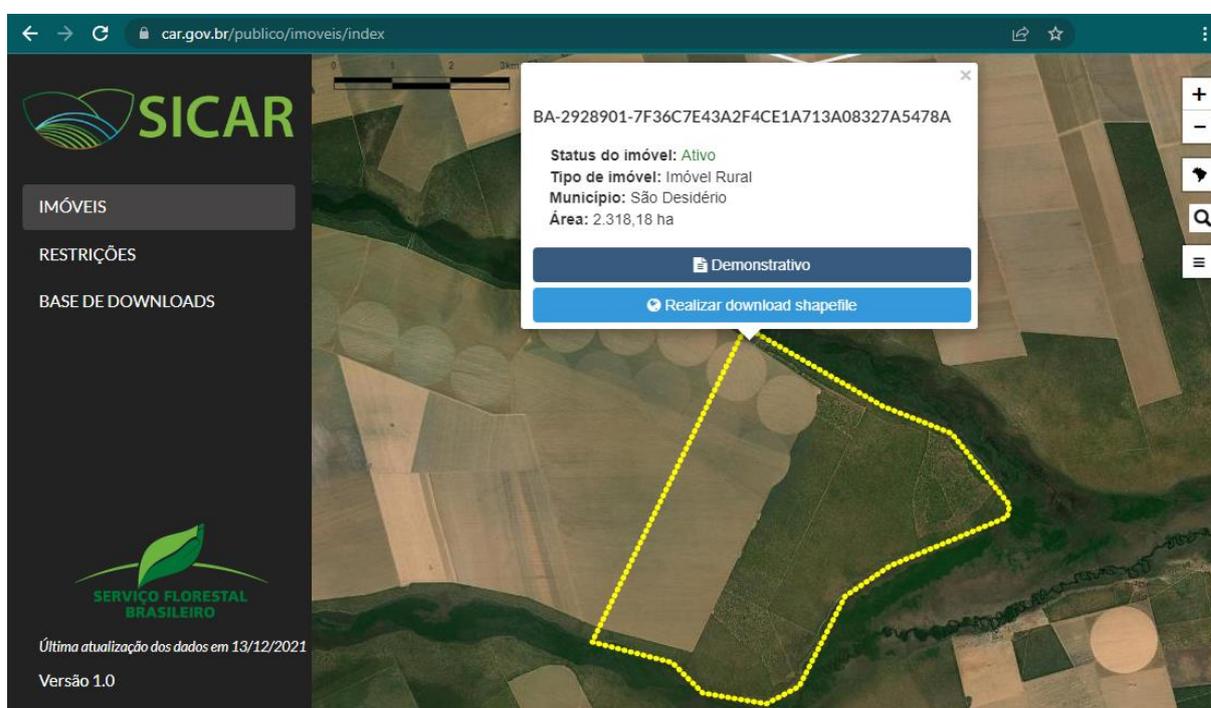


Figura 13 Tela de seleção do imóvel no SiCAR em São Desidério – BA.

Fonte: Adaptado de SiCAR (2022).

As informações vetoriais são disponibilizadas em arquivo ZIP contendo 4 arquivos *shapefile* referentes às classes de uso e cobertura do solo (UCS) do imóvel, vetores de reserva legal (RL), cobertura do solo, área do imóvel e área de preservação permanente (APP). Do arquivo vetorial de cobertura do solo foram extraídos os vetores de área não classificada pelo SiCAR e de área total de remanescentes de vegetação nativa (RVN) (Figura 14). Subtraindo da área total do imóvel o somatório das informações das áreas de RVN, RL e APP constantes no demonstrativo do imóvel obtém-se a área não classificada e comparando

com as informações vetoriais, é possível identificar a área de uso consolidado do imóvel (Tabela 6).



Figura 14 Classes de cobertura do solo SiCAR do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.

Tabela 6 Informações documentais constantes no demonstrativo do CAR do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.

Cobertura do Solo	Área (ha)	% Total
Área total de remanescentes de vegetação nativa	888,67	38,3
Área de reserva legal (averbada vetorizada)	463,57	20,0
Áreas de preservação permanente	113,12	4,9
Área não classificada (uso consolidado)	852,82	36,8
Área total do imóvel	2.318,18	100

Com o vetor do imóvel foram realizados 2 processos para testar o *pipeline* de MD para classificação de áreas com culturas anuais. Um dos processos consistiu em realizar um geoprocesso de recorte dos limites do imóvel sobre o mapeamento de classificação com culturas anuais da mesorregião Oeste da Bahia, calculando a área mapeada dentro do imóvel

(Caso M1). Enquanto, no outro processo obteve-se o recorte dos limites do imóvel em todas as imagens de STIV e seus atributos da combinação, com resultados mais satisfatórios na tarefa de MD, desse derivaram-se os casos M2 e M3 (Figura 15).

A biblioteca Python Joblib foi utilizada para importação e aplicação do modelo de classificação RF já treinado, com intuito de realizar o processo de MD sobre a região do imóvel e identificar as áreas com culturas anuais dentro do imóvel (Caso M2). O processo de simplificação geométrica foi aplicado sobre a região classificada com culturas anuais dentro do imóvel (Caso M3), em seguida foram realizados cálculos das áreas mapeadas no imóvel antes e após o processo de simplificação, realizando uma comparação das informações de área do demonstrativo, das geometrias e dos processos de classificação.



Figura 15 Recorte de imagem de atributo derivado de STIV-Trend com base no limite do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A.

Para a etapa de avaliação dos mapeamentos com culturas anuais no imóvel, foram gerados pontos no centroide de cada *pixel* de uma imagem MODIS de NDVI, dentro dos limites do imóvel, resultando 385 pontos. A geometria de uso consolidado do imóvel, onde é

possível inspecionar a área de agricultável do imóvel via imagens OLI em composição falsa-cor 564 podem ser visualizados na Figura 16. Com o processo de seleção espacial, com suporte do *software* QGIS, os pontos que intersectavam a geometria receberam a classe culturas anuais (143 pontos) e os demais pontos rotulados como outras classes. Esses dois conjuntos de pontos foram utilizados como referência de verdadeiro (Figura 16).

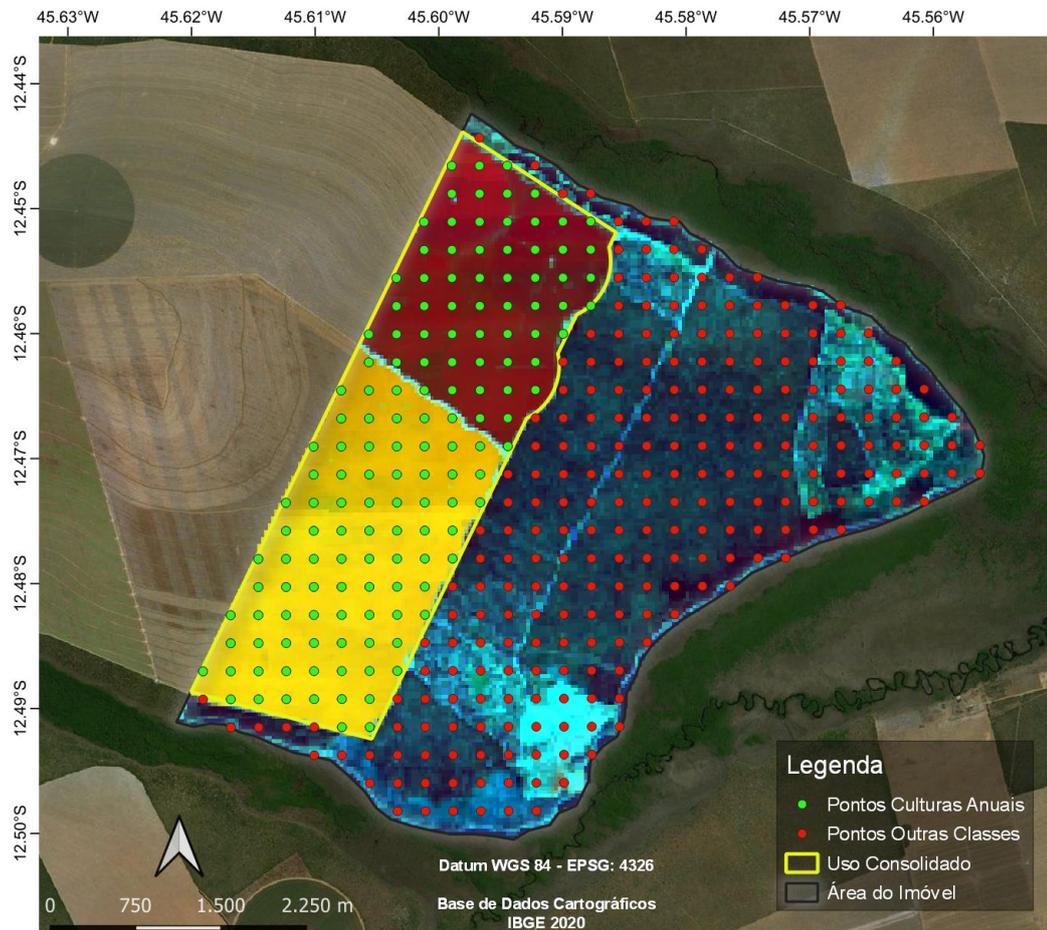


Figura 16 Pontos referência de verdadeiro para as classes de culturas anuais (em verde) e outras classes (em vermelho) no imóvel BA-2928901-7F36C7E43A2F4CE1A713A08 327A5478A, sobre imagens OLI falsa-cor 564.

O processo de validação e cálculo de métricas foi realizado para os casos de mapeamento: Caso M1 – recorte do imóvel no mapeamento classificado; Caso M2 – processo de MD com o modelo aplicado sobre as imagens de STIV-Trend, antes da simplificação geométrica; Caso M3 – processo de MD com o modelo aplicado sobre as imagens de STIV-Trend, após simplificação geométrica.

5 RESULTADOS E DISCUSSÃO

Na Figura 17, apresenta-se uma STIV de um *pixel* selecionado com culturas anuais, antes e após a suavização aplicando o filtro SG, como parte da etapa de pré-processamento. Os dados brutos (STIV-Raw, em vermelho) possuem maior quantidade de picos em relação aos dados reconstruídos STIV suavizada (STIV-SG, em azul), observando-se uma janela de três datas.

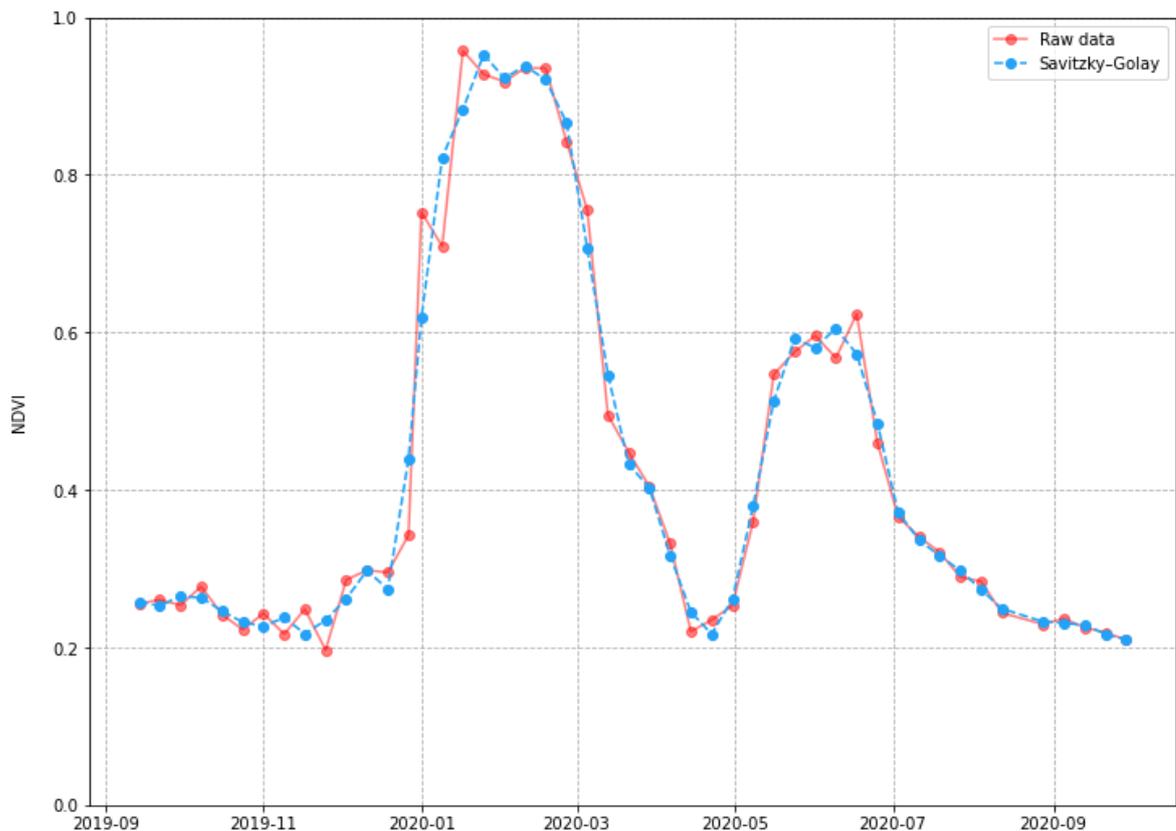


Figura 17 STIV-Raw de culturas anuais (vermelho) e após suavização com filtro SG (azul) STIV-SG.

A simplificação da STIV-SG foi realizada extraindo-se os componentes da série T, S e R, utilizando-se o modelo STL (Figura 18). O período de análise de tendência aplicado no modelo STL foi de quatro datas ($period = 4$, argumento do modelo). As ocorrências sazonais, variações do nível de base (médio) e o comportamento de longo prazo das séries temporais podem ser analisados pela componente de tendência.

As componentes de tendência de cada STIV-SG foram armazenadas em STIV-Trend após a extração, análise e seleção com o modelo STL. A componente de tendência da série não apresenta fatores de sazonalidade dentro do ano-safra e nem interferências aleatórias, devido à robustez do modelo STL.

A decomposição da STIV-SG utilizando STL evitou as ocorrências de sazonalidade intrínsecos do método estatístico de suavização SG, assim como observado por Ouyang, Ravier e Jabloun (2021), que utilizaram o modelo STL como uma etapa de pré-processamento, em séries temporais, melhorando os resultados das previsões realizadas.

Na Figura 19 é possível verificar os componentes T, S e R da STIV para outras classes, nesse caso o eucalipto. Constatou-se menor variação na série temporal de tendência, o que pode auxiliar na diferenciação entre STIV com culturas anuais, devido ao ciclo vegetativo da agricultura em relação ao ciclo das culturas perenes e alvos permanentes, como a vegetação nativa do cerrado, característica na região de estudo.

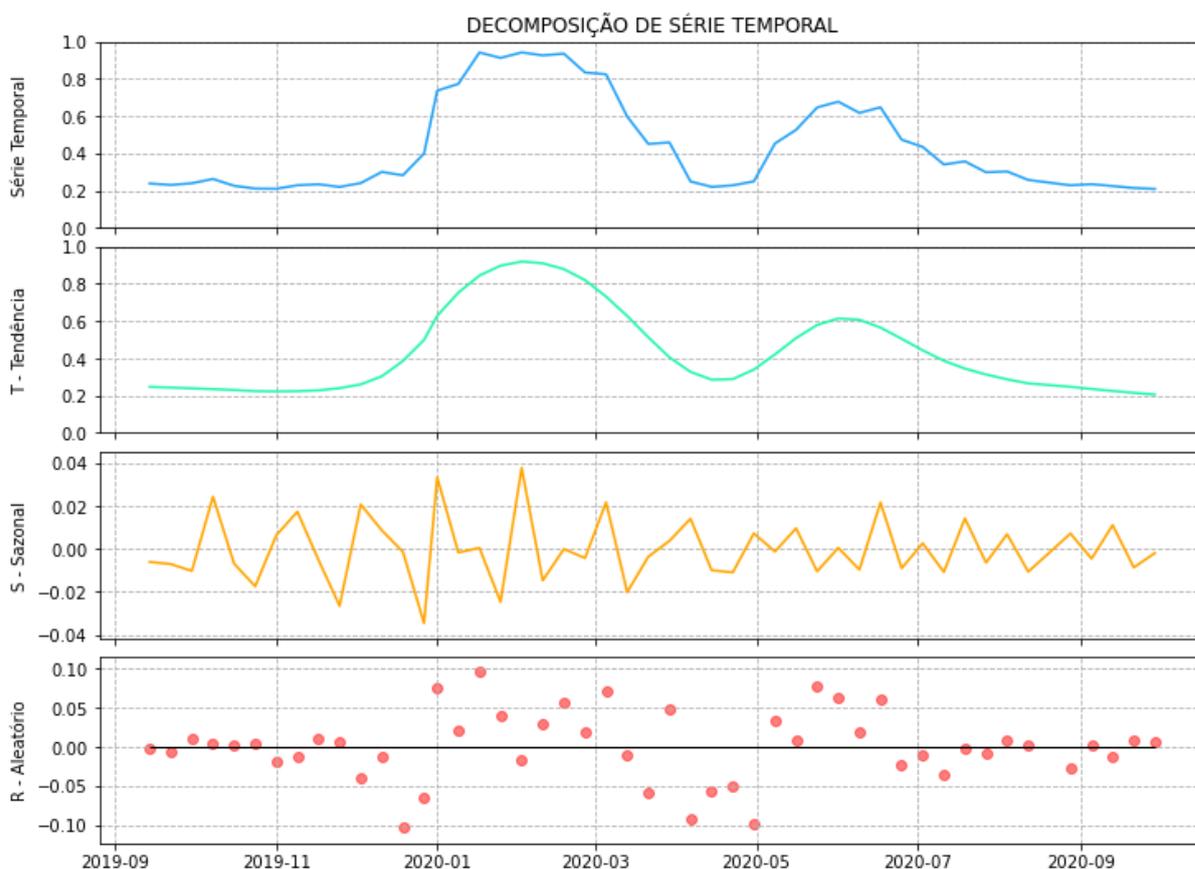


Figura 18 Decomposição da STIV-SG de um *pixel* contendo culturas de ciclo anual no ano-safra 2019/2020, em componente de tendência (T), sazonal (S) e aleatório (R), utilizando o modelo STL.

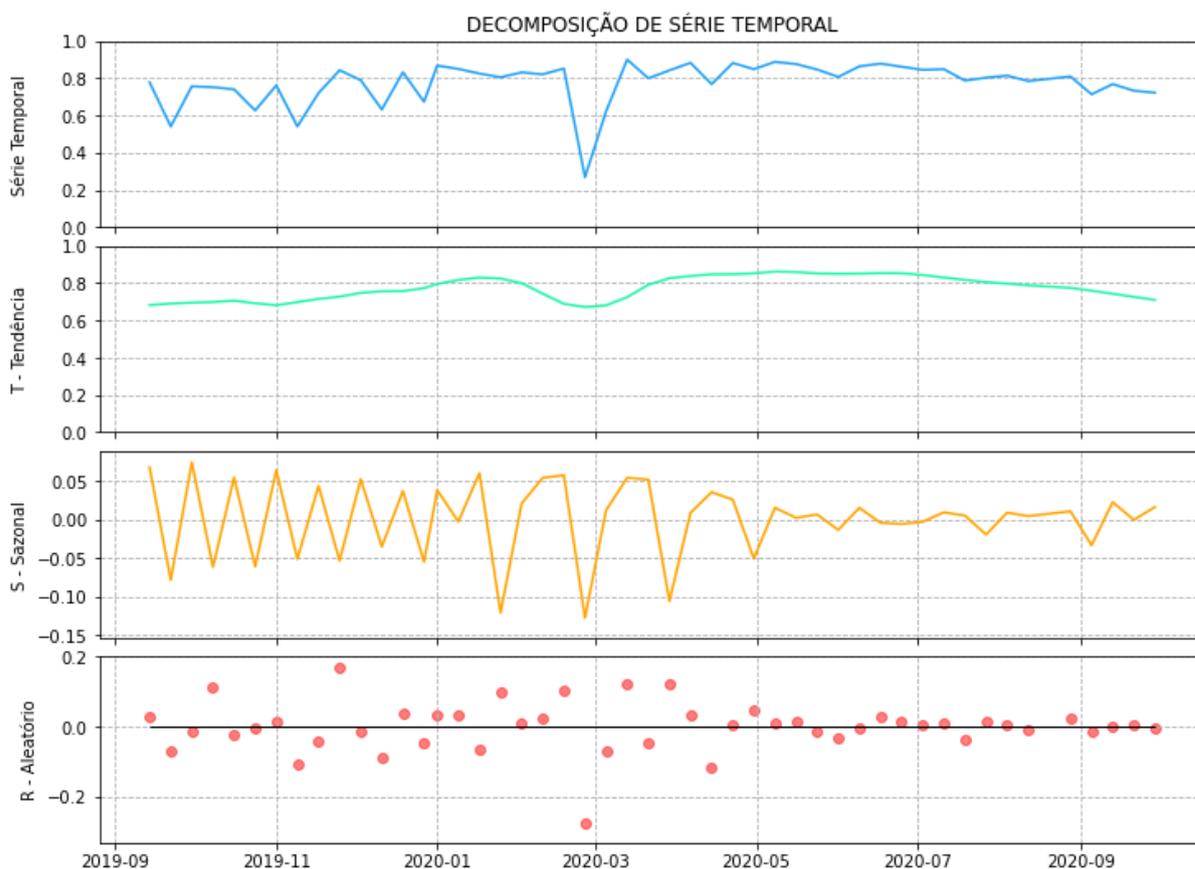


Figura 19 Decomposição da STIV-SG de um *pixel* contendo outras classes (eucalipto) no ano-safra 2019/2020, em componente de tendência (T), sazonal (S) e aleatório (R), utilizando o modelo STL.

Após a projeção da STIV-SG e STIV-Trend na forma polar, em que os valores de índice de vegetação foram transformados para ângulos no intervalo $[0, 2\pi]$ (KÖRTING; FONSECA; CÂMARA, 2013; SOARES et al., 2020), foram extraídas métricas básicas e polares das STIV. A representação da forma polar de quatro *pixels* de STIV-Trend com as classes de culturas anuais (soja e algodão) e outras classes (eucalipto e pastagem) pode ser verificada na Figura 20.

A representação na forma polar (Figura 20) permite a visualização do padrão do ciclo da vegetação no ano-safra 2019/2020. Para extrair métricas fenológicas de séries temporais, são necessários limites temporais adequados para definir o início e o final da temporada (OUYANG; RAVIER; JABLOUN, 2021). A abordagem baseada nas datas de referência das estações do ano-safra 2019/2020, padronizou as transformações das STIV para a forma polar, como sugerido por Körting, Fonseca e Câmara (2013).

O relacionamento de cada quadrante com as quatro estações do ano, em que o quadrante (Figura 20-A, *summer*) com a maior área hachurada é associada ao período de desenvolvimento das culturas anuais de verão, com início do ciclo na primavera (*spring*). STIV com culturas anuais apresenta forma polar mais irregular. Enquanto, *pixels* com STIV de outras classes, que em geral é constituída por alvos permanentes ou culturas perenes (Figura

20-B e Figura 20-D), apresenta forma polar com menor desvio padrão de um objeto em relação a um círculo (CSI), circunferência e maior área hachurada.

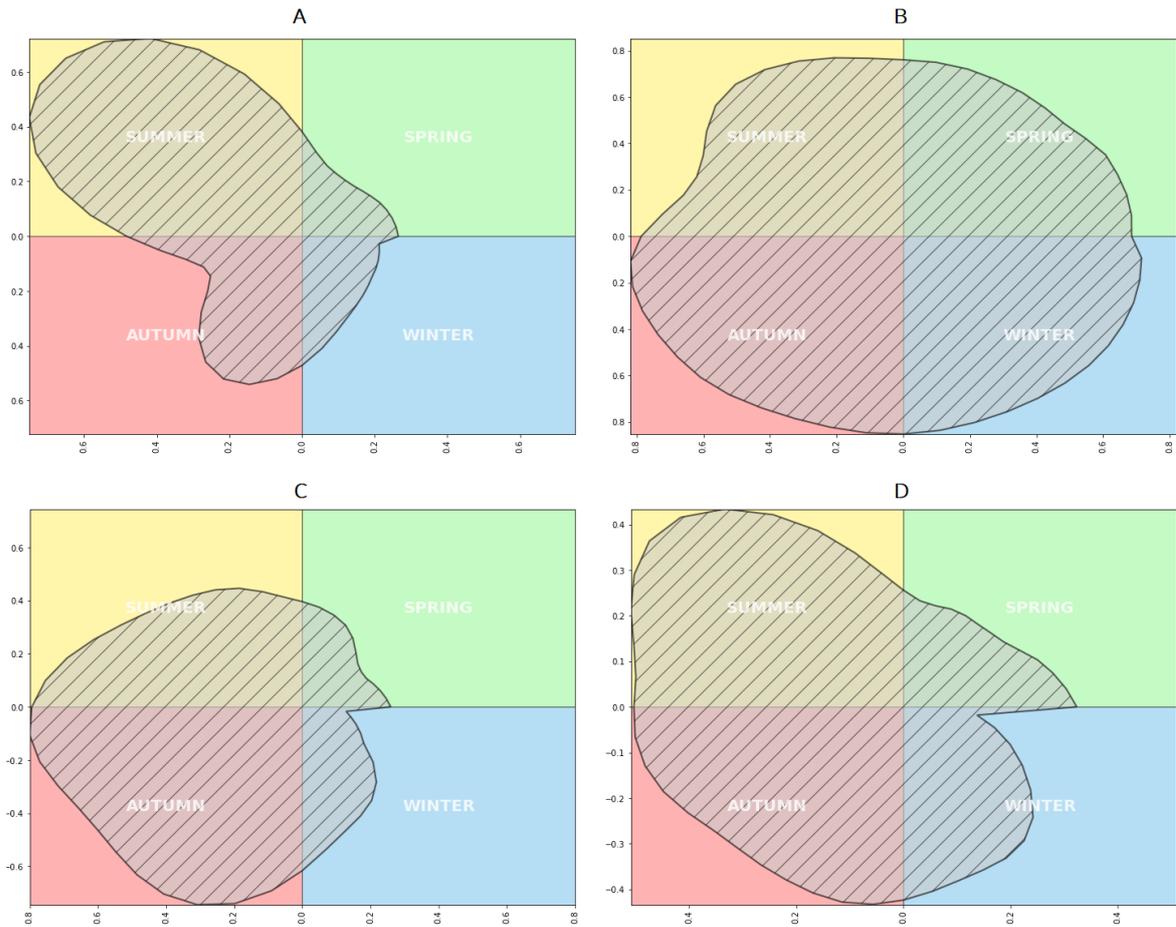


Figura 20 Representação da STIV-Trend na forma polar de *pixels* com as classes de culturas anuais (A – soja e C – algodão) e outras classes (B – eucalipto e D – pastagem).

As imagens de atributos derivados (AD) de STIV foram construídas com as métricas básicas e polares para cada conjunto das STIV-SG e STIV-Trend. As métricas básicas possibilitaram a construção de 15 imagens de AD, visualizadas na Figura 21. Enquanto, com as métricas polares foram construídas 10 imagens de AD utilizando a partir das STIV (Figura 22).

A inspeção visual das Figura 21 e Figura 22 permite identificar diferenças no padrão de resposta (valores) de cada *pixel* nos diferentes atributos gerados. Os mapas construídos com as métricas foram padronizados para que apresentação de valores altos fosse realizada em tons de cor mais escura, enquanto valores menores em tons de cores mais claras nas imagens.

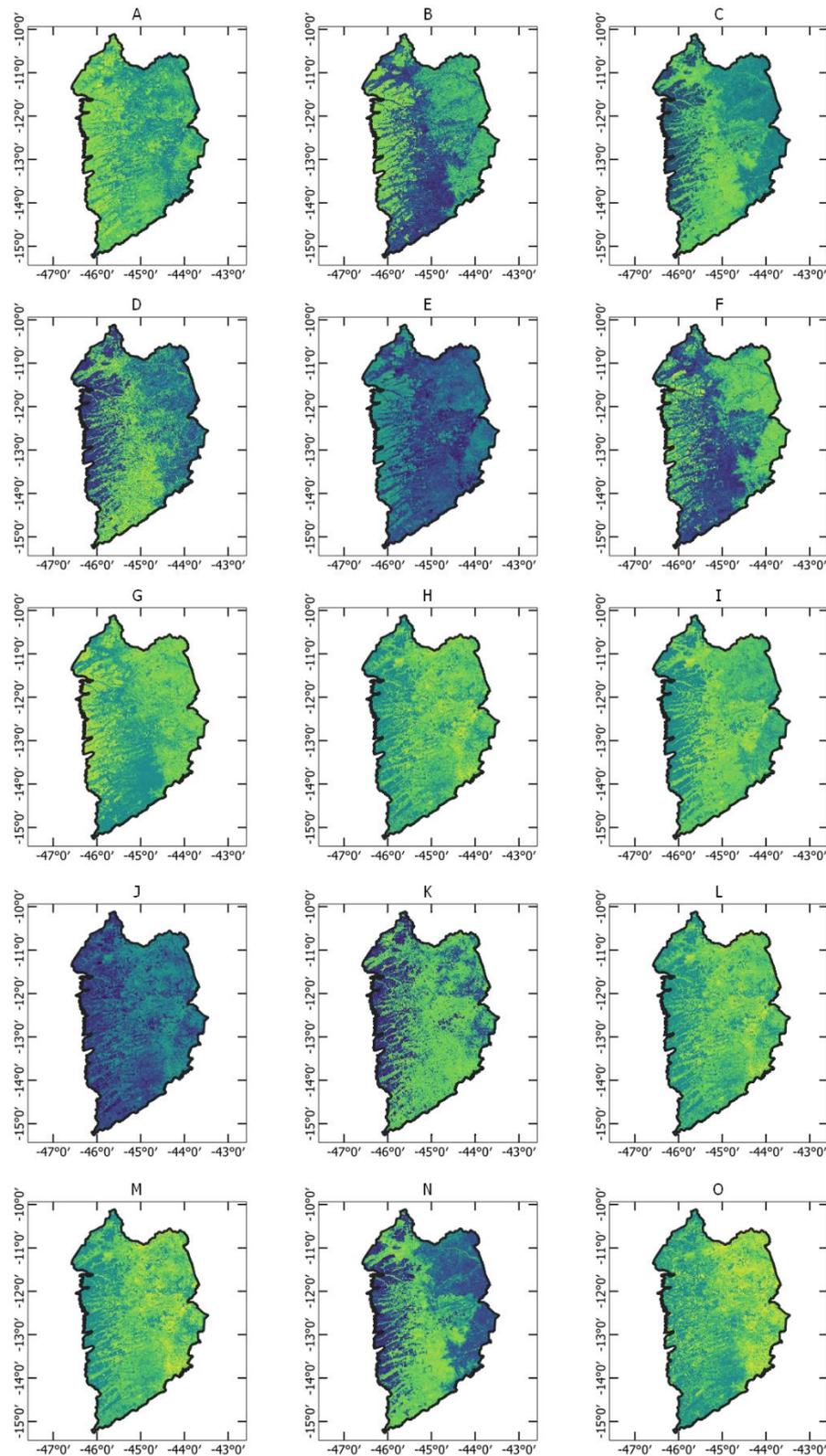


Figura 21 Imagens dos atributos derivados de STV-Trend com métricas básicas.

Nota: A - Soma absoluta dos valores ao longo da ST, B - Derivada média absoluta, C - Diferença entre os valores máximo e mínimo da ST, D - Valor máximo da primeira inclinação do ciclo, E - Primeiro quartil da ST, F - Intervalo interquartil da ST, G - Valor máximo da ST, H - Valor médio da ST, I - Valor mínimo da ST, J - Energia espectral média, K - Assimetria da ST, L - Soma dos valores ao longo do ciclo, M - Segundo quartil da ST, N - Desvio padrão da ST, O - Terceiro quartil da ST.

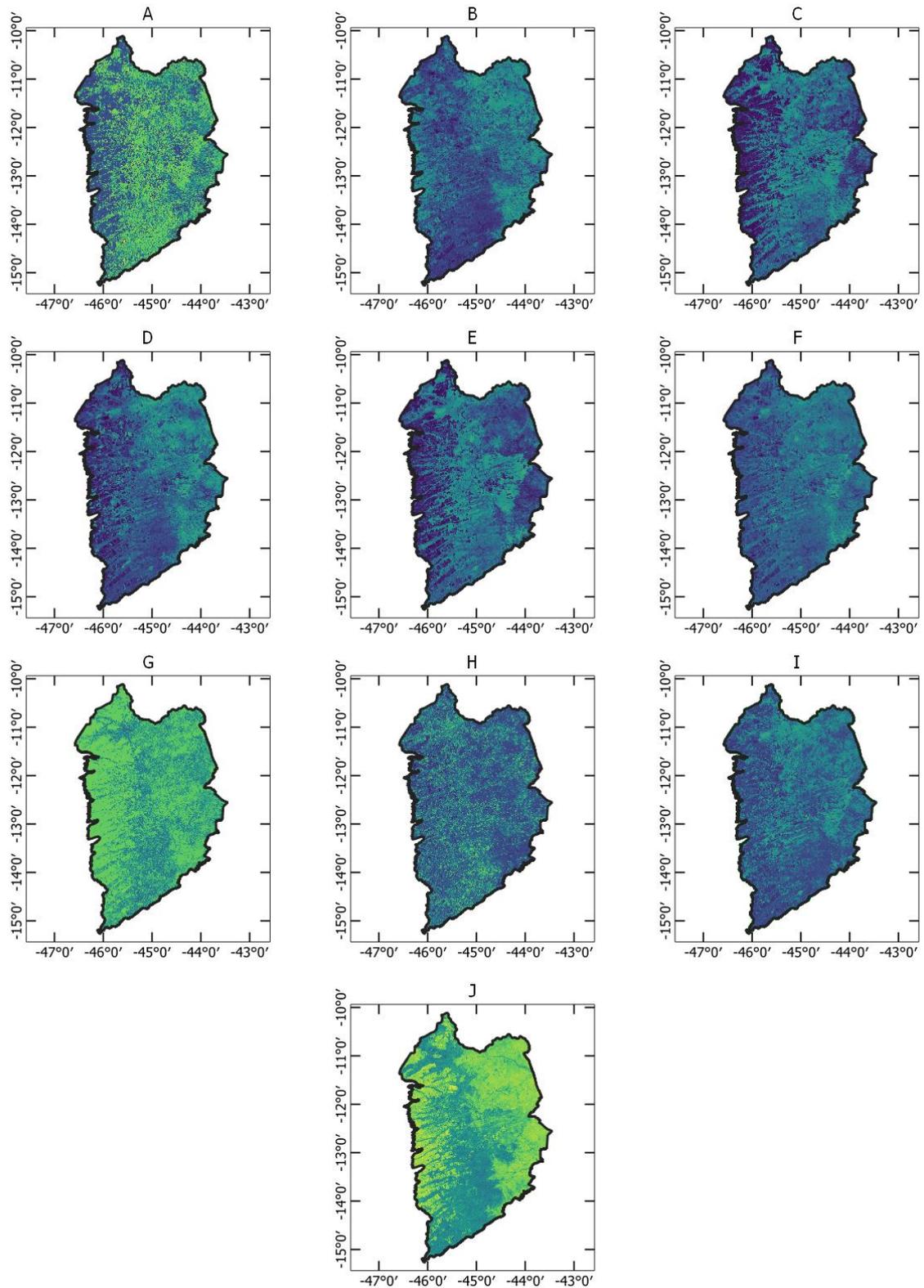


Figura 22 Imagens dos atributos derivados de STIV-Trend com métricas polares.

Nota: A - Ângulo principal da ST fechada na forma polar, B - Área da forma polar sobre o primeiro quadrante, C - Área da forma polar sobre o segundo quadrante, D - Área da forma polar sobre o terceiro quadrante, E - Área da forma polar sobre o quarto quadrante, F - Área total da forma polar da ST, G - desvio padrão de um objeto em relação a um círculo, H - Excentricidade, I - Distância média dos pontos ao centroide da forma, J - Equilíbrio polar, desvio padrão das áreas por estação.

Após agrupamento das classes do *Dataset* LEM+, foram extraídos os valores de STIV-SG (SG) e STIV-Trend (TD) para as novas rotulações, culturas anuais e outras classes para análise de variabilidade e dissimilaridade das STIV. Os gráficos boxplot (Figura 23) para os valores de índice de vegetação e de análise de dissimilaridade das classes agrupadas.

A classe com culturas anuais possui uma variabilidade maior no índice de vegetação ao longo do tempo em relação a outras classes, devido ao comportamento sazonal das culturas agrícolas. Enquanto a dissimilaridade das STIV é menor nas culturas anuais, mesmo com variação no passo da curva de NDVI existe similaridade na classe, devido à capacidade do DTW comparar séries com intervalos temporais diferentes.

O agrupamento das classes de UCT do *Dataset* LEM+ foi realizado com intuito de minimizar problemas com amostragens irregulares e grandes distorções temporais. Essas irregularidades podem ser causadas pelo imageamento dos sensores dos satélites ou pelo comportamento característico das classes, haja vista a diferença em datas de plantio de uma mesma cultura ou estágio vegetativo.

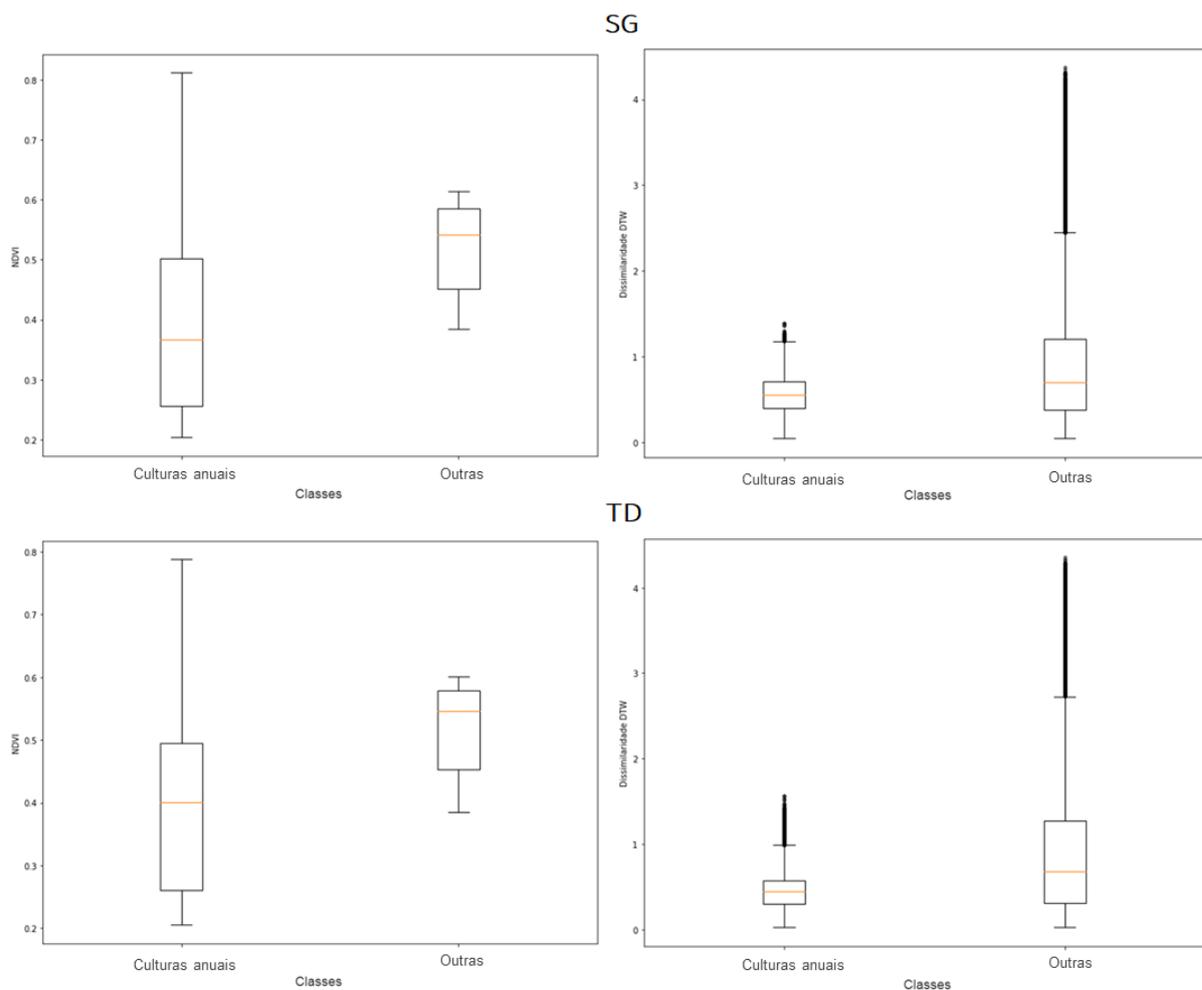


Figura 23 Boxplot de distribuição do valor de NDVI e de distribuição da dissimilaridade das classes amostrais agrupadas em culturas anuais e outras classes.

A avaliação da importância atributos com o algoritmo Boruta permitiu a otimização do processo de seleção de atributos que melhor descreve as classes de amostragem. O processo também reduziu a quantidade de atributos utilizados na tarefa de classificação em cada uma das dez combinações de atributos (Figura 24).

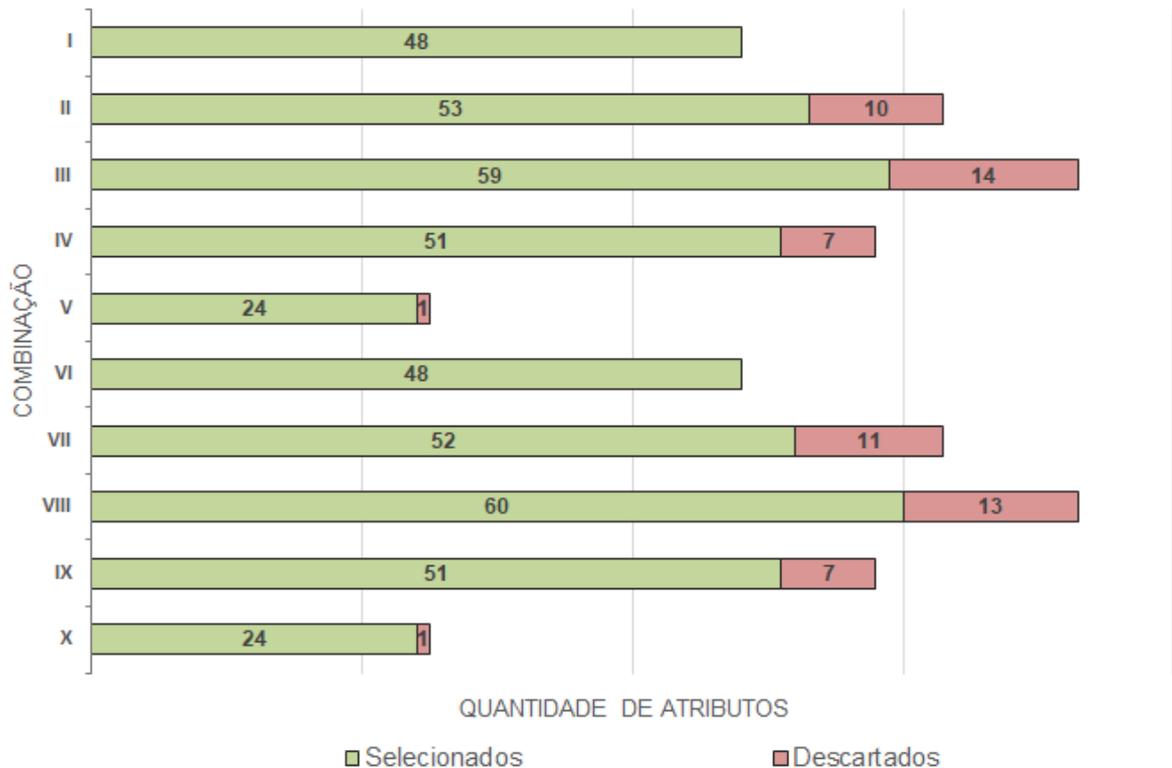


Figura 24 Quantidade de atributos selecionados e descartados em cada combinação de atributos testada.

O *ranking* da importância relativa dos atributos com maior relevância, para todas as combinações analisadas pelo algoritmo Boruta é apresentado na Tabela 7. O *ranking* foi construído pela soma acumulada da importância relativa de cada atributo, com limite de até 80% de relevância total, com base no princípio de Pareto. Nesse caso, a média percentual de causas, em cada combinação foi de 34%, ou seja, do total de atributos utilizados menos da metade são necessários para compor uma importância de até 80% na classificação.

Tabela 7 Importância acumulada dos atributos mais relevantes até 80%, pelo algoritmo Boruta em cada combinação de atributos

Tipo	Atributo	I	II	III	IV	V	VI	VII	VIII	IX	X
TS	20190914	6,6	3,2	4,5	7,9		6,9	2,7	2,4	7,6	
	20190922	8,2	6,3	4,6	8,3		9,7	6,9	6,4	10,3	
	20190930	4,9	5	3,6	6,2		6,6	6,6	6,2	6,9	
	20191008	6,1	6,2	2,4	6,8		4,6	5,7	3,2	7	
	20191016	7,9	4,3	6	4,2		5,7	3,9	4,2	4,7	
	20191024	3,8	4	1,4	6		4,5	3,1		6	
	20191101	3,6	3,2	2,1	2,9		5,1	4,1	2,5	4,2	
	20191109	2,3					2,8		1,7		
	20191117						4,2	1,5		2,5	
	20191125	2,8	1,7	2			4,2	3,1	1,7	2,4	
	20191203	3,9	3,7	2,3	2,3		5,6	3,6	1,6	3,4	
	20191211	1,7			3,2		1,8	1,5	1,3	2,4	
	20191219	3,2									
	20200202							1,6			1,6
	20200210	2,1						2,5			
	20200218	3,2	1,4		1,9		3,2	1,7	1,5	3,3	
	20200226	3,2	1,4	1,3	2,2		3,1			2,1	
	20200406	2,3	1,5								1,9
	20200414			1,9	2,2				1,7		
	20200422	2,6			1,7						1,8
	20200804							1,6			
	20200812				1,8						
	20200905			1,4							
	20200913	4,7	1,2	2,4	2		2		1,3		
	20200921	7,5	3,8	4,4	2,9		4,6				
	20200929			1,6	3,6					2,3	3,1
BS	AMD							7,9	11,6		22,1
	Amplitude		9,2	5,4		11,1		7,4	4,2		8
	F. Slope							7	10,7		15,5
	FQR		9,1	10,8		15,2		3,9	3,4		7,1
	IQR								1,6		3,6
	Max		2,9	2,4		4,9		1,6			
	Min			2,1		5,5			2		3,6
	Skew		2,3			5		1,6			
	SQR			1,8		3,9					
	Std		9,8	9,3		18,7		7,6	7,4		10,8
PL	Area Q1			1,5	2,3						
	Area Q2				4	10,4				2,8	5,5
	G. Rad.			1,4							
	Polar Balance			4,2	8,6	7,9			1,2	6,6	5,3
	Acumulado	80,6	80,2	80,8	81,1	82,5	80,4	81,2	80,1	80,5	81,5

A combinação VIII apresentava 73 atributos dos quais 60 foram selecionados para treinamento do modelo RF de classificação na *pipeline* de MD, com descarte de 13 atributos pelo algoritmo Boruta, durante o processo de análise de importância e seleção dos atributos em cada combinação. A identificação dos atributos descartados nas combinações pode ser visualizada na (Figura 25), com exceção das combinações (I e VI) que possuíam apenas atributos de STIV, as demais apresentaram descarte de dados que faziam parte da série temporal.

Tipo	Atributo	Combinação									
		I	II	III	IV	V	VI	VII	VIII	IX	X
TS	20200516		■	■		■		■	■		■
	20200524		■	■		■		■	■		■
	20200601		■	■		■		■	■	■	■
	20200609		■	■	■	■		■	■	■	■
	20200617		■	■		■		■	■	■	■
	20200625		■	■	■	■		■	■		■
	20200703		■	■	■	■		■	■	■	■
	20200711		■	■	■	■		■	■	■	■
	20200719		■	■	■	■		■	■	■	■
	20200727		■	■		■		■	■		■
20200804			■		■			■		■	
BS	tqr	■			■		■	■		■	
PL	angle	■	■	■	■	■	■	■	■	■	■
	area_q4	■	■	■			■	■	■		
	csi	■	■	■	■		■	■			

Legenda

■ Atributo descartado (Boruta)
■ Não utilizado na combinação

Figura 25 Atributos descartados pelo algoritmo Boruta em cada combinação de atributos testada na *pipeline* de mineração de dados.

Nota: TS – imagens de série temporal de índice de vegetação; BS – imagens de métricas básicas derivadas de STIV com STMetric; PL – imagens de métricas polares derivadas de STIV; TQR – imagem da métrica de terceiro quartil da STIV; ANGLE – imagem do ângulo principal da STIV fechada na forma polar; AREA_Q4 – área da forma polar sobre o quarto quadrante; CSI - Índice de forma celular - medida quantitativa adimensional de morfologia, desvio padrão de um objeto em relação a um círculo.

Com o *GridSearchCV*, foi possível obter os melhores hiperparâmetros em cada combinação de atributos otimizando o processo de modelagem. Esse processo possibilitou a identificação do número ideal de estimadores. Utilizando a função *GridSearchCV* para otimização do desempenho de modelos (WANG; ZANG; TIAN, 2020), foram obtidos os melhores resultados e hiperparâmetros de maneira automatizada, em uma tarefa de classificação. Analisando-se a Tabela 8, nota-se que as combinações II, IV, VIII e X

necessitaram de uma menor quantidade de estimadores (500 estimadores) dentre as demais combinações.

Tabela 8 Melhores hiperparâmetros identificados com *GridSearchCV* para a tarefa de classificação com RF

Combinação	Bootstrap	Criterion	Max_features	N_estimators
I	True	gini	log2	2000
II	False	entropy	log2	500
III	False	entropy	sqrt	1500
IV	False	gini	log2	500
V	True	entropy	log2	1500
VI	True	entropy	sqrt	1500
VII	True	entropy	sqrt	1000
VIII	True	gini	sqrt	500
IX	False	gini	log2	2000
X	True	gini	sqrt	500

Observando-se as métricas dispostas na Tabela 9, as combinações de atributos que apresentam valores acima de 98% para o Kappa são as II, III, VIII, que apresentam os menores erros de comissão e maiores valores de F1-score e acurácia acima de 99%. Além disso, a combinação VIII apresenta o menor erro de omissão, indicando que essas combinações apresentam, de maneira geral, os melhores resultados de classificação na tarefa de MD.

Tabela 9 Métricas de avaliação geradas a partir da matriz de confusão dos testes de classificação, utilizando RF em cada combinação de atributos de STIV

Combinação	Kappa	Recall	Precisão	Erro de omissão	Erro de comissão	F1-Score	Acurácia
I	0,960	0,972	0,988	0,028	0,012	0,980	0,980
II	0,980	0,988	0,992	0,012	0,008	0,990	0,990
III	0,980	0,988	0,992	0,012	0,008	0,990	0,990
IV	0,968	0,988	0,980	0,012	0,020	0,984	0,984
V	0,968	0,980	0,988	0,020	0,012	0,984	0,984
VI	0,968	0,988	0,980	0,012	0,020	0,984	0,984
VII	0,976	0,988	0,988	0,012	0,012	0,988	0,988
VIII	0,984	0,992	0,992	0,008	0,008	0,992	0,992
IX	0,976	0,988	0,988	0,012	0,012	0,988	0,988
X	0,972	0,980	0,992	0,020	0,008	0,986	0,986

O mapeamento obtido com a classificação foi gerado utilizando-se a combinação de atributos com as melhores métricas de avaliação, geradas a partir da MC, para o conjunto de dados de treinamento em relação ao conjunto de dados de teste (Tabela 9). As métricas de avaliação demonstram maiores precisão e acurácia em combinações com STIV e seus atributos derivados, tanto básico, quanto polares.

Com a combinação VIII, que contempla a STIV-Trend (TD), os AD com métricas básicas e polares, gerou-se o mapeamento das culturas anuais. A combinação VIII apresentou métricas de avaliação mais vantajosas, enquanto a modelagem otimizada do RF contou com um baixo número comparativo de estimadores no processo de aprendizagem ($n_{\text{estimators}} = 500$). Com base nos testes realizados verifica-se que, combinações de atributos contando apenas as STIV (I e VI) demandaram maior número de estimadores no processo de classificação, assim como combinações com número reduzido de atributos (IV, IX).

A combinação VIII apresentou um valor de Kappa de 98,4% que, segundo Landis e Koch (1977), são valores excelentes para qualidade de mapeamento. A combinação também obteve precisão e acurácia de 99,2%, assim como menores métricas de erros de omissão e de comissão em 0,8%. No entanto, as métricas de avaliação calculadas após o processo de inspeção visual de pontos aleatoriamente sobrepostos. Resultaram em Kappa de 86%, com Precisão de 92,2% e Acurácia em 92% para o mapeamento das culturas de anuais.

Os valores obtidos nas métricas de avaliação são semelhantes aos encontrados por Wu et al. (2021) que, ao mapearem o uso da terra, tiveram precisão de classificação de 98,38% utilizando RF em séries temporais de NDVI para classificação e extração de terras agrícolas (áreas agricultáveis) no território da China. Entretanto, Viana, Girão e Rocha (2019) não auferiram valores consistentes de precisão no mapeamento de classes agrícolas, pois no mapeamento de lavouras temporárias a precisão foi acima de 74% e as classes do estudo envolviam sistemas agroflorestais (agricultura em conjunto com floresta), mapeados com STIV provenientes de imagens de satélite com resolução temporal de 16 dias e espacial de 30 m.

O método RF também foi utilizado por Wang, Zang e Tian (2020), com STIV de imagens MODIS, adicionando parâmetros fenológicos e informações sobre umidade do solo. Os autores realizaram o mapeamento de áreas com arroz em casca do norte da planície de Songnen. A combinação de atributos utilizada teve precisão e Kappa superiores a 96% e 90%, respectivamente, em seu estudo.

Em seguida, foi aplicado o filtro majoritário para eliminar *pixels* isolados nos confrontos de classes e redução do efeito quadriculado dos *pixels* nos limites dos talhões com a simplificação geométrica, para suavização das bordas. É possível notar (Figura 26), por inspeção visual, a suavização dos limites das geometrias com menor efeito *pixelado*, contornos mais suaves e a supressão de *pixels* isolados. Os processos aplicados aumentam a eficiência de visualização do mapeamento resultante em sistemas de informações geográficas (ZHU et al., 2019), definindo melhor os talhões quando se comparam com imagens de maior resolução, como o mapa de base do *Google Maps* sob o município de Formosa do Rio Preto – BA.

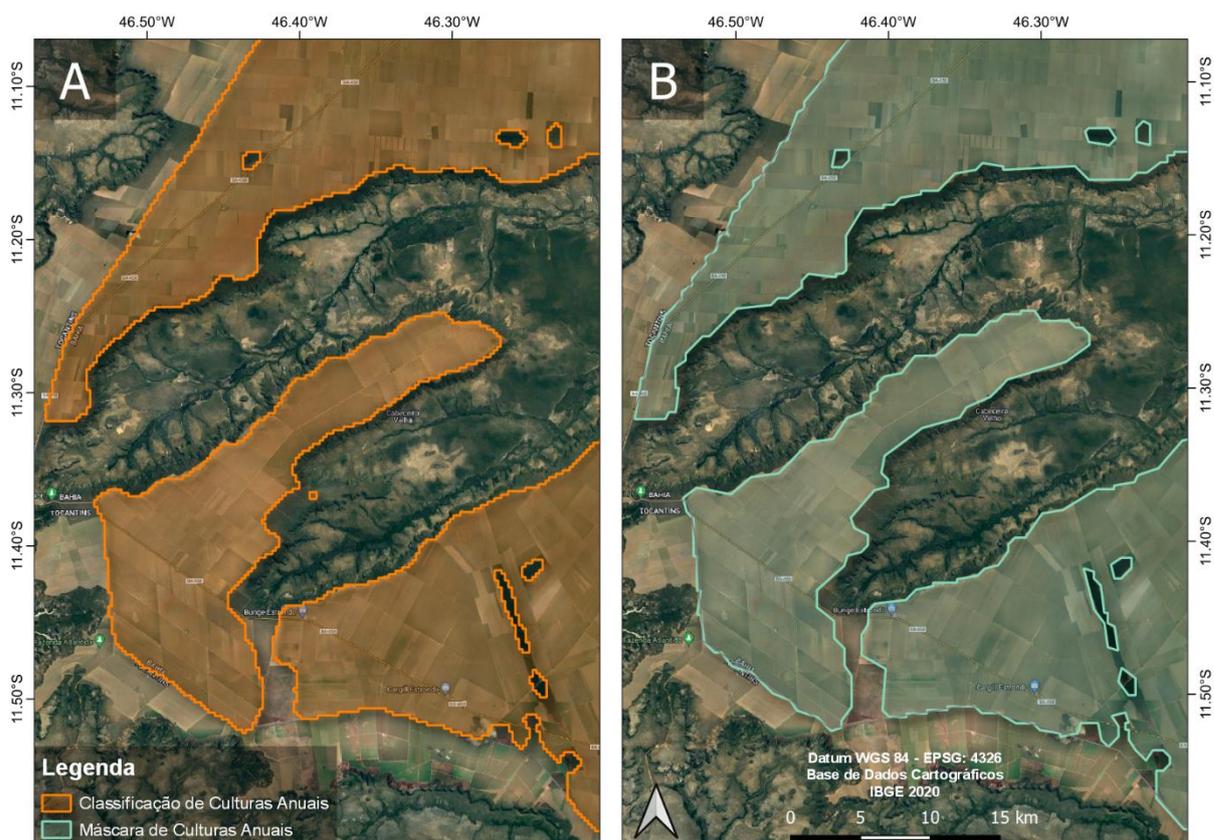


Figura 26 Comparativo classificação de culturas anuais (A), máscara de culturas anuais (B), após aplicação do filtro majoritário e simplificação geométrica.

Os processos também auxiliaram na eficiência de armazenamento estático dos arquivos físicos (ZHU et al., 2019). A classificação em formato TIFF ocupava 3.135 KB; após a poligonização para formato geoJSON o tamanho espaço utilizado estava em 2.221 KB. Com a simplificação geométrica houve uma redução de 63,08%, obtendo-se uma máscara em formato geoJSON e 820 KB de espaço físico. A redução de armazenamento necessário se dá ao menor número de vértices que constituem as geometrias, esse fator impacta positivamente em novas operações computacionais e análises espacial com a máscara (ZHU et al., 2019), como recorte ou extração de dados municipais.

O mapeamento das áreas com culturas anuais para o ano-safra 2019/2020 da mesorregião Oeste da Bahia é apresentado na Figura 27. A área identificada no mapeamento foi de 2.533.573 hectares, o que representa uma diferença de 12,5% em relação aos dados oficiais do IBGE, com área de 2.251.661 ha.

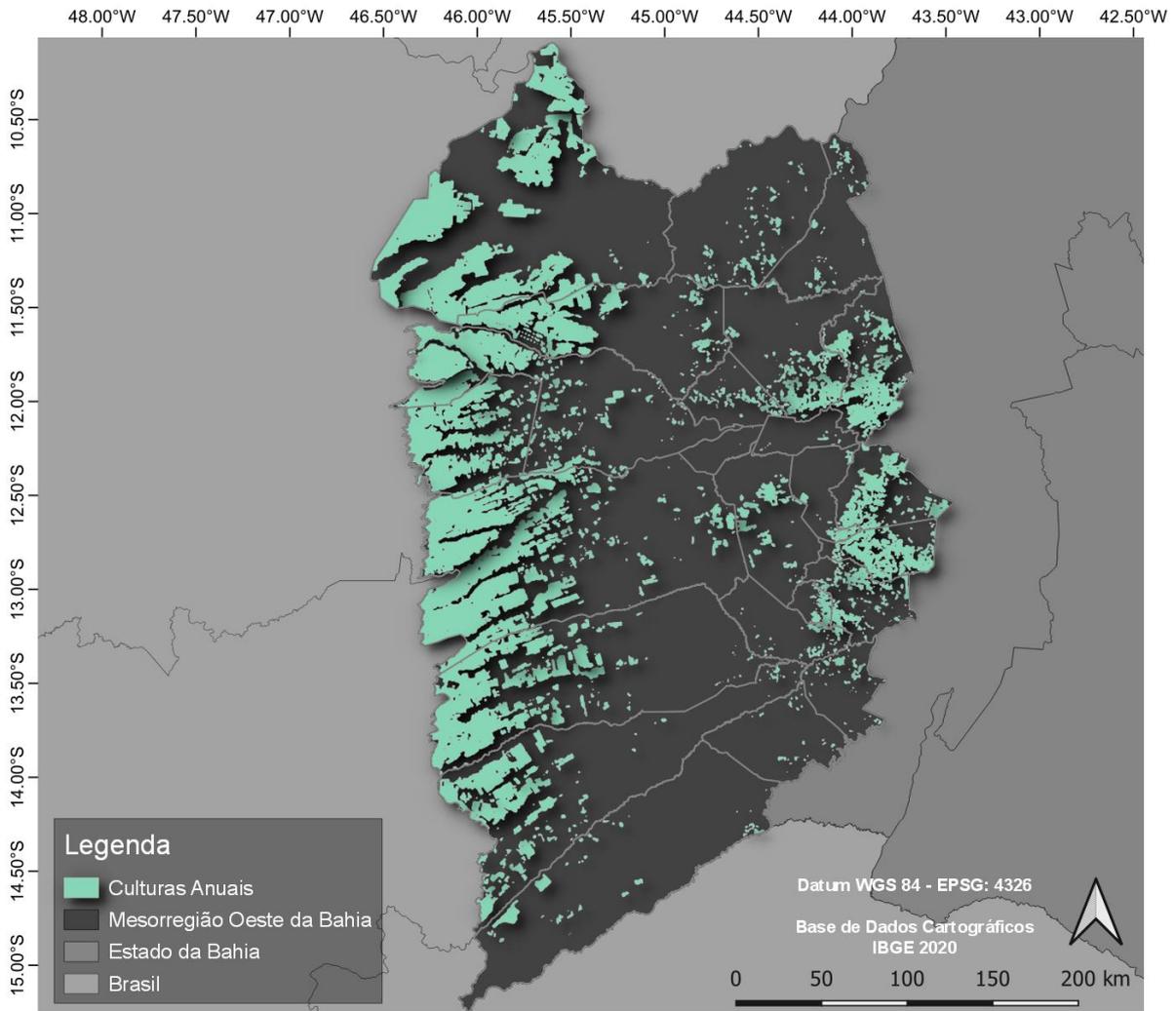


Figura 27 Mapa com culturas anuais para o ano-safra 2019/2020 da mesorregião Oeste da Bahia.

Comparando as áreas extraídas pela máscara de classificação e dados oficiais do IBGE (Tabela 10) notam-se valores variando de -93,4% a 46.468%, amplitude devida a fatores regionais. A falta de amostragem de culturas no Dataset LEM+ pode impactar e gerar superestimativa de valores ou uma divergência na base de dados do IBGE em alguma das culturas, podendo impactar em subestimativa de área. Valores acima de 100% indicam superestimativa e inferiores apontam subestimativa dos dados da máscara, em relação aos dados oficiais do IBGE.

Tabela 10 Área mapeada, em hectares (ha), com distribuição municipal extraída da máscara de classificação em comparativo com dados oficiais do IBGE

id	Município	Mapeamento (ha)	IBGE (ha)	$\left(\frac{\text{Mapeamento} - \text{IBGE}}{\text{IBGE}}\right)$
1	Angical	5.342	2.290	133,3%
2	Baianópolis	25.924	22.861	13,4%
3	Barreiras	231.263	253.150	-8,6%
4	Brejolândia	74.044	159	46.468,7%
5	Canápolis	13.594	570	2.284,9%
6	Catolândia	30	461	-93,4%
7	Cocos	24.420	27.809	-12,2%
8	Coribe	3.594	815	340,9%
9	Correntina	291.414	269.338	8,2%
10	Cotegipe	65.246	3.600	1.712,4%
11	Cristópolis	1.765	2.450	-28,0%
12	Formosa do Rio Preto	518.448	514.792	0,7%
13	Jaborandi	173.918	148.192	17,4%
14	Luís Eduardo Magalhães	207.038	227.047	-8,8%
15	Mansidão	13.441	1.330	910,6%
16	Riachão das Neves	153.935	173.385	-11,2%
17	Santa Maria da Vitória	11.319	1.870	505,3%
18	Santa Rita de Cássia	6.272	4.924	27,4%
19	Santana	42.338	4.510	838,8%
20	São Desidério	544.860	584.218	-6,7%
21	São Félix do Coribe	9.783	1.298	653,7%
22	Serra Dourada	6.218	1.250	397,4%
23	Tabocas do Brejo Velho	2.740	2.250	21,8%
24	Wanderley	106.626	3.095	3.345,1%
Total		2.533.573	2.251.661	12,5%

O teste de normalidade de Anderson-Darling foi realizado nos dados de diferença entre área mapeada e dados do IBGE. O teste indicou que os dados não apresentam normalidade, com p-valor = 0,701 a 5% de significância. Portanto, foi utilizado o coeficiente de correlação de Spearman [rs] para análise da relação entre os dados de mapeamento e oficiais do IBGE. A correlação identificada entre dados oficiais e os dados extraídos da máscara foi de 0,7496, com p-valor = $2,49e^{-6}$.

Quando as informações utilizadas são de uma única imagem elas se tornam ineficientes; portanto o uso de uma série de imagens associada a índices de vegetação em serie temporal tem um potencial considerável na utilização para classificação de imagem (NIAZMARDI et al., 2018). Os componentes da STIV para um *pixel* com cultura de ciclo anual, permitem a identificação do padrão das safras (crescimento e senescência da planta) no período analisado, facilitando o processo de classificação de STIV e suas métricas.

Na Figura 29 podem ser visualizadas as máscaras de classificação de áreas com culturas anuais no imóvel para os testes de mapeamento caso M1, caso M2 e caso M3. Os três casos de classificação realizados em um imóvel comparando com sua área de uso

consolidado (Figura 29) demonstram as máscaras de classificação da área, nos diferentes casos: Caso M1 – recorte do imóvel na máscara classificada; Caso M2 – processo de MD com o modelo aplicado sobre as imagens de STIV-Trend, antes da simplificação geométrica; Caso M3 – processo de MD com o modelo aplicado sobre as imagens de STIV-Trend, após simplificação geométrica.

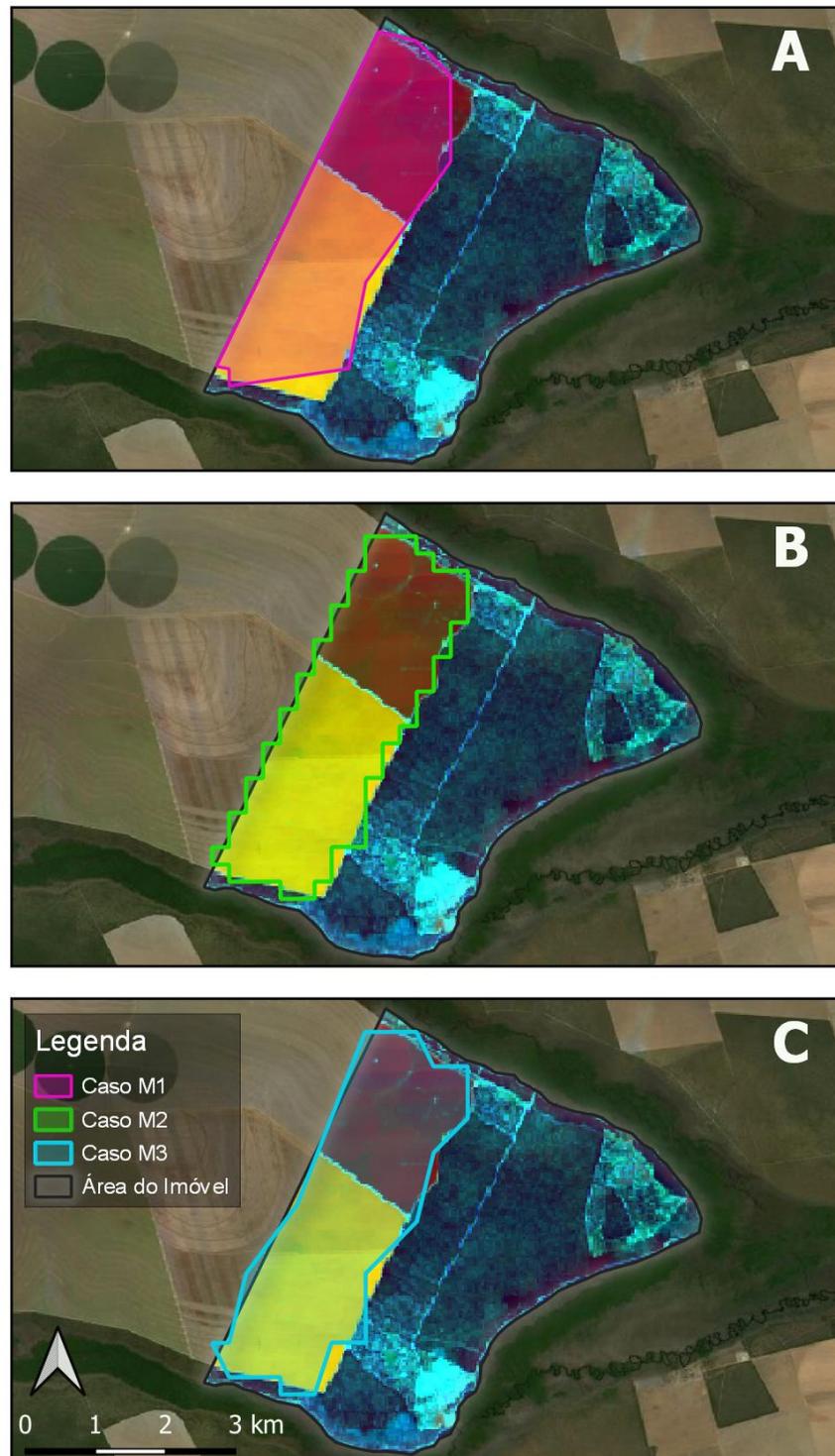


Figura 28 Máscaras de classificação de área com culturas anuais no imóvel, nos 3 casos de mapeamento realizados: A – caso M1, B – caso M2; C – caso M3.

As métricas de avaliação calculadas para aferição dos mapeamentos com culturas anuais no imóvel nos casos de mapeamento M1, M2 e M3 podem ser verificadas na Tabela 11.

Tabela 11 Métricas de avaliação geradas a partir da matriz de confusão dos mapeamentos do imóvel BA-2928901-7F36C7E43A2F4CE1A713A08327A5478A

Mapeamento	Kappa	Recall	Precisão	Erro de Omissão	Erro de Comissão	F1-Score	Acurácia
Caso M1	0,9263	0,9091	1	0,0909	0	0,9524	0,9662
Caso M2	0,9832	0,979	1	0,021	0	0,9894	0,9922
Caso M3	0,9550	0,9441	1	0,0559	0	0,9712	0,9792

Observou-se que o mapeamento segue os limites do imóvel, devido ao recorte do imóvel da máscara de culturas anuais realizado para a região Oeste da Bahia. Contudo, os limites desse mapeamento apresentam concavidades mais acentuadas, reduzindo a área mapeada. Essa ocorrência se deve à simplificação geométrica aplicada, pois o processo é realizado em cada polígono contido na classificação e é diretamente afetado pelo seu tamanho e quantidade de vértices.

Á área de uso consolidado no imóvel é utilizada para o cultivo agrícola com culturas anuais totalizando de 852,82 hectares, com base no demonstrativo do imóvel. No cálculo das áreas mapeadas em cada máscara, a diferença percentual em relação à área de uso consolidado do imóvel é apresentada na Tabela 12.

Tabela 12 Área calculada nas máscaras mapeadas com culturas anuais, em comparativo com área de uso consolidado do imóvel

Geometria / mapeamento	Área (ha)	$\left(\frac{\text{Caso} - \text{Uso consolidado}}{\text{Uso consolidado}}\right)$
Uso consolidado	852,820	
Caso M1	792,770	-7,0%
Caso M2	848,145	-0,5%
Caso M3	827,509	-3,0%

O caso M1 apresentou a maior diferença percentual de área mapeada em relação à área de uso consolidada do imóvel. Nesse caso, nota-se o impacto da diminuição dos vértices que causa a geração de um falso-negativo, ocasionando erros de omissão na classificação. Esse caso apresentou o maior valor para erro de omissão em 9,09%. Entretanto, o caso M2 apresentou o melhor desempenho com valores de Kappa em 98,32%, precisão aproximada de 100% e acurácia de 99,22% (Tabela 11), em relação aos outros casos de mapeamento, nos quais foi possível obter a maior área mapeada com culturas anuais mapeada dentro do imóvel.

O caso M2 resulta do mapeamento utilizando recorte das imagens de STIV-Trend com atributos da combinação VIII e aplicação do processo de MD com importação do classificador RF. Verifica-se na sobreposição com as imagens OLI falsa-cor 564 que mantém um bom alinhamento dos limites da região com área agricultável do imóvel em relação aos limites do polígono. Também é possível observar que não houve mapeamento em outras regiões do imóvel, evidenciando o baixo valor para erros de comissão, quando no processo de classificação são mapeadas classes erradas pelo algoritmo, ou seja, baixo índice de falso-positivo.

Com 3% de diferença percentual no tamanho da área mapeada, em relação à área de uso consolidado do imóvel, o caso M3 apresenta uma boa delimitação da área com culturas anuais, porém devido à simplificação geométrica, resolução espacial das imagens MODIS utilizadas no mapeamento e a quantidade de vértices no polígono, houve subestimativa da área da máscara para o imóvel. Com nível de acurácia intermediário, em relação aos outros casos estudados (LANDIS; KOCH 1977), o mapeamento resultante do caso M3 ainda apresenta valores excelentes para Kappa: 95,5%.

De acordo com Luo et al. (2023), é possível a geração de mapas de cultivo, em escalas compatíveis com os fragmentos de campos agrícolas e levando em conta a complexidade da estrutura da região com dados de satélite e algoritmos de aprendizagem de máquina. Aliados a isso, os classificadores de aprendizado de máquina combinados com a seleção de recursos fornecem um método eficiente, preciso e prático para discriminação de culturas com alta resolução, usando séries temporais de índice de vegetação em regiões agrícolas com sistemas de cultivo misto (LUO et al., 2023).

Em todas as linhas de crédito agrícola as instituições financeiras têm exigido os recibos de inscrição do imóvel rural no CAR para localizar e fiscalizar a aplicação dos recursos financeiros, nos últimos quatro anos (SICAR, 2022). As técnicas aplicadas para geração de mapeamentos com culturas anuais, e em seguida aplicadas ao caso M1, podem ser replicadas para auferir informações acerca das áreas de uso consolidado e destinadas ao plantio de empreendimentos fomentados com custeio agrícola.

Com o mapeamento de uma região e máscara de áreas agricultáveis é possível realizar o recorte do imóvel e consultar essa variável. Esse processo embora mais ágil e útil, pode ter menor precisão quanto aos valores mapeados. O processo pode ser adaptado para a etapa de pré-concessão de crédito agrícola, com intuito de validar a existência da área de plantio no imóvel e área mínima compatível com o recurso solicitado.

A linha de crédito para custeio agrícola demanda por informações para fiscalizar as áreas de plantio dos empreendimentos pós-contratação do crédito. Essas áreas devem ser analisadas e monitoradas pelas concedentes, sendo possível o uso de sensoriamento remoto (MCR 2-7-4-“a”-II) (BACEN, 2021). Técnicas de mineração de dados e análise de séries temporais de índice de vegetação podem suportar as demandas de técnicas do setor.

As instituições financeiras podem, ao longo do tempo, formar um grande banco de dados para amostragem e treinamento de algoritmos com a tarefa de classificação realizando a perpetuidade dos modelos já treinados e validados. Esses modelos podem ser aplicados diretamente nas imagens de STIV e métricas para detectar com maior precisão as áreas com culturas anuais de cada imóvel, assim como realizado no caso M2, com custeio agrícola concedido e utilizar essa informação no monitoramento e fiscalização dos empreendimentos.

O CAR tem mais de 10 anos, desde sua criação, e ainda é um desafio enfrentado pelos órgãos ambientais e governamentais na análise e validação dos imóveis em todo território brasileiro, devido ao déficit nos processos de validação, com base na legislação ambiental (BRASIL, 2022). Os processos de validação podem ser otimizados utilizando técnicas semelhantes as desta pesquisa, melhoradas com a utilização de plataformas com tecnologias com maior capacidade de escala. Infraestruturas com arquitetura em nuvem como o *Google Earth Engine*[™] (GEE) e a *Amazon Web Services*[™] (AWS) podem prover dados de STIV, realizar etapas de mineração de dados e disponibilizar informações para consumo analítico.

6 CONCLUSÕES

Considerando os objetivos e a metodologia empregada nesta pesquisa, é possível concluir que:

A mineração de dados em séries temporais de índices de vegetação e técnicas de engenharia e extração de recursos forneceu uma compreensão ágil com melhor precisão na distinção de áreas com culturas anuais, o que possibilita a criação de processos e modelos replicáveis.

O mapeamento das culturas anuais permitiu a estimativa de áreas e melhoria da gestão do território no oeste do estado da Bahia.

O levantamento de áreas agrícolas utilizando métodos de aprendizagem de máquina e séries temporais de índice de vegetação é de extrema importância para mitigar impactos em regiões de expansão agrícola do Brasil.

Mapeamentos georreferenciados da agricultura, que podem ser utilizados em diversas plataformas de sistemas de informações geográficas, estudos ambientais e logísticos podem ser realizados de maneira simplificada.

O mapeamento de áreas com culturas anuais em uma região pode beneficiar empresas e órgãos públicos diretamente. As empresas podem melhorar e acelerar seu planejamento estratégico, dimensionando a equipe técnica, os estoques e a infraestrutura de armazenagem de grãos com base nas informações sobre suas áreas de atendimento. A agilidade na geração de dados sobre regiões agrícolas pode favorecer órgãos públicos que podem preparar políticas mais precisas para o agronegócio.

As técnicas e processos aplicados nesta pesquisa podem auxiliar no procedimento de análise de concessão do crédito agrícola. As técnicas de mineração de dados em séries temporais de índices de vegetação podem auxiliar na aferição e validação de áreas de uso consolidado em imóveis rurais que estejam destinadas aos empreendimentos com culturas anuais no custeio agrícola. Enquanto, outros tipos de operação de crédito agrícola podem ganhar mais agilidade nos processos, tendo em vista o potencial escalável das técnicas.

REFERÊNCIAS

- AJADI, O. A.; BARR, J.; LIANG, S.; FERREIRA, R.; KUMPATLA, S.P.; PATEL, R.; SWATANTRAN, A. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 97, A. 102294, 2021.
- ARTUZO, F. D.; FOGUESATTO, C. R.; SOUZA, A. R. L.; SILVA, L. X. Gestão de custos na produção de milho e soja. **Revista Brasileira de Gestão de Negócios**. São Paulo- SP, v. 20, n. 2, p. 273-294, 2018.
- ASSOCIAÇÃO DE AGRICULTORES E IRRIGANTES DO OESTE DA BAHIA – AIBA. **Principais culturas**. Dados e pesquisa. 2019. Disponível em: <<https://aiba.org.br/principais-culturas/>>. Acesso em: 26 jul. 2022.
- BANCO CENTRAL DO BRASIL – BACEN. **Manual de crédito rural**. Brasília – DF: Bacen. 7 de julho de 2021. Disponível em: <<https://www3.bcb.gov.br/mcr/completo>>. Acesso em: 8 de jul. 2021.
- BRASIL. Ministério da Economia. Banco Central do Brasil – BACEN. Resolução CMN Nº 4.895, de 26 de fevereiro de 2021. Atualiza e consolida as regras aplicáveis à fiscalização das operações de crédito rural pelas instituições financeiras e consolida as regras sobre desclassificação e reclassificação. **Diário Oficial da União**, Brasília – DF. 01/03/2021. Edição 39, Seção 1, p, 84, 2021. Disponível em: <<https://www.in.gov.br/en/web/dou/-/resolucao-cmn-n-4.895-de-26-de-fevereiro-de-2021-305690055>>. Acesso em: 8 Jul. 2021.
- BARGIEL, D. Capabilities of high-resolution satellite radar for the detection of semi-natural habitat structures and grasslands in agricultural landscapes. **Ecological Informatics**. v. 13, p. 9-16, 2013.
- BELDA, S.; PIPIA, L.; MORCILLO-PALLARÉS, P.; RIVERA-CAICEDO, J. P.; AMIN, E.; GRAVE, C.; VERRELST, J. DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection. **Environmental Modelling e Software**, v. 127, 14 p., 2020.
- BELIK, W. O financiamento da agricultura familiar no Brasil na década de 2000. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL, 52., 2014, Goiânia. **Anais...** Brasília: Sober 2014. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/3407/1/td_2028.pdf>. Acesso em: 16 de out. 2019.
- BEM, P. P.; CARVALHO JÚNIOR, O. A.; CARVALHO, O. L. F.; GOMES, R. A. T.; GUIMARÃES, R. F.; PIMENTEL, C. M. M. Irrigated rice crop identification in southern Brazil using convolutional neural networks and Sentinel-1 time series. **Remote Sensing Applications: Society and Environment**, v. 24, A. 100627, 2021.
- BENDINI, H.; FONSECA, L. M. G.; SCHWIEDER, M.; KÖRTING, H. S.; RUFIN, P.; SANCHES. I. A.; LEITÃO, P. J.; HOSTERT, P. Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 82, A. 101872, 2019.

BERVEGLIERI, A.; IMAI, N. N.; CHRISTOVAM, L. E.; GALO, M. L. B. T.; TOMMASELLI, A. M. G.; HONKAVAARA, E. Analysis of trends and changes in the successional trajectories of tropical forest using the Landsat NDVI time series. **Remote Sensing Applications: Society and Environment**, v. 24, A. 101872. 2021.

BRASIL. Conselho Nacional do Ministério Público – CNMP. **Cadastro ambiental rural**. Diretrizes para atuação do Ministério Público. Brasília – DF. 2022. Disponível em: <https://www.cnmp.mp.br/portal/images/Publicacoes/documentos/2022/cadastro_ambiental_rural.pdf>. Acesso em: 26 de dez. 2022.

BREIMAN, L. Random Forests. **Machine Learning**. v. 45, p. 5-32, 2001.

BUDNIK, M.; GUTIERREZ-GOMEZ, E.-L.; SAFADI, B.; PELLERIN, D.; QUÉNOT, G. Learned features versus engineered features for multimedia indexing. **Multimedia Tools and Applications**, v. 76, n. 9, p. 11941–11958, 2017.

BUITINCK, LARS, LOUPPE, G.; BLONDEL, M.; PEDREGOSA, F.; MUELLE, A. et al. API design for machine learning software: experiences from the scikit-learn project. **European conference on machine learning and principles and practices of knowledge discovery in databases**, Prague, Czech Republic. Sep. 2013. Workshop: Languages for data mining and machine learning. 2013. p. 108-122. Disponível em: <<https://arxiv.org/abs/1309.0238>>. Acesso em: 4 out. 2021.

CAI, Y.; GUAN, K.; PENG, J.; WANG, S.; SEIFERT, C.; WARDLOW, B.; LI, Z. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. **Remote Sensing of Environment**, v. 210, p. 35-47, 2018.

CLEVELAND, R. B.; CLEVELAND, W. S.; MC RAE, J. E.; TERPENNING, I. STL: A seasonal-trend decomposition procedure based on Loess. **Journal of Official Statistics**, v. 6, n. 1, p. 3-73, 1990.

DEFEO, L. F.; ESQUERDO, J. C. D. M. Utilização de índices de vegetação NDVI e EVI do sensor MODIS para caracterização espectro-temporal de coberturas vegetais no Pantanal. SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 6, Geopantanal. Cuiabá – MT. 2016. **Anais...** Cuiabá, MT, Brasil: Embrapa Informática Agropecuária/INPE, 2016. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/150198/1/Geopantanal-Defeo-141.pdf>>. Acesso em: 27 out. 2019. p. 141-149.

DUARTE, L.; TEODORO, A. C.; MONTEIRO, A. T.; CUNHA, M.; GONÇALVES, H. QPhenoMetrics: An open source software application to assess vegetation phenology metrics. **Computers and Electronics in Agriculture**, v. 148, p. 82–94, 2018.

EGER, G. Z. S.; SILVA JUNIOR, G. C.; MARQUES, E. A.G.; LEÃO, B. R. C.; DA ROCHA, D. G.T.B.; GILMORE, T. E.; DO AMARAL, L. G. H.; SILVA, J. A. O.; NEALE, C. Recharge assessment in the context of expanding agricultural activity: Urucuia Aquifer System, western State of Bahia, Brazil. **Journal of South American Earth Sciences**, v. 112, A. 103601, 2021.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Zoneamento Agrícola de Risco Climático**, 1995. Disponível em: <<https://www.embrapa.br/busca-de-solucoes-tecnologicas/-/produto-servico/3933/zoneamento-agricola-de-risco-climatico---zarc>>. Acesso em: 29, abr. 2021.

ESQUERDO, J. C. D. M.; ANTUNES, J. F.G.; COUTINHO, A. C.; SPERANZA, E. A.; KONDO, A. K.; DOS SANTOS, J. L. SATVeg: A web-based tool for visualization of MODIS vegetation indices in South America. **Computers and Electronics in Agriculture**, v. 175, A. 105516, 2020.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, 1996a.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volume of data. **Communication of the ACM**, v. 39, n. 11, p. 27-34, 1996b.

FAYYAD, U.; STOLORZ, P. Data mining and KDD: Promise and challenges. **Future Generation Computer Systems**, v. 13, p. 99-115, 1997.

FONSECA, L. M.G.; KÖRTING, T. S.; BENDINI, H. DO N.; GIROLAMO-NETO, C. D.; NEVES, A. K.; SOARES, A. R.; TAQUARY, E. C.; MARETTO, R. V. Pattern recognition and remote sensing techniques applied to Land Use and Land Cover mapping in the Brazilian savannah. **Pattern Recognition Letters**, v. 148, p. 54-60, 2021.

FREITAS, P. L.; POLIDORO, J. C.; SANTOS, H. G.; PRADO, R. B.; CALDERANO, S. B.; GREGORIS, G.; MANZATTO, C. V.; DOWICH, I.; BERNARDI, A. C. C. Identificação e caracterização físico-química de latossolos de textura arenosa e média da região Oeste da Bahia. **Cadernos de Geociências**, Slavador – BA, v., n. 1-2, p. 83-93, 2014.

GIORGINO, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. **Journal of Statistical Software**, v. 31, p. 1-24, 2009.

HE, R.; ZHANG, L.; CHEW, A. W. Z. Modeling and predicting rainfall time series using seasonal-trend decomposition and machine learning. **Knowledge-Based Systems**, v. 251, A. 109125, 2022.

HOKAMA, R. H.; MIRANDA, C. G. Mitigação de riscos em operações crédito rural com o uso de sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO – SBSR, 19, 2019 Santos - SP. **Anais...** Santos, SP, Brasil: Instituto Nacional de Pesquisas Espaciais (INPE), 2019. Disponível em: <<https://proceedings.science/sbsr-2019/papers/mitigacao-de-riscos-em-operacoes-de-credito-rural-com-o-uso-de-sensoriamento-remoto>>. Acesso em: 16, out. 2019.

HOPPE, F.; HOHMANN, J.; KNOLL, M.; KUBIK, C.; GROCHE, P. Feature-based supervision of shear cutting processes on the basis of force measurements: evaluation of feature engineering and feature extraction. **Procedia Manufacturing**, v. 34, p. 847-856, 2019.

HUETE, A. R.; JUSTICE, C.; LIU, H. Development of vegetation and soil indices for MODIS-EOS. **Remote Sensing of Environment**, v. 49, n. 3, p. 224-234, 1994.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E. P.; GAO, X.; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices, **Remote Sensing of Environment**, v. 83, p. 195-213, 2002.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 2nd ed. Melbourne: OTexts, 2018. Disponível em: <<https://otexts.com/fpp2/>>. Acesso em: 09 set. 2021.

INSTITUTO DE ASTRONOMIA, GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS – IAG, Departamento de Astronomia. USP. **Início das estações do ano (2005-2020)**. 2021. Disponível em: <<https://www.iag.usp.br/astronomia/inicio-das-estacoes-do-ano>>. Acesso em: 04 out. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Coordenação de agropecuária. **Pesquisas agropecuárias**. 3. ed. Rio de Janeiro: IBGE, 2018. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101552.pdf>>. Acesso em: 26 maio 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Biomass continentais do Brasil**. 2021. Disponível em: https://geofp.ibge.gov.br/informacoes_ambientais/estudos_ambientais/biomass/documentos/Sintese_Descricao_Biomass.pdf; Acesso em: 04 out. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Sidra – Produção Agrícola Municipal (PAM - 2021). 2022. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/pam/tabelas>>. Acesso em: 10 de out. 2022.

JÖNSSON, P.; EKLUNDH, L. TIMESAT- a program for analyzing time-series of satellite sensor data. **Computers e Geosciences**, v. 30, n. 8, p. 833-845, 2004.

JUSTICE, C. O.; VERMOTE, E.; TOWNSHEND, J. R. G.; DEFRIES, R.; ROY, P. D.; HALL, D.K.; SALOMONSON, V.; PRIVETTE, J. L.; RIGGS, G.; STRAHLER, A.; LUCHT, W.; MYNENI, R. B.; KNYAZIKHIN, Y.; RUNNING, S. W.; NEMANI, R. R.; WAN, Z.; HUETE, A. R.; LEEUWEN, W. V.; WOLFE, R. E.; GIGLIO, L.; MULLER, J. P.; LEWIS, P.; BARNSLEY, M. The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. **IEEE Transactions on Geoscience and Remote Sensing**, v. 36, n. 4, p. 1228-1247, 1998.

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA—Geographic Data Mining Analyst. **Computers & Geosciences**, v. 57, p. 133-145, 2013.

KURSA, M.; RUDNICKI, W. Feature selection with the Boruta package. **Journal of Statistical Software**, v. 36, p. 1-13, 2010.

LAKSHMIPADMAJA, D.; VISHNUVARDHAN, B. Classification performance improvement using random subset feature selection algorithm for data mining. **Big Data Research**, v. 12, p. 1-12, 2018.

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

LEPRINCE, J.; MILLER, C.; ZEILER, W. Data mining cubes for buildings, a generic framework for multidimensional analytics of building performance data. **Energy and Buildings**, v. 248, A. 111195, 2021,

LIEN, L-C; DOLGORSUREN, U. Rule-based knowledge discovery of satellite imagery using evolutionary classification tree. **Journal of Parallel and Distributed Computing**, v. 147, p. 132-139, 2021.

LU, H.; RAUPACH, M. R.; McVICAR, T. r.; BARRETT, D. Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. **Remote Sensing of Environment**, v. 86, p. 1-18, 2003.

LUO, K.; LU, L.; XIE, Y.; CHEN, F.; YIN, F.; LI, Q. Crop type mapping in the central part of the North China Plain using Sentinel-2 time series and machine learning. **Computers and Electronics in Agriculture**, v. 205, A. 107577, 2023.

MEERT, W.; HENDRICKX, K.; VAN CRAENENDONCK, T.; ROBBERECHTS, P. **DTAIDistance**. Version v. 2. 2022. Disponível em: <<https://zenodo.org/record/7158824>>. Acesso em: 20 mar. 2022.

MERCURE, J.-F.; PAIM, M.A.; BOCQUILLON, P.; LINDNER, S.; SALAS, P.; MARTINELLI, P.; BERCHIN, I.I.; ANDRADE GUERRA, J.B.S.O. De; DERANI, C.; ALBUQUERQUE JUNIOR, C.L. De; RIBEIRO, J.M.P.; KNOBLOCH, F.; POLLITT, H.; EDWARDS, N.R.; HOLDEN, P.B.;

FOLEY, A.; SCHAPHOFF, S.; FARACO, R. A.; VINUALES, J. E. System complexity and policy integration challenges: The Brazilian Energy- Water-Food Nexus. **Renewable and Sustainable Energy Reviews**, v. 105, p. 230-243, 2019.

MERLO, T. P. **Aplicação de geotecnologias na gestão de títulos e garantias no crédito agrícola**. 2016. [s.n.]. Trabalho de Conclusão de Curso (Graduação em Engenharia Agrícola) Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola. Campinas, 2016.

MILLARD, K.; DARLING, S.; PELLETIER, N.; SCHULTZ, S. Seasonally decomposed Sentinel-1 backscatter time-series are useful indicators of peatland wildfire vulnerability. **Remote Sensing of Environment**, v. 283, A. 113329, 2022

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO - MAPA. **Zoneamento agrícola de risco climático** - ZARC. 2019. Disponível em: <<http://indicadores.agricultura.gov.br/zarc/index.htm>>. Acesso em: 20 fev. 2021.

MULLA, D. Twenty-five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. **Biosyst. Eng**, v.114, p. 358–371, 2013.

MÜLLER, H. RUFIN, P.; GRIFFITHS, P.; SIQUEIRA, A. J. B.; HOSTERT, P. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, p. 490-499, 2015.

NIAZMARDI, S.; HOMAYOUNI, S.; SAFARI, A.; MCNAIRN, H.; SHANG, J.; BECKETT, K. Histogram-based spatio-temporal feature classification of vegetation indices time-series for crop mapping. **International Journal of Applied Earth Observation and Geoinformation**, v. 72, p. 34-41, 2018.

OLDONI, L. V.; CATTANI, C. E. V.; MERCANTE, E.; JOHANN, J. A.; ANTUNES, J. F. G.; ALMEIDA, L. Annual cropland mapping using data mining and OLI Landsat-8. **R. Bras. Eng. Agríc. Ambiental**, Campina Grande - PB, v. 23, n.12, p. 952-958, 2019.

OLDONI, L. V.; SANCHES, I.D.; PICOLI, M.C.A.; COVRE, R.M.; FRONZA, J.G. LEM+ dataset: For agricultural remote sensing applications. **Data in Brief**, v. 33, p. 1-7, 2020.

OLIVEIRA, L. P.; PACHECO, M. G. Geographical analysis: relations of use and coverage to land morphology from moderate resolution imaging spectroradiometer (modis) and shuttle radar topography mission (srtm) for west baiano, Bahia, Brazil. **Diálogo Andino**, n. 63, p. 285-297, 2020.

OLIVEIRA, S. N.; CARVALHO JÚNIOR, O. A.; GOMES, R. A. T.; GUIMARÃES, R. F.; MCMANUS, C. M. Deforestation analysis in protected areas and scenario simulation for structural corridors in the agricultural frontier of Western Bahia, Brazil, **Land Use Policy**, v. 61, p. 40-52, 2017.

OUYANG, Z.; RAVIER, P.; JABLON, M. STL decomposition of time series can benefit forecasting done by statistical methods but not by machine learning ones. **Engineering Proceedings**, v. 5, p. 1-10, 2021.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B. Scikit-Learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

PERKTOLD, J.; SEABOLD, S.; TAYLOR, J. 2019. **Installing Statsmodels**. Disponível em: <<https://www.statsmodels.org/devel/install.html>>. Acesso em: 06 jun. 2021.

POTGIETER, A. B.; APAN, A.; HAMMER, G.; DUNN, P. Early-season crop area estimates for winter crops in NE Australia using MODIS satellite imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n. 4, p. 380–387, 1 jul. 2010.

ROUSE J. W.; HAAS, R. H.; SCHELL J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. In: Earth Resources Technology Satellite-1 Symposium, 3., Washington, 1973. **Proceedings...** Washington: NASA, 1974, v.1, p.309-317.

RUFIN, P.; MÜLLER, H.; PFLUGMACHER, D.; HOSTERT, D. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 41, p. 1–10, 1 set. 2015.

SAVITZKY, A.; GOLAY, M.J.E. Smoothing and differentiation of data by simplified least squares procedure. **Analytical Chemistry**, v. 36, p. 1627-1639, 1964.

BAHIA. Secretaria da Agricultura, Pecuária, Irrigação, Pesca e Aquicultura - SEAGRI. **Importância da Bahia cresce em estudo que traz o Brasil como um dos gigantes mundiais do agro.** Disponível em: <<http://www.seagri.ba.gov.br/noticias/2021/06/04/importancia-da-bahia-cresce-em-estudo-que-traz-o-brasil-como-um-dos-gigantes>>. Acesso em: 20, jul. 2021.

SISTEMA NACIONAL DE CADASTRO AMBIENTAL RURAL – SICAR. **O que é o SICAR?** 2022. Disponível em: <<https://www.car.gov.br/#/sobre>>. Acesso em: 22 de jul 2022.

SILVA, C. R.; MACHADO, S. L. D.; ARAÚJO, A. A.; ABREU JUNIOR, C. A. M. Analysis of the phenology dynamics of brazilian caatinga species with NDVI time series. **Cerne**, v. 24, n. 1, p. 48-58, 2018.

SIMÕES, R.; CAMARA, G.; QUEIROZ, G.; SOUZA, F.; ANDRADE, P. R.; SANTOS, L.; CARVALHO, A.; FERREIRA, K. Satellite image time series analysis for big earth observation data. **Remote Sens.**, v. 13, n. 2428, p.1-20, 2021.

SINGH, M.; TYAGI, K. D. *Pixel* based classification for Landsat 8 OLI multispectral satellite images using deep learning neural network. **Remote Sensing Applications: Society and Environment**, v. 24, n. October, p. 1-15, 2021.

SOARES, A. R.; BENDINI, H. N.; VAZ, D. V.; UEHARA, T. D.; NEVES, A. K.; LECHLER, S.; KÖRTING, T. S.; FONSECA, L. M. G. STMETRICS: **A Python package for satellite image time-series feature extraction.** Conference: 2020 IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2020. Waikoloa, HI, USA, 2020, p. 2061-2064, DOI: 10.1109/IGARSS39084.2020.9323346.

SOUZA, P. M.; NEY, M.G.; PONCIANO, N. J. Análise da distribuição dos financiamentos rurais entre os estabelecimentos agropecuários brasileiros. **Revista de Economia e Sociologia Rural**, Piracicaba – SP, v. 53, n.2, p. 251-270, abr/jun 2015. Disponível em: <<http://www.scielo.br/pdf/resr/v53n2/1806-9479-resr-53-02-00251.pdf>>. Acesso em: 16, out. 2019.

TORMENE, P.; GIORGINO, T.; QUAGLINI, S.; STEFANELLI, M. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. **Artificial Intelligence in Medicine**, v. 45, p. 11-34, 2008.

VAROQUAUX, G. Joblib: running Python functions as *pipeline* jobs. **Readthedocs**, 2021. Disponível em: < <https://joblib.readthedocs.io/en/latest/> > Acesso em: 11, set. 2021.

VIANA, C. M.; GIRÃO, I.; ROCHA, J. Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. **Remote Sensing**, v. 11, p.1-22, 2019.

WANG, Y.; ZANG, S.; TIAN, Y. Mapping paddy rice with the random forest algorithm using MODIS and SMAP time series. **Chaos, Solitons & Fractals**, v. 140, p 1-11, 2020.

WATTS, L.M; LAFFAN S.W. Effectiveness of the BFAST algorithm for detecting vegetation response patterns in a semi-arid region. **Remote Sensing of Environment**, v. 154, p. 234-245, 2014.

WEISS, M.; JACOB, F.; DUVEILLER, G. Remote sensing for agricultural applications: A meta-review. **Remote Sensing of Environment**, v. 236, A. 111402, jan. 2020.

WU, Z.; LIU, Y.; HAN, Y.; ZHOU, J.; LIU, J.; WU, J. Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data. **Science of the Total Environment**, v. 754, p. 1-12, 2021.

XUN, L.; ZHANG, J.; CAR, D.; WANG, J.; ZHANG, S.; YAO, F. Mapping cotton cultivated area combining remote sensing with a fused representation-based classification algorithm. **Computers and Electronics in Agriculture**, v. 181, p. 1-13, 2021.

ZHANG Z.; TANG, P.; ZHANG, W.; TANG, L. Satellite image time series clustering via time adaptive optimal transport. **Remote Sensing**, v.13, p. 1-23., 2021.

ZHU, J.; WANG, X.; CHEN, M.; WU, P.; KIM, M. J. Integration of BIM and GIS: IFC geometry transformation to shapefile using enhanced open-source approach, **Automation in Construction**, v.106, A. 102859, 2019.