



EDUARDA THAIS GOULART

**MODELAGEM PARA PROJEÇÃO DE PRODUTIVIDADE EM
LAVOURAS CAFEEIRAS UTILIZANDO *RANDOM FOREST***

**LAVRAS - MG
2025**

EDUARDA THAIS GOULART

**MODELAGEM PARA PROJEÇÃO DE PRODUTIVIDADE EM LAVOURAS
CAFEIRAS UTILIZANDO *RANDOM FOREST***

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, para a obtenção do título de Mestre.

Prof. Dr. Fausto Weimar Acerbi Junior
Orientador
Prof. Dr. Allan Pereira Arantes
Coorientador

**LAVRAS - MG
2025**

Ficha catalográfica elaborada pela Catalogação da Biblioteca Universitária da UFLA

Goulart, Eduarda Tais.

Modelagem para projeção de produtividade em lavouras
cafeeiras utilizando *Randon Forest* / Eduarda Tais Goulart. - 2025.
84 p. : il.

Orientador(a): Fausto Weimar Acerbi Junior.
Coorientador(a): Allan Pereira Arantes.

Dissertação (mestrado) - Universidade Federal de Lavras, 2025.
Bibliografia.

1. Produção. 2. Coffea arábica. 3. Série temporal. I. Acerbi
Junior, Fausto Weimar. II. Arantes, Allan Pereira. III. Título.

Ficha elaborada por André Felipe Calsavara (CRB6/3137)

EDUARDA THAIS GOULART

**MODELAGEM PARA PROJEÇÃO DE PRODUTIVIDADE EM LAVOURAS
CAFEIIRAS UTILIZANDO *RANDOM FOREST*
MODELING FOR PROJECTING PRODUCTIVITY IN COFFEE FARMS USING
RANDOM FOREST**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, para a obtenção do título de Mestre.

APROVADA em 30 de janeiro de 2025.

Dra. Mônica Canaan Carvalho

Dr. Inacio Thomaz Bueno

Dr. Kalill Jose Viana Da Pascoa

Dra. Michele Duarte De Menezes

University of Florida

UFLA

UFLA

Prof. Dr. Fausto Weimar Acerbi Junior

Orientador

Prof. Dr. Allan Pereira Arantes

Co-orientador

LAVRAS – MG
2025

*Aos meus pais, Maria e Ivan Goulart, que, com muito amor,
trabalharam arduamente para que eu chegasse até este momento.*

Meu maior exemplo de vida.

Dedico.

AGRADECIMENTOS

Ao concluir este trabalho, sinto-me profundamente grata por todos aqueles que, de diferentes maneiras, contribuíram para esta jornada de aprendizado e crescimento.

Agradeço primeiramente a Deus, fonte de sabedoria, força e inspiração, por guiar meus passos e iluminar o caminho até aqui.

À minha família, em especial ao meu pai, minha mãe e minha irmã, que sempre me deram apoio incondicional, amor e encorajamento, sendo pilares fundamentais em todos os momentos.

Aos amigos e companheiros de curso, cuja convivência, troca de ideias e incentivo foram essenciais para a conclusão desta pesquisa.

Ao meu orientador, Prof. Dr. Fausto Weimar Acerbi Junior, e ao coorientador, Prof. Dr. Allan Arantes Pereira, por suas orientações precisas, paciência, dedicação e confiança, que enriqueceram imensamente este trabalho e a minha formação.

À Fazenda Sertãozinho, cuja parceria e generosidade ao fornecer os dados necessários foram de extrema importância para o desenvolvimento desta pesquisa.

À Universidade Federal de Lavras, por proporcionar o ambiente acadêmico e os recursos necessários para a realização do curso e da pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Agradecemos à Capes por conceder a bolsa de estudos que viabilizou este projeto e o fortalecimento da ciência agrária no Brasil.

A todos vocês, minha sincera gratidão por serem parte dessa trajetória.

*Portanto, quer comais quer bebais, ou façais
outra qualquer coisa, fazei tudo para glória de Deus.*

1 Coríntios 10. 31.

RESUMO GERAL

O objetivo desta pesquisa foi criar um modelo para projeção de produtividade de lavouras cafeeiras. O estudo foi desenvolvido nas fazendas Sertãozinho e Laranjal, nos municípios de Botelhos e Poços de Caldas, em Minas Gerais. Os valores de produtividade dos talhões de café para os anos de 2017 a 2024 foram fornecidos pelas fazendas. As variáveis espectrais utilizadas no modelo foram as medidas de máxima, mínima, mediana, média, desvio padrão e moda dos índices de vegetação NDVI, NDWI, EVI e SAVI extraídas de imagens Sentinel-2, no período de agosto dos anos de 2016 a 2023. As características de terreno foram obtidas de um Modelo Digital de Elevação (MDE) Alos Palsar utilizando o Software Saga Gis e foram importadas para a plataforma Google Earth Engine (GEE), onde foram calculadas as médias destas variáveis para cada um dos talhões. As variáveis de clima foram adquiridas da coleção Eras-5 Land, disponível para acesso no GEE. Foram considerados os dados médios mensais de precipitação, radiação solar e volume de água no solo, além de dados de mínima, máxima e média mensais de temperatura, dos meses de agosto de 2016 a julho de 2024. Os dados de clima, de terreno e espectrais foram exportados do GEE em formato tabelar e unidos aos dados de produtividade. Para a modelagem, foi realizada uma análise de correlação de Pearson entre as variáveis independentes e dependente e gerou-se uma matriz de correlação. Para compor o banco de dados de entrada do modelo, foi calculada a diferença percentual da produtividade de um ano a outro. O desenvolvimento do modelo *Random Forest* contou com a técnica de validação cruzada para treinar o algoritmo. Para a execução do *Random Forest*, adotou-se como 100 o número de árvores de decisão a serem construídas. O modelo gerou estimativas para os anos de 2018 a 2023, cujos os valores reais foram informados para treinamento do algoritmo, e projetou a produtividade para o ano de 2024. Foi avaliado o desempenho do algoritmo através das métricas de R^2 , MAPE e RMSE. Dentre as variáveis de entrada do modelo, aquelas que obtiveram maiores correlações com a produtividade de café foram os IVs EVI, NDWI e NDVI, a radiação solar, temperatura mínima e precipitação. As variáveis de maior importância para o modelo foram a diferença percentual da produtividade entre um ano e outro, a temperatura mínima dos meses de março e novembro, a temperatura média do mês de outubro, a radiação solar média dos meses de agosto, outubro, fevereiro e junho, o índice Convergence index (CI) e as medidas de moda, mediana e mínima do índice NDWI do mês de agosto. O modelo obteve melhor desempenho quando foram consideradas apenas as variáveis de importância acima de 200, apresentando R^2 de 93,74%, MAE de 2,49 sc.ha⁻¹ e RMSE de 10,99 sc.ha⁻¹. Uma das principais limitações do modelo foi a pequena base de dados de produtividades disponíveis para treinamento e teste do modelo de projeção de safras criado, que prejudicou a previsão de valores de produtividade extremos.

Palavras-Chave: Produção; *Coffea arabica*; Google Earth Engine; Aprendizagem de máquina.

ABSTRACT

The objective of this research was to develop a model for projecting the productivity of coffee plantations. The study was conducted on the Sertãozinho and Laranjal farms, located in the municipalities of Botelhos and Poços de Caldas, Minas Gerais, Brazil. The productivity values of coffee plots from 2017 to 2024 were provided by the farms. The spectral variables used in the model were the maximum, minimum, median, mean, standard deviation, and mode of the vegetation indices NDVI, NDWI, EVI, and SAVI, extracted from Sentinel-2 images during August from 2016 to 2023. Terrain characteristics were obtained from an Alos Palsar Digital Elevation Model (DEM) using Saga GIS software and imported into the Google Earth Engine (GEE) platform, where the averages of these variables were calculated for each plot. Climate variables were acquired from the Era-5 Land collection, available for access on GEE. Monthly averages of precipitation, solar radiation, and soil water volume were considered, along with minimum, maximum, and mean monthly temperatures, from August 2016 to July 2024. Climate, terrain, and spectral data were exported from GEE in tabular format and merged with the productivity data. For modeling, Pearson correlation analysis was performed between the independent and dependent variables, generating a correlation matrix. To compose the model's input database, the percentage difference in productivity from one year to the next was calculated. The Random Forest model was developed using the cross-validation technique to train the algorithm. For *Random Forest* execution, the number of decision trees to be constructed was set to 100. The model generated estimates for the years 2018 to 2023, for which actual values were provided to train the algorithm, and projected productivity for the year 2024. The algorithm's performance was evaluated using R^2 , MAPE, and RMSE metrics. Among the model's input variables, those with the highest correlations with coffee productivity were the vegetation indices EVI, NDWI, and NDVI, as well as solar radiation, minimum temperature, and precipitation. The most important variables for the model were the percentage difference in productivity from one year to the next, the minimum temperature in March and November, the mean temperature in October, the average solar radiation in August, October, February, and June, the Convergence Index (CI), and the mode, median, and minimum values of the NDWI index in August. The model performed best when only variables with an importance score above 200 were considered, achieving an R^2 of 93.74%, MAE of 2.49 sc.ha⁻¹, and RMSE of 10.99 sc.ha⁻¹. One of the main limitations of the model was the small database of productivity values available for training and testing the projection model, which hampers the prediction of extreme productivity values.

Keywords: Production; *Coffea arabica*; Google Earth Engine; Machine Learning.

INDICADORES DE IMPACTO

Esta dissertação tem como objetivo estudar a integração entre o sensoriamento remoto e o algoritmo *Random Forest* no desenvolvimento de um modelo de estimativa de produtividade para lavouras cafeeiras do sul de Minas Gerais, avaliando o desempenho do modelo e sua precisão frente à produção real de café. O uso do sensoriamento remoto combinado a técnicas de *machine learning* mostra-se uma abordagem promissora no contexto agrícola, permitindo o monitoramento da produção de diversas commodities em grandes áreas com maior precisão, contribuindo para a otimização do manejo e a redução de incertezas no planejamento da safra. O modelo desenvolvido apresentou resultados satisfatórios, com R^2 de 93,74%, MAE de 2,49% e RMSE de 10,99%. As maiores discrepâncias identificadas entre as estimativas dos anos de 2018 a 2023 foram de 8,50 sc.ha⁻¹ a mais e 6,85 sc.ha⁻¹ a menos que a produtividade real colhida pela fazenda, demonstrando que o modelo é capaz de projetar a produção com bom grau de precisão. Embora em 2024 tenham sido observadas diferenças mais elevadas em alguns talhões com produtividades excepcionalmente altas, o modelo ainda mostrou capacidade de captar o caráter genético de bienalidade da cultura do café. Logo, o modelo desenvolvido contribui para o aprimoramento da estimativa de produtividade cafeeira, ferramenta estratégica para a gestão agrícola, planejamento econômico, políticas públicas e fortalecimento do setor exportador, e pode ser de interesse de instituições responsáveis pelo monitoramento da cafeicultura nos municípios brasileiros, como a Embrapa, EPAMIG, CONAB e Prefeituras. O público diretamente beneficiado inclui produtores rurais da região, técnicos agrícolas e gestores públicos que, com acesso a projeções mais precisas de produtividade, podem tomar decisões mais assertivas sobre manejo, colheita e comercialização. Os impactos do trabalho podem ser classificados na área temática de tecnologia e produção (7), contribuindo também para o alcance de pelo menos três Objetivos de Desenvolvimento Sustentável (ODS) da ONU: ODS 2 (Fome Zero e Agricultura Sustentável), ODS 8 (Trabalho Decente e Crescimento Econômico), e ODS 12 (Consumo e Produção Responsáveis, alinhando se às metas da Agenda 2030 ao promover inovação no campo e aumento da resiliência das cadeias produtivas agrícolas

IMPACT INDICATORS

This dissertation aims to study the integration of remote sensing and the Random Forest algorithm in the development of a yield estimation model for coffee plantations in southern Minas Gerais, evaluating the model's performance and accuracy compared to the actual coffee production. The use of remote sensing combined with machine learning techniques proves to be a promising approach in the agricultural context, allowing for the monitoring of crop production across large areas with greater precision. This contributes to optimized crop management and reduced uncertainty in harvest planning. The model developed showed satisfactory results, with an R^2 of 93.74%, MAE of 2.49%, and RMSE of 10.99%. The largest discrepancies identified between the estimates from 2018 to 2023 ranged from 8.50 sc.ha⁻¹ above to 6.85 sc.ha⁻¹ below the actual yield harvested by the farm, demonstrating the model's ability to project production with a good degree of accuracy. Although higher differences were observed in some plots in 2024, which had exceptionally high yields, the model still showed the capacity to capture the genetic characteristic of biennial bearing in coffee crops. Therefore, the model contributes to improving coffee yield estimation—an essential tool for agricultural management, economic planning, public policy development, and strengthening the export sector. It may be of interest to institutions responsible for monitoring coffee production in Brazilian municipalities, such as Embrapa, EPAMIG, CONAB, and local governments. The directly benefited audience includes local farmers, agricultural technicians, and public managers, who, with access to more accurate yield projections, can make more informed decisions regarding crop management, harvesting, and marketing. The impacts of this work can be classified within the thematic area of technology and production (7), also contributing to the achievement of at least three United Nations Sustainable Development Goals (SDGs): SDG 2 (Zero Hunger and Sustainable Agriculture), SDG 8 (Decent Work and Economic Growth), and SDG 12 (Responsible Consumption and Production), aligning with the goals of the 2030 Agenda by promoting innovation in agriculture and increasing the resilience of agricultural value chains.

LITA DE ILUSTRAÇÕES

Figura 1 - Representação das estruturas da espécie <i>Coffea arabica</i> L. mostrando a disposição de ramos, folhas, gemas e frutos.....	21
Figura 2 - Esquematização das seis fases fenológicas do cafeeiro (<i>Coffea arabica</i> L.) durante dois anos do ciclo.....	22
Figura 3 - Espectro eletromagnético.....	28
Figura 4 – Comportamento Espectral das Cultivares de <i>Coffea arabica</i> L. Acaiá Cerrado - MG 1474, Mundo Novo IAC 379 – 19, Yellow Bourbon IAC J10, Catuaí Vermelho IAC 99, Topazio MG 1190, Acauã New, IAC 125 RN e Paradise MG H 419-1.....	32
Figura 5 – Reflectância das bandas RED e NIR durante as 6 fases fenológicas do café.....	33
Figura 6 - Comportamento do índice NDVI durante as 6 fases fenológicas do café.....	33
Figura 7 - Configuração da órbita dos satélites Sentinel-2.....	36
Figura 8 - Representação esquemática do <i>Random Forest</i>	40
Figura 9 – Fluxograma de atividades.....	42
Figura 10 – Localização das lavouras de café alvo da pesquisa.....	43
Figura 11 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis espectrais NDVI, SAVI, EVI e NDWI.....	51
Figura 12 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis de terreno.....	53
Figura 13 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis climáticas.....	54
Figura 14 – Ranking das primeiras 30 variáveis mais importantes para o modelo.....	55
Figura 15 - Ranking das variáveis de maior importância quando consideradas apenas aquelas com valores acima de 7%.....	56
Figura 16 – Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da temperatura média de outubro, da temperatura mínima de novembro e da temperatura mínima de março durante os anos de estudo.....	58

Figura 17 – Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da radiação solar dos meses de agosto, outubro, fevereiro e junho.....	60
Figura 18 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento dos valores de mínimo, moda e mediana do NDWI.....	61
Figura 19 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento das variáveis precipitação do mês de outubro e precipitação do mês de fevereiro.....	63
Figura 20 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da variável volume de água disponível no solo.....	64
Figura 21 – Relação entre a produtividade real e a produtividade estimada.....	66
Figura 22 – Resíduos do modelo <i>Random Forest</i> para projeção de safra.....	67
Figura 23 – Projeção de produtividade ao longo dos anos dos talhões Figueira 2018, Gordurinha 2018 e Jequitibá 2018.....	70

LISTA DE TABELAS

Tabela 1 – Bandas das imagens multiespectrais dos satélites Sentinel 2A e Sentinel 2B.....	37
Tabela 2 - Talhões de estudo e suas respectivas cultivares, área (ha), quantidade de plantas por hectare, número de plantas total e espaçamento entre a linha de plantio e entre plantas.....	44
Tabela 3 - Data de plantio e produtividade (sc/ha) anual de cada talhão de estudo.....	45
Tabela 4 - Variáveis do terreno derivadas do MDE através do software Saga Gis.....	47
Tabela 5 - Variáveis meteorológicas da pesquisa, as bandas da coleção ERA5-Land utilizadas para obtenção destas variáveis, as medidas estatísticas calculadas e o período de obtenção destes dados.....	48
Tabela 6 – Medidas de desempenho MAE, R ² e RMSE para os modelos gerados.....	64
Tabela 7 – Produtividade real e prevista pelo modelo (em Sc ha ⁻¹) para os talhões e os anos de estudo e a diferença entre ambas.....	68
Tabela 8 – Produtividades reais e projetadas para o ano de 2024 por talhão, em Sc ha ⁻¹	69

SUMÁRIO

1.	INTRODUÇÃO	17
2.	REFERENCIAL TEÓRICO	19
2.1.	Origem, Botânica e Fases Fenológicas do Cafeeiro	20
2.2.	Fatores que Interferem na Produção do Cafeeiro	23
2.3.	Monitoramento da Produção Agrícola	25
2.4.	Sensoriamento Remoto e Índices de Vegetação	27
2.4.1.	Assinatura espectral da vegetação e Comportamento espectral do cafeeiro.	31
2.5.	Imagens Sentinel – 2	34
2.6.	<i>Machine Learning</i>	37
2.6.1.	<i>Random Forest</i>	39
3.	MATERIAL E MÉTODOS	41
3.1.	Caracterização da área de estudo	43
3.2.	Talhões	44
3.3.	Organização do Banco de Dados	45
3.3.1.	Variáveis espectrais	45
3.3.2.	Variáveis do terreno	46
3.3.3.	Variáveis climáticas	47
3.4.	Modelo <i>Random Forest</i> para Projeção de Safra	48
4.	RESULTADOS E DISCUSSÕES	50
4.1.	Correlação Entre as Variáveis Independentes e a Variável Dependente	50
4.2.	Importância das Variáveis	54
4.2.1.	Comparação entre o comportamento das variáveis de maior importância e o comportamento da produtividade nos anos de produção	57
4.2.1.1.	Temperatura do ar	57
4.2.1.2.	Radiação solar na superfície terrestre	59
4.2.1.3.	NDWI – Presença de água no dossel	61

4.2.1.4. Volume de água no Solo e Precipitação	62
4.3. Análise de desempenho do modelo preditivo	64
4.4. Projeção de Safra dos anos de 2018 a 2024	67
4.5. Limitações do Modelo.....	71
REFERÊNCIAS	72

1. INTRODUÇÃO

Ao longo das décadas, a cultura do café se consolidou como uma das principais *commodities* agrícolas comercializadas no mercado nacional e mundial. A produção de grãos na safra 2023/24 foi de 168 milhões de sacas de 60 quilos, que devem contribuir para suprir o consumo mundial de 170,63 milhões de sacas de café até a próxima colheita. O Brasil se consolidou como maior produtor de café mundial, seguido por países como Vietnã e Colômbia. Sua participação na próxima safra é estimada em 69,9 milhões de sacas de 60 kg, das quais quase 46,65 milhões serão destinadas a países importadores, sendo os principais os países da União Europeia, os Estados Unidos e a China (BRAINER, 2019; USDA, 2023).

Minas Gerais tem forte impacto na quantidade final de café produzido no Brasil. O estado alcançou produtividade de 25,2 sc.ha⁻¹ na última safra (2023/24), que corresponde a uma produção de 28 milhões de sacas de café arábica e robusta beneficiadas, valores relativamente superiores que qualquer outro estado brasileiro. A região Sul e Centro-Oeste de Minas Gerais corresponde à maior produção do café do país, uma vez que atingiu nessa mesma safra aproximadamente 13,6 milhões de sacas de café, representando cerca de 24,84% da produção brasileira (CONAB, 2024).

Para a alta produção de grãos, o cafeeiro necessita de uma forte base genética e exige condições ótimas ambientais e de cultivo para minimizar os efeitos dos diversos fatores bióticos e abióticos que influenciam direta ou indiretamente a produtividade da lavoura (FILHO, 2018). Segundo Soares *et al.* (2020), o clima da região exerce grande impacto nas fases de desenvolvimento da planta de café e afeta sua produção. Aspectos meteorológicos como temperatura, precipitação, radiação solar, fotoperíodo e umidade relativa do ar são responsáveis, em grande parte, pelo vingamento das flores, crescimento e amadurecimento dos frutos e boa formação das sementes (BRASIL, 2021; FILHO, 2018). Outros fatores, como sistema de cultivo (a pleno sol ou sombreado), altitude da área, espaçamento de plantio adotado, presença de irrigação, compactação do solo, disponibilidade de água, macro e micronutrientes no solo, capacidade produtiva da cultivar implantada e idade da lavoura também apresentam significativos efeitos sobre o crescimento e desenvolvimento das plantas e produção da lavoura cafeeira (FREITAS *et al.*, 2020; MACHADO *et al.*, 2020; JARA *et al.*, 2019; SILVA *et al.*, 2020; VIANA *et al.*, 2019).

Devido à ampla quantidade de fatores que interferem na produção da planta, o monitoramento e a estimativa de safras de café são dificultados. A estimativa de produtividade

de qualquer cultura agrícola exerce influência no planejamento do setor econômico, na estabilidade de preços e transações do mercado de importação e exportação, são base para criação e execução de políticas públicas pelo Estado, é fundamental para a gestão integrada das propriedades rurais e tem grande importância na tomada de decisões de produtores rurais e cooperativas agrícolas (BOLFE *et al.*, 2023; BATISTELLA *et al.*, 2023).

Em busca de previsões eficientes para a produção cafeeira, diferentes metodologias foram desenvolvidas. Dentre elas, o uso de modelagens estatísticas tem se tornado popular devido à maior aplicabilidade em grandes áreas, à capacidade de manipulação de vários dados e à rapidez e confiabilidade na obtenção de resultados (AMARO, 2023; MIQUELLUTI, 2015). Esses modelos utilizam diferentes variáveis que estão intimamente associadas à queda ou ao aumento da produção agrícola e simulam as relações entre elas e delas com o fator produção a fim de gerar uma previsão adequada de colheita de grãos na próxima safra.

Para melhores estimativas de produção, os modelos de previsão de safra podem ser construídos utilizando dados históricos de produção da lavoura. O Ministério da Agricultura, Pecuária e Abastecimento (MAPA), por exemplo, realizou, em 2022, uma série de estimativas de produção para diferentes culturas agrícolas, como grãos, frutas, produtos derivados da silvicultura e produtos derivados da pecuária, para os anos agrícolas de 2021/22 a 2031/32, utilizando dados de produção desde o ano de 1994 (MAPA, 2022). Nestes casos, quando usa-se informações de produtividade de anos anteriores para estimativas de produtividade dos anos seguintes, o termo geralmente empregado é “projeção de produtividade”.

Modelos de previsão ou projeção de safras atuais aplicados a culturas agrícolas como feijão, soja, milho, trigo, algodão e cana-de-açúcar utilizam dados de sensoriamento remoto para relacionar o estado fisiológico da cultura com sua produção futura (ANDRADE *et al.*, 2022; CLARINDO *et al.*, 2022; COSTA; RABELO, 2023; MORAIS; BREDEMEIER, 2019; SANDRINI *et al.*, 2019; SILVA *et al.*, 2019). No sensoriamento remoto, a intensidade energética da radiação eletromagnética emitida pelo sol é medida por sensores acoplados a satélites que orbitam a Terra. Esses sensores captam a radiação que chega ao objeto na superfície terrestre e é emitida ou refletida por ele. As imagens geradas permitem a associação da radiação solar refletida com o estado de biomassa de lavouras agrícolas e possibilitam o estudo do estado da lavoura em relação à próxima safra (MENESES; ALMEIDA; BAPTISTA, 2019).

Para o desenvolvimento e construção de modelos de projeção agrícola, o algoritmo de aprendizagem de máquina *Random Forest* tem grande aplicabilidade. Desenvolvido por

Breiman em 2001, o *Random Forest* é um método de aprendizado de conjunto composto por múltiplas árvores de decisão utilizado para classificação ou predição de amostras. As árvores de decisão são modelos de predição construídos a partir de variáveis/amostras de entrada ao algoritmo, e possui uma estrutura hierárquica composta por subconjuntos de características das amostras, também chamados de nós raiz, nós intermediários e nós folha. Deste modo, as árvores avaliam um ou mais atributos da amostra até alcançar um nó folha e determinar o resultado final. Na classificação final, é utilizada uma votação, onde a classe mais votada é a escolhida. Para problemas de regressão, geralmente é usada a média (BIAU; SCORNET, 2016; YIN et al., 2021).

No estudo de Pereira, Beker e Schmoeller (2024), o algoritmo *Random Forest* apresentou alto desempenho em realizar estimativas de produtividade da soja em lavouras ao oeste do Paraná. Foram usadas como variáveis de entrada para o modelo dados meteorológicos, como evapotranspiração da cultura, temperatura média, radiação solar, umidade relativa do ar e velocidade do vento. O *Radom Forest* atingiu um R^2 de 86% e erro quadrático médio (RMSE) de 309,33 k/ha para o conjunto de treino do modelo, e R^2 de 81% e RMSE de 410,18 kg/ha para o conjunto de teste, sendo superior aos demais algoritmos testados. Os autores afirmam que modelos de *machine learning* podem ser uma ferramenta útil para estimar a produtividade de culturas agrícolas e que as estimativas de produtividade geradas por esses modelos podem auxiliar os agricultores a tomar decisões mais acertadas sobre o manejo de suas lavouras, contribuindo para aumentar a eficiência e a sustentabilidade da produção agrícola.

Tendo em vista os desafios que dificultam a estimativa de produção de grãos para a cultura do café e o uso crescente de tecnologias como o sensoriamento remoto e *machine learning* para previsões de colheita de demais culturas agrícolas, o objetivo desta pesquisa é estudar a integração entre o sensoriamento remoto e o algoritmo *Random Forest* no desenvolvimento de um modelo de estimativa de produtividade para lavouras cafeeiras do sul de Minas Gerais, avaliando o desempenho do modelo e sua precisão frente à produção real de café. Além disso, o estudo busca analisar a correlação entre as variáveis de entrada e a produtividade, bem como identificar as variáveis de maior importância para a construção e ramificação das árvores de decisão do algoritmo.

2. REFERENCIAL TEÓRICO

2.1. Origem, Botânica e Fases Fenológicas do Cafeeiro

O café, conhecido na França, em Portugal, na Espanha e no Brasil como *café*; na Inglaterra como "*coffée*"; na Holanda como "*koffy*"; na Alemanha como "*kaffe*"; e no Egito como "*ellkarié*", foi originalmente encontrado nas vegetações de sub-bosque das florestas tropicais do sul da Etiópia, no continente africano (ANTHONY, 1992; LUNZ; BERNARDES; RIGHI, 2009; MOREIRA, 1873). A planta pertence ao Reino *Plantae*, Divisão *Magnoliophyta*, Classe *Magnoliopsida*, Ordem *Gentianales*, Família *Rubiaceae* e Gênero *Coffea*. A família *Rubiaceae* abrange mais de 10 mil espécies agrupadas em 630 gêneros. Dentre as espécies do gênero *Coffea*, apenas a *Coffea arábica* e a *Coffea canephora* são exploradas comercialmente (FERRÃO *et al.*, 2017; MELO; SOUSA, 2011).

Embora a base genética da espécie *Coffea arábica* L. seja bastante estreita, uma série de fatores, como mutações, cruzamentos naturais, adaptações ambientais e intervenções do melhoramento genético, levou a modificações morfológicas na estrutura anatômica do cafeeiro, resultando na ampla gama de cultivares de diferentes portes e características fisionômicas encontradas atualmente (CARVALHO, 2007).

De modo geral, o cafeeiro é um arbusto perene que pode atingir de dois a seis metros de altura. Seu sistema radicular é constituído por uma raiz principal pivotante, raízes laterais e axiais, e raízes laterais de superfície e de profundidade. Esse conjunto pode atingir até 2 metros de profundidade no solo, sendo que cerca de 40% a 50% das raízes estão nos primeiros 60 cm de solo, e grande parte se concentra na projeção da saia do cafeeiro (CARVALHO, 2007).

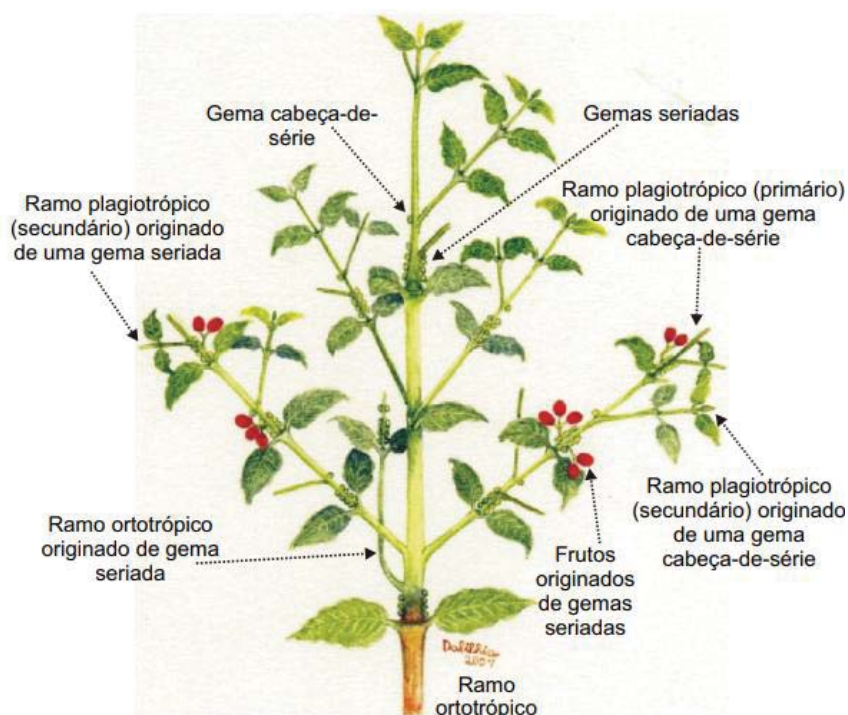
A parte aérea da planta é formada por um ramo vertical, denominado ortotrópico, e ramos laterais, conhecidos como plagiotrópicos. Em ambas as estruturas, encontram-se uma série de gemas localizadas nas axilas foliares, onde a primeira gema é chamada de “cabeça-de-série”, e as demais são conhecidas como gemas seriadas. Em ramos ortotrópicos, as gemas são responsáveis pelo desenvolvimento de estruturas vegetativas, como ramos e folhas. Já as gemas de ramos plagiotrópicos possuem comportamento diferente, em que as gemas cabeças-de-séries desenvolvem ramos secundários e as gemas seriadas podem gerar ramos, folhas, botões florais e, posteriormente, frutos (WINTGENS, 2004).

As folhas do cafeeiro são brilhantes, com coloração verde-escuro na face adaxial e verde mais claro na face abaxial. São elípticas, com comprimento médio de 5 cm, de 4 a 14 pares de nervuras e grande quantidade de estômatos em sua parte inferior. São dispostas em filotaxia oposta cruzada quando a planta é jovem e oposta quando adulta (VIEIRA, 2017).

As inflorescências da planta se desenvolvem na axila foliar e dão origem a até quatro flores, em uma estrutura denominada glomérulo. As flores possuem coloração branca e apresentam anteras e estigmas proeminentes, e estilo longo. É importante ressaltar que a espécie *Coffea arábica* é cleistógama, ou seja, as flores são autofecundadas antes de sua abertura na floração (CARVALHO, 2007).

O fruto do cafeeiro possui uma estrutura conhecida como pericarpo, formada por três camadas de células: o epicarpo, o mesocarpo e o endocarpo. O pericarpo tem uma parte externa delgada (exocarpo ou epicarpo), que é a casca propriamente dita. O mesocarpo é carnoso e rico em mucilagem (pectinas e açúcares), ocorrendo mais em frutos de café arábica. O endocarpo, ou pergaminho, apresenta-se fibroso e lignificado no final da sua formação, em particular na fase de expansão, durante o crescimento do fruto (MELO, SOUSA, 2011). A figura 1 apresenta a disposição dos ramos, folhas, gemas e frutos do cafeeiro.

Figura 1 - Representação das estruturas da espécie *Coffea arábica* L. mostrando a disposição de ramos, folhas, gemas e frutos.



Fonte: Carvalho (2007).

As sementes, produto final da cultura do café, são esverdeadas, cobertas por uma película prateada aderente e constituídas quimicamente por carboidratos, cafeína, proteínas, óleos, ácido clorogênico e outras moléculas. Sementes proeminentes de frutos que desenvolveram dois óvulos são denominadas grãos chatos, enquanto frutos que desenvolveram

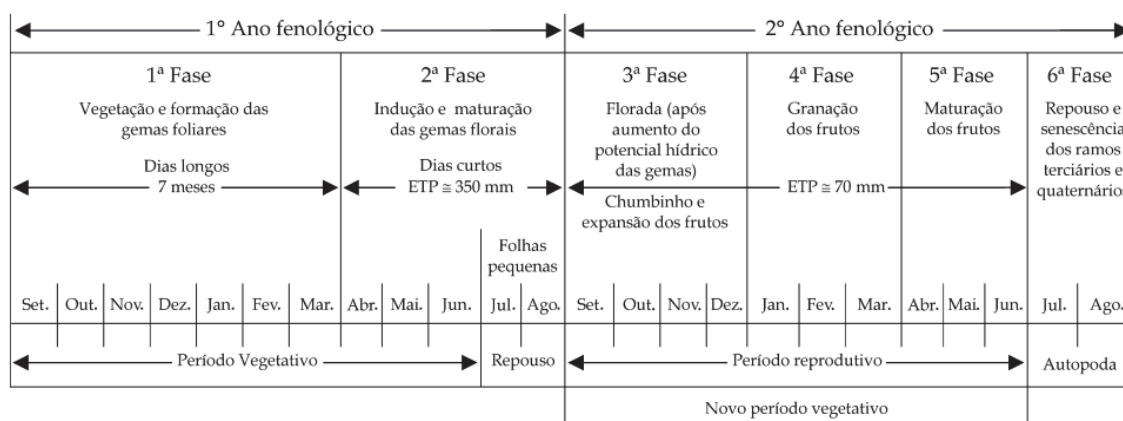
um único ovário, devido ao abortamento do segundo, geram sementes arredondadas, conhecidas como grão moka (VIEIRA, 2017).

O cafeeiro exige um período de aproximadamente 3 anos para o crescimento e desenvolvimento de todas as suas estruturas vegetativas e reprodutivas. Apenas após esse período o cafeeiro é capaz de produzir quantidades relevantes de frutos. Quando bem manejada, essa planta pode chegar a produzir por até 80 anos, porém, para não ocorrer perdas de vigor e diminuição da produção, as lavouras atuais são renovadas em até 30 anos (WINTGENS, 2004).

Após os primeiros anos de crescimento da planta, o ciclo de produção de grãos do cafeeiro leva dois anos para se completar, passando por seis fases distintas. A primeira fase acontece durante os meses de setembro a março e é responsável pela formação de gemas foliares e pelo desenvolvimento dos estágios de vegetação. A segunda fase se inicia em abril, com dias curtos (menos de 13 horas de luz efetiva), que intensifica a indução das gemas foliares existentes para gemas florais. Essas gemas florais vão amadurecendo e, quando maduras, entram em dormência e ficam prontas para a antese. A dormência só é quebrada com o aumento do potencial hídrico nas gemas dormentes, causado principalmente por irrigação, precipitação ou elevação da umidade do ar, desencadeando a florada (CAMARGO; CAMARGO, 2001).

O segundo ano fenológico e a 3ª parte do ciclo do cafeeiro iniciam-se com a florada e a formação dos chumbinhos, que precedem a expansão dos grãos até atingir o tamanho normal. Em seguida, ocorre a granação dos frutos e a fase de maturação. Finalmente, advém a senescência, morte dos ramos plagiotrópicos terminais, e a conhecida auto poda. Na primavera do ano civil seguinte, brotam novos ramos vegetativos que se transformam em reprodutivos, permitindo nova produção (CAMARGO; CAMARGO, 2001; MEIRELES *et al.*, 2009).

Figura 2 - Esquemática das seis fases fenológicas do cafeeiro (*Coffea arabica* L.) durante dois anos do ciclo.



Fonte: Camargo; Camargo (2001).

2.2. Fatores que Interferem na Produção do Cafeeiro

Existem diferentes fatores que influenciam a produtividade das lavouras cafeeiras. O clima e condições meteorológicas da região estão intrinsecamente ligados à produção de grãos e à qualidade da bebida. Entre os elementos climáticos, a temperatura do ar e a precipitação se destacam como os de maior impacto sobre o rendimento da colheita no final da safra (RIBEIRO; PEDROSA; FERREIRA, 2022). Segundo o último Zoneamento Agrícola de Risco Climático para a cultura do café arábica, de modo geral, são consideradas condições ótimas para o cafeeiro a precipitação anual entre 1.200 a 1.800 mm e a temperatura média anual na faixa de 18 a 21°C. Temperaturas superiores nas fases iniciais do período reprodutivo levam ao não vingamento da florada e ao mau desenvolvimento de frutos, enquanto baixas temperaturas levam a ocorrência de geadas, ocasionando danos à cultura e até à morte das plantas (BRASIL, 2021).

Características do terreno também estão relacionadas a fatores climáticos e de condução da lavoura e, portanto, à produção do cafezal. Segundo Ribeiro, Pedrosa e Ferreira (2022), a altitude e as faces de exposição da lavoura em relação à incidência da radiação solar ao longo do dia e do ano podem interferir no desenvolvimento das estruturas vegetativas e reprodutivas da planta. Kamal *et al.* (2021) estudaram os efeitos de diferentes altitudes na qualidade e quantidade de cafés produzidos. Para tanto, o estudo foi desenvolvido em lavouras cafeeiras com altitudes entre 800 a 1.600 m. Como conclusão, os autores apontam que o aumento da altitude refletiu diretamente na quantidade de grãos produzidos e na diminuição do teor de cafeína nos grãos, uma vez que temperaturas ambientes mais baixas durante o período de maturação do café proporcionaram maiores teores de sacarose e acúmulo de carboidratos nos grãos, tornando-os mais pesados e densos.

Já a face do terreno em que as lavouras foram implantadas reflete na maturação tardia dos grãos e proporciona elevado percentual de frutos verdes no momento da colheita. Lavouras com face de exposição solar ao norte recebem maior quantidade de radiação solar, de modo que há maior acúmulo de fotoassimilados e compostos químicos nos grãos, auxiliando no processo de amadurecimento dos mesmos. Quando a colheita é homogênea e os frutos estão em plena maturação (estádio cereja), o rendimento de grãos é maior quando comparado a uma colheita onde os frutos estão verdes e menores, que resulta em menos unidades de sacas por hectare (CUSTÓDIO *et al.*, 2015).

Assim como o clima e terreno da região, a capacidade produtiva da cultivar explorada está profundamente relacionada à produção final da lavoura cafeeira. Deste modo, o melhoramento genético tem trabalhado no desenvolvimento de cultivares mais produtivas, a fim de aumentar a produtividade da lavoura (sc.ha^{-1}), sem que haja aumento de área plantada (CARVALHO *et al.*, 2022). Outros aspectos agrônômicos que também refletem na quantidade e qualidade dos grãos de café estão sendo estudados pelos melhoristas para diminuir a influência desses fatores e aumentar a estabilidade da produção da lavoura, como maior adaptabilidade ao ambiente, melhor uniformidade de maturação em épocas diferenciadas, tolerância à seca e maior resistência a diferentes pragas e doenças (FERRÃO *et al.*, 2021).

Além destes fatores, existe uma ampla gama de variáveis que estão relacionadas à produção final do cafezal. A nutrição da lavoura se enquadra nessa afirmação. O cafeeiro necessita de nutrientes essenciais, os quais são classificados em macronutrientes (N, P, K, Ca, Mg e S) e micronutrientes (Fe, Mn, Zn, Cu, B, Cl e Mo), para assegurar o correto desenvolvimento vegetativo da planta e a alta produção de flores e frutos (FERREIRA *et al.*, 2021). Assim, é fundamental que o produtor conheça o estado nutricional dos solos e do cafeeiro, a fim de fornecer o melhor manejo de fertilidade e equilíbrio nutricional à lavoura, que reflete não apenas no aumento da produtividade, mas também em maior lucro ao produtor e na conservação do solo, da água e do meio ambiente (FERREIRA *et al.*, 2021; PIRES *et al.*, 2022).

Segundo Moreira *et al.* (2019), práticas de manejo e técnicas de condução da lavoura, presentes desde a mecanização e automação do plantio até a colheita dos grãos, são fatores indiretos que influenciam nos ganhos de produção quando aplicados corretamente à lavoura. O adensamento da lavoura é uma prática que vem sendo discutida na literatura, uma vez que reflete na quantidade de frutos produzidos por pé e na produtividade da área total. Estudos desenvolvidos por Verdin Filho *et al.* (2019) mostram que maiores espaçamentos na linha de cultivo e entre pés de café resultam em plantas mais baixas e com maior produção unitária. Já em espaçamentos menores, observa-se maior produtividade da lavoura devido ao acréscimo no número de plantas por área.

Pesquisas e estudos mostram os diversos benefícios da prática de consorciamento do café com outras culturas, como leguminosas, forrageiras, espécies arbóreas, madeireiras e frutíferas (FONSECA *et al.*, 2023; MOREIRA, 2004; OLIVEIRA *et al.*, 2024; PARAJARA *et al.*, 2020). Entre as vantagens do plantio consorciado, destacam-se a ciclagem de diferentes nutrientes, aumento e conservação do teor de umidade do solo, favorece a areação do solo,

proporciona maior controle de plantas daninhas devido à alelopatia ou competitividade, e mantém a matéria orgânica entre linhas do cafeeiro, além de contribuir para maior qualidade da bebida dos grãos (FREITAS *et al.*, 2020; SIQUEIRA *et al.*, 2020). O consorciamento também pode interferir na produção da lavoura. O trabalho de Zacarias *et al.* (2020) mostrou que o cultivo de café associado com leguminosas (feijão-guandu, mucuna preta; feijão-de-porco e margaridão mexicano) proporcionou maior rentabilidade à cultura e contribui significativamente para a produção dos pés de café, independentemente da espécie de leguminosa consorciada.

Se tratando ainda sobre a produção do cafeeiro, a adoção de sistemas de irrigação na área agrícola facilita o manejo da umidade no solo e no ar e minimiza as incertezas quanto à disponibilidade de água à lavoura ao longo do seu ciclo de cultivo. A umidade do solo e o teor de água disponível para as plantas têm profunda relação com a germinação de sementes, o desenvolvimento de mudas, o crescimento das estruturas vegetativas, a indução floral, florescimento, e o amadurecimento dos frutos, sendo necessário o correto manejo e fornecimento desse elemento nos diferentes estágios de qualquer cultura agrícola. Além disso, a umidade do solo pode contribuir para a estabilidade de produção da lavoura cafeeira, diminuindo a bienalidade de produção da cultura entre os anos agrícolas (BASSOI *et al.*, 2019; GUEDES *et al.*, 2019; VEIGA *et al.*, 2023).

Vários outros aspectos ao longo do ciclo do cafeeiro podem interferir na produção das plantas e diminuir a produtividade da lavoura no final do ano safra. As pesquisas e trabalhos científicos contribuem para minimizar os fatores controláveis que podem afetar esse aspecto da cultura agrícola. O monitoramento da produtividade é essencial nesse sentido e será discutido no próximo tópico.

2.3. Monitoramento da Produção Agrícola

O monitoramento da cultura agrícola no campo possibilita ao agricultor acompanhar as interações do sistema solo-planta-atmosfera e detectar e gerir os fatores que levam à redução do rendimento das plantas e às perdas na produção (QUEIROZ *et al.*, 2020).

Nesse sentido, o monitoramento das condições ambientais e dos parâmetros agrícolas permite a antecipação e prevenção de incidência de pragas, doenças, plantas invasoras e de impactos climáticos na lavoura, como mudanças na temperatura, precipitação, umidade do solo, umidade do ar, períodos de escassez hídrica e quantidade de radiação solar suficiente (BARROS *et al.*, 2021; GUTIERRES; NEVES, 2021; NÓIA JUNIOR *et al.*, 2022; RODRIGUES FILHO

et al., 2023; SANTOS *et al.*, 2020). A prevenção da ocorrência desses fatores é essencial para que os agricultores tomem decisões acertadas em relação à irrigação, adubação, controle fitossanitário e manejo da lavoura em geral, a fim de mitigar os danos causados durante o crescimento, desenvolvimento e colheita dos produtos agrícolas (BARROS *et al.*, 2021; RODRIGUES FILHO *et al.*, 2023).

O monitoramento e a estimativa da produtividade da cultura também são realizados com base nas informações de clima, solo, fitossanidade e condução da lavoura durante todo o ciclo da planta. Para estimativas eficientes de produção da cultura, vários métodos foram desenvolvidos e aprimorados ao longo dos anos. Os métodos tradicionais de monitoramento da produção são baseados na aplicação de questionários e relatórios aos produtores rurais e na coleta de informações por técnicos, agrônomos, órgãos, entidades públicas e privadas e demais colaboradores. Os dados colhidos são traduzidos em estatísticas de previsão e acompanhamento de safras a nível municipal, regional ou estadual (BACH *et al.*, 2020; IBGE, 2024a).

A Companhia Nacional de Abastecimento (CONAB) apresenta regularmente levantamentos de previsão e estimativa de safras para diversas *commodities* brasileiras, além de informações a respeito de exportações, importações, área plantada, cultivares exploradas e outros aspectos econômicos e sociais da produção agropecuária. Para tanto, a CONAB aplica a metodologia tradicional para recolhimento de informações de uma amostra de produtores de diferentes regiões do Brasil e conta com aproximadamente 80 técnicos para a realização dessa tarefa. A instituição utiliza modelos estatísticos e pacotes tecnológicos modais para associar as informações concedidas a dados agrometeorológicos e espectrais da cultura, a fim de obter resultados eficientes e próximos ao real (CONAB, 2021).

Atualmente, as entidades e instituições agrícolas procuram formas mais rápidas e menos exaustivas para a obtenção de dados e simulação de produção nas próximas safras. Nesse contexto, a aplicação de modelos matemáticos e estatísticos no monitoramento de produtividade tem se tornado popular, uma vez que reúnem e analisam as interações entre as variáveis ambientais, genéticas e demais parâmetros agrônômicos e sua relação com a produção da lavoura, apresentando como resultado uma simulação dos rendimentos a serem obtidos nas próximas safras (SILVA; ARAÚJO, 2020). As vantagens desse método são inúmeras, entre as quais se destacam a capacidade de lidar com um grande volume de variáveis, a economia de tempo, a aplicabilidade em grandes áreas e o alto desempenho de processamento (BUSKE *et al.*, 2019; SANTOS; BEKO; LEITHARDT, 2022).

Um dos modelos matemáticos mais antigos e eficientes foi elaborado por Doorenbos e Kassam (1979) como uma proposta de avaliar o rendimento de culturas anuais sob fornecimento adequado ou limitado de água. O modelo desenvolvido se baseia na seguinte equação:

$$(1 - Y_a/Y_m) = k_y (1 - ET_a/ET_m);$$

onde: Y_a = Produção real; (1)

Y_m = Produção máxima;

k_y = fator de resposta à produtividade

ET_a = Evapotranspiração real

ET_m = Evapotranspiração máxima

Posteriormente, o modelo de Doorenbos e Kassam foi adaptado à cultura do café por Rosa *et al.* (2010), que adicionaram variáveis espectrais com o objetivo de abordar mais fatores que alteram a produtividade do cafeeiro e amenizar os efeitos da bienalidade da cultura no modelo. Apesar do novo modelo, denominado de agrometeorológico-espectral, apresentar eficácia de até 80% na estimativa de produção do cafeeiro, os autores afirmam que as diferenças entre os resultados reais e os resultados estimados ocorrem porque nem todos os fatores que interferem na variável dependente (produção) foram abordados no modelo.

Assim como o modelo proposto por Rosa *et al.* (2010), muitos outros estudos têm sido desenvolvidos para adicionar as tecnologias da agricultura de precisão ao monitoramento e à estimativas de safras de diferentes culturas agrícolas. O SR é uma ferramenta potente que vem sendo explorada nesse sentido (LEIVAS *et al.*, 2022a). O modelo de Nascimento (2019) correlaciona o vigor vegetativo das plantas de café com a produtividade de grãos do ano anterior para estimar a produção de safras futuras. Para tanto, o autor utilizou uma série de imagens dos satélites China-Brazil Earth Resources Satellite (Cbbers-4), Land Remote Sensing Satellite (Landsat-8), Resourcesat-2 e Sentinel-2 para compilar dados sobre a resposta espectral de todo o ciclo fenológico do cafeeiro. Como conclusão, Nascimento (2019) aponta a eficiência do modelo, que pode ser aplicado com até três meses de antecedência para a estimativa de produção na próxima safra.

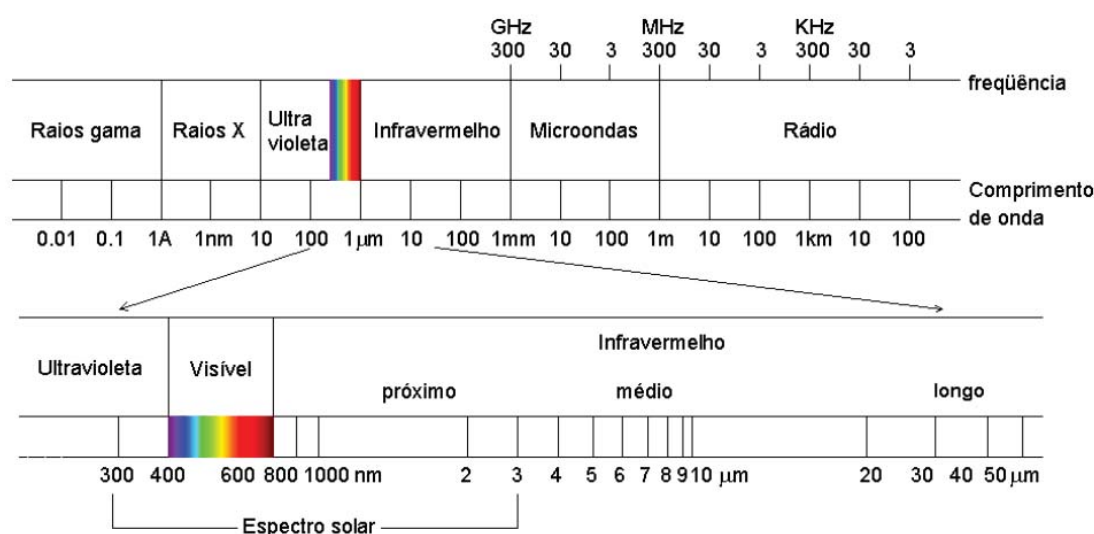
No tópico a seguir, será discutida a fundo a definição e aplicação do SR dentro da agricultura, enfatizando o recorrente uso dessa tecnologia na modelagem para previsão/projeção de produtividade de culturas agrícolas, inclusive o cafeeiro.

2.4. Sensoriamento Remoto e Índices de Vegetação

O SR é capaz de obter informações sobre a superfície terrestre por meio da interação da radiação eletromagnética (REM) com alvos de interesse, sem que haja contato direto com esses alvos. Sensores remotos passivos utilizam os raios solares como fonte de REM medindo a quantidade de radiação que é refletida pelo objeto-alvo (CRUSIOL *et al.*, 2024).

A REM ou energia eletromagnética é emitida por qualquer objeto da face da Terra com temperatura acima de zero absoluto ($-273,15^{\circ}\text{C}$, $-459,67^{\circ}\text{F}$ ou 0 K), tornando-se, portanto, uma fonte de REM (GONÇALVES, 2022). O termo “Espectro Eletromagnético” é o arranjo da REM ordenada de maneira contínua em função de seu comprimento de onda ou de sua frequência (Figura 3) (GOMES; CUBAS, 2021). A maioria dos sistemas passivos de SR utilizados para levantamentos de características terrestres atuam na faixa do visível e do infravermelho, variando entre $0,38\mu\text{m}$ e 1 mm . No caso dos sistemas ativos de sensoriamento (radares), a faixa utilizada é a de micro-ondas, que pode variar de 1 mm a 100 cm (GONÇALVES, 2022).

Figura 3 - Espectro eletromagnético.



Fonte: Dornelles (2008).

No SR aplicado à agricultura de precisão, geralmente são utilizadas as faixas do azul (450 nm), verde (550 nm), vermelho (650 nm), borda vermelha (700 nm) e infravermelho próximo (800–900 nm) do espectro eletromagnético. A maioria dos sensores possuem bandas espectrais que captam a irradiância do objeto nessas seções do espectro eletromagnético e são categorizados em sensores multiespectrais, com 10 a 15 bandas espectrais, e hiperespectrais, que possuem mais de 50 bandas (GIOVOS *et al.*, 2021).

As bandas multiespectrais são usadas para calcular os Índices de Vegetação (IVs) da cultura. Os IVs são o principal produto do SR e são obtidos através da combinação de ao menos

duas bandas espectrais. Esses índices são projetados para maximizar as características da lavoura agrícola e minimizar os fatores externos a ela, como solo e atmosfera, e podem fornecer informações espectrais e temporais relevantes sobre as plantas (TSOUROS; BIBI; SARIGIANNIDIS, 2019).

Os IVs são aplicados em diferentes estágios de desenvolvimento das plantas e são utilizados em diferentes áreas importantes para a condução da lavoura, como a avaliação da condição fisiológica das plantas, percepção de estresses causados por secas, estimativas de índice de área foliar e de clorofila presente nas folhas e permite o monitoramento do potencial produtivo das plantas, do estado fitossanitário da lavoura e da colheita dos produtos agrícolas, entre outras aplicações (CUI *et al.*, 2019; DAVIDSON *et al.*, 2022; DONG *et al.*, 2019; NDUKU *et al.*, 2023).

Para a utilização das bandas de sensores remotos e, posteriormente, o cálculo de IVs, é necessário que as bandas sejam corrigidas em relação à interferência atmosférica no momento de captação da irradiância. A correção atmosférica remove efeitos da atmosfera e da variação da iluminação solar e fornece os valores reais de reflectância dos objetos terrestres. A calibração e o ajuste das condições atmosféricas nas imagens obtidas por sensores são essenciais para o correto uso das mesmas (DELAVARPOUR *et al.*, 2021). Assim, muitos satélites fornecem as imagens já corrigidas e calibradas, uma vez que esse processo pode ser complexo e trabalhoso para o usuário.

Com a importância dos IVs e suas múltiplas aplicações, vários deles foram desenvolvidos ao longo dos anos com objetivos e funções diferentes, totalizando 519 índices criados atualmente (INDEX DATABASE, 2024). O NDVI (sigla de *Normalized Difference Vegetation Index*) é um dos mais conhecidos nas ciências agrárias, pois pode ser relacionado à biomassa da vegetação em diferentes períodos de crescimento e geralmente é utilizado para estudos das fases vegetativas da cultura, mudanças no uso do solo e condição fisiológica das plantas (BARROS; FARIAS; MARINHO, 2020). O NDVI explora duas bandas das imagens espectrais, sendo elas a vermelha (Red) e o infravermelho próximo (NIR), através do cálculo a seguir:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \quad (2)$$

Os resultados deste cálculo podem variar entre -1 e 1, sendo os valores negativos relacionados a áreas sem nenhuma vegetação ou áreas úmidas, como solo, rios, lagoas e áreas

urbanas, e valores próximos de 1 sugerem a presença de alta biomassa vegetal (BARROS; FARIAS; MARINHO, 2020).

Outros índices foram desenvolvidos para ajustar o NDVI às condições terrestres. O EVI (*Enhanced Vegetation Index*) minimiza a influência da atmosfera, do solo e dos efeitos de aerossóis da atmosfera nas bandas das imagens de satélite. Esse índice é amplamente utilizado no monitoramento e detecção de doenças e perdas de biomassa em lavouras com alto índice foliar (CASTRO *et al.*, 2018; CASTRO, 2019). O cálculo do EVI é apresentado a seguir:

$$EVI = \frac{G * (NIR - Red)}{(NIR + C1) * (Red - C2) * (Blue + L)} \quad (3)$$

Onde:

G: fator de ganho (2,5);

C1 e C2: são coeficientes de ajuste com valores de 6 e 7,5, respectivamente;

L: fator de ajuste para o solo. Varia entre 0 e 1;

NIR: banda do Infravermelho Próximo;

Red: banda do vermelho;

Blue: banda do azul.

Já o Índice SAVI (*Soil Adjusted Vegetation Index*) acrescenta um fator de correção para a influência do solo, o que resulta em uma representação mais fiel da vegetação se comparado ao NDVI. O SAVI é uma alternativa para áreas com vegetação mais aberta como o cerrado, minimizando a influência da reflectância dos solos ou do sub-bosque no cálculo do índice. A fórmula para obtenção desse índice considera as bandas do NIR e do vermelho, e um fator de correlação (L), que varia entre 0 (áreas com muita vegetação) a 1 (áreas com pouca vegetação) (LEIVAS *et al.*, 2022b).

$$SAVI = \frac{NIR - Red}{NIR + Red + L} * (1+L) \quad (4)$$

Outro índice muito utilizado no meio agrícola é chamado de NDWI (*Normalized Difference Water Index*), que está relacionado com a umidade e presença de água no ambiente. Sua aplicação possibilita a avaliação de estresse hídrico na vegetação e pode ser calculado pela seguinte fórmula (ALVES *et al.*, 2020):

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (5)$$

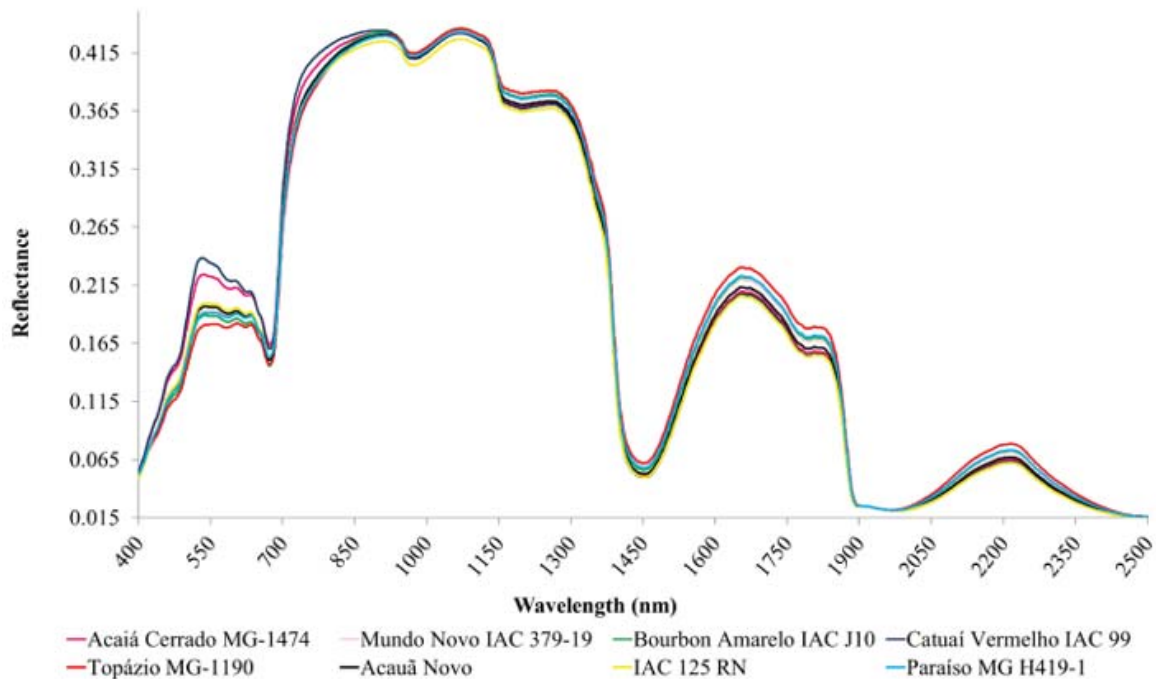
Um dos usos mais recorrentes dos IVs é estudar a relação entre os IVs e a produção futura da lavoura, como o estudo realizado por Almeida *et al.* (2023) que demonstraram a alta correlação entre a produção de milho com os IVs NDVI, SAVI e EVI, principalmente durante o estágio vegetativo V8 da cultura. Resultados semelhantes foram encontrados por Leda, Gonçalves e Lima (2019), quando relacionaram os índices NDVI, EVI e MSAVI (derivado do SAVI) com a produção de lavouras de Cana de Açúcar e concluíram que os IVs mostram-se promissores na explicação da produtividade agrícola.

2.4.1. Assinatura espectral da vegetação e Comportamento espectral do cafeeiro.

Como apresentado anteriormente, na agricultura são exploradas frequentemente as bandas correspondentes aos comprimentos de onda do azul, verde, vermelho, borda vermelha e infravermelho próximo das imagens de satélite. O interesse nessas bandas em particular se dá porque compõem as características espectrais das plantas na superfície terrestre, ou de corpos próximos a áreas de vegetação, como estradas, cursos d'água, áreas urbanas, etc. A quantidade de radiação que é absorvida por qualquer elemento nas diferentes faixas do espectro eletromagnético é chamada de “comportamento ou assinatura espectral” e demonstra as características espectrais nas faixas do visível e do infravermelho do objeto alvo (SANTOS, 2015).

No caso de uma vegetação sadia, a absorção de ondas na região do visível é maior, uma vez que a clorofila, pigmento presente nas folhas das plantas, retém comprimentos de onda azuis e, principalmente, vermelhos para a realização de fotossíntese e geração de energia para a planta, ocasionando em baixa reflectância dessas faixas do espectro eletromagnético. Já os comprimentos de onda verde e infravermelho próximo, correspondentes a região de 550 nm e 800–900 nm, respectivamente, não são de interesse para o vegetal, e são refletidos por ele devido a morfologia celular e estrutura da vegetação (FERREIRA; FERREIRA; FERREIRA, 2008; INÁCIO, 2016; LOURENÇO; LANDIM, 2004; SANTOS, 2015) A Figura 04 foi desenvolvida por Freitas *et al.* (2023) e mostra a assinatura espectral de algumas cultivares de *Coffea arabica* L.

Figura 4 – Comportamento Espectral das Cultivares de *Coffea arabica* L. Acaiá Cerrado MG 1474, Mundo Novo IAC 379 – 19, Yellow Bourbon IAC J10, Catuai Vermelho IAC 99, Topazio MG 1190, Acauã New, IAC 125 RN e Paradise MG H 419-1.



Fonte: Freitas *et al.* (2023).

A fim de estudar o comportamento espectral do cafeeiro, Carrijo, Borges e Souza (2021) procuraram entender as respostas das bandas RED e NIR de imagens do satélite Sentinel 2A e 2B para cada fase fenológica de desenvolvimento do pé de café (Figura 5). Segundo os autores, os períodos de maior reflectância NIR correspondem a 1º, a 3º e a 4º fase do ciclo fenológico do café, desenvolvido por Camargo e Camargo (2001), quando a planta está passando por processos de enfolhamento e formação de gemas foliares, desenvolvimento de flores e granação dos frutos. Em relação a banda RED, a 2º, a 3º e a 4º fase, que ocorrem durante os meses de abril a agosto, setembro a dezembro e janeiro a março, são caracterizadas por menor reflectância dessa banda, mostrando que, nesses períodos, a planta possui maior capacidade de absorver comprimento de ondas do vermelho. Provavelmente, esse fato está associado à pluviosidade de região, que se inicia em outubro e termina em meados de abril do próximo ano, e à fase de vegetação do pé de café.

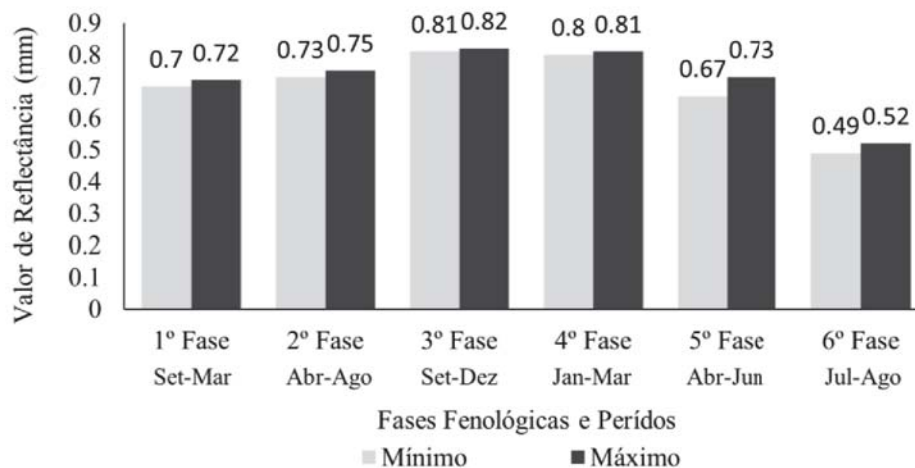
Figura 5 – Reflectância das bandas RED e NIR durante as 6 fases fenológicas do café.



Fonte: Carrijo, Borges, Souza (2021). Adaptado.

Além disso, Carrijo, Borges e Souza (2021) demonstram o comportamento de NDVI durante os anos vegetativos e reprodutivos da planta (Figura 6). Verifica-se que os maiores valores de NDVI são encontrados quando ocorre a florada do cafeeiro, nos primeiros meses do 2º ano fenológico. Já os menores valores de NDVI são encontrados em sua 6ª fase, período pós-safra, quando ocorre um alto desfolhamento causado pela colheita dos frutos e inicia-se a auto-poda.

Figura 6 – Comportamento do índice NDVI durante as 6 fases fenológicas do café.



Fonte: Carrijo, Borges, Souza (2021). Adaptado.

Nota-se, na figura 6, que as primeiras quatro fases do ciclo fenológico do cafeeiro são épocas de alto NDVI. Este resultado concorda com a afirmação de Bernardes *et al.* (2013, p.

0720), que explicam o aumento de biomassa expressiva no primeiro ano fenológico do café, ano de baixa bienalidade de produção de frutos.

Em anos de alta produção, a planta direciona maior parte dos fotoassimilados para os frutos em detrimento de novos ramos que seriam responsáveis pela produção do ano seguinte. Em anos de menor produção a planta trabalha preferencialmente para geração de novos ramos (aumento da biomassa foliar) os quais produzirão no ano seguinte.

(BERNARDES *et al.*, 2013, p. 0720).

Os resultados do estudo de Bernardes *et al.* (2013) evidenciam a possibilidade de desenvolver um modelo de estimativa de produção cafeeira utilizando índices de vegetação associados a características biofísicas da planta, e refletem a dependência da produtividade em função da biomassa foliar.

Moreira, Adami e Rudorff (2004) estudaram a assinatura espectral do cafeeiro em relação a floresta nativa e pastagem. As análises demonstraram que, apesar de serem estruturalmente diferentes, o café em formação e a pastagem possuem comportamento semelhantes nas bandas verde e infravermelho, tornando difícil sua separação. Porém, quando em produção, o café difere da classe de floresta nativa e pastagem em qualquer uma das bandas estudadas (verde, vermelho e infravermelho). Além disso, a banda do infravermelho próximo é muito útil para diferenciar o café, tanto em formação quanto em produção, das demais classes de uso da terra, uma vez que difere estatisticamente de ambas as vegetações.

2.5. Imagens Sentinel – 2

Os satélites Sentinel foram desenvolvidos inicialmente pelo programa de Monitorização Global do Ambiente e da Segurança (GMES) em parceria com a Comissão Europeia (*European Commission* - EC) e com a Agência Espacial Europeia (*European Space Agency* – ESA) (BERTINI *et al.*, 2012). Em 2014, o GMES passou a ser chamado de Programa Europeu de Observação da Terra, ou programa Copernicus, e firmou associação com os Estados-Membros da UE, a Organização Europeia para a Exploração de Satélites Meteorológicos (EUMETSAT), o Centro Europeu de Previsões Meteorológicas de Médio Prazo (ECMWF), e a Mercator Ocean International (ESA, 2024; PHIRI *et al.*, 2020).

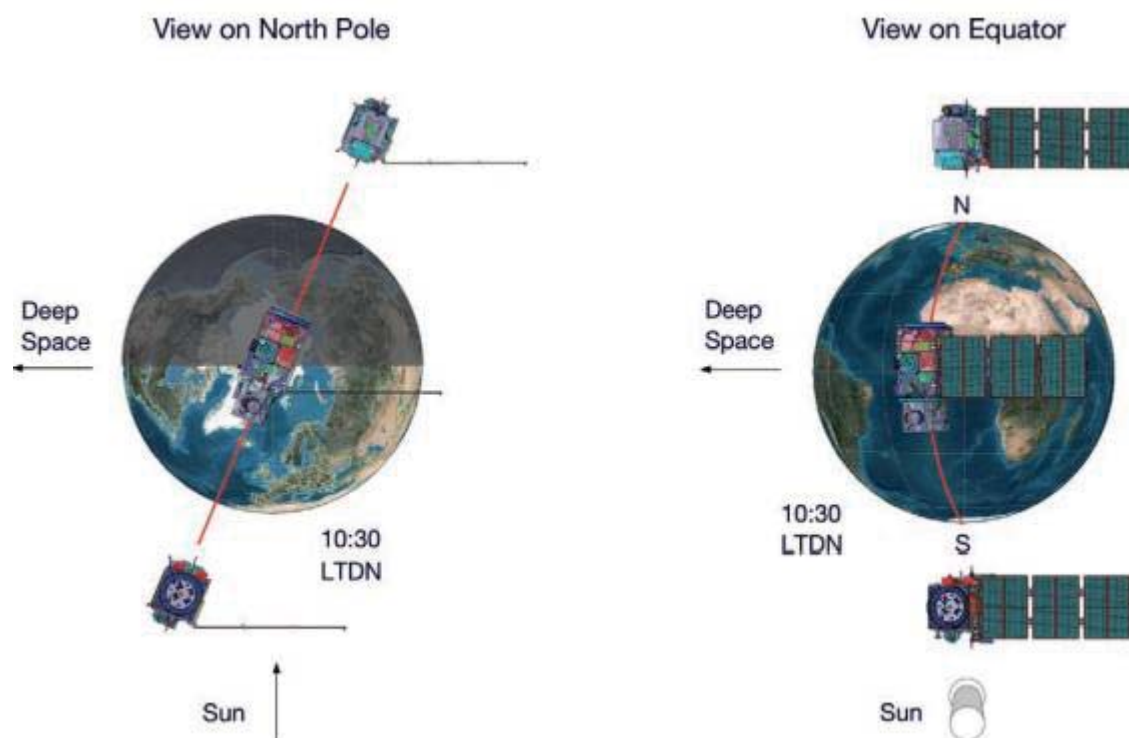
As missões Sentinel são sustentadas pelo Copernicus Ground Segment, que supervisiona suas operações e a aquisição, produção e disseminação de dados. A Copernicus lançou uma série de constelações de satélites gêmeos e satélites individuais para a aquisição de diferentes dados do planeta Terra. A primeira constelação de satélites foi formada pelos satélites Sentinel 1A e 1B, lançados nos anos de 2014 e 2016, respectivamente, ambos com o objetivo

de capturar informações sobre a superfície terrestre e sobre o comportamento oceânico. Quase que simultaneamente, foram lançados os satélites da missão Sentinel-2 em 2015 e 2017. Ainda nesse período, o programa Copernicus lançou a missão Sentinel-3, dedicado à oceanografia, em 2016 e 2017. Também em 2017, foi lançado o único satélite da missão Sentinel-5, com o intuito de monitorar a atmosfera do planeta e adquirir dados sobre a poluição do ar. Os satélites da constelação Sentinel-4 ainda estão em construção, assim como o satélite Sentinel-1C, que irá substituir o satélite Sentinel 1B, cuja missão foi encerrada em julho de 2022 (ESA, 2024).

Os satélites da missão Sentinel-2 foram construídos com a finalidade de complementar as observações dos satélites Landsat e SPOT (*Satellite Pour l'Observation de la Terre*) e melhorar a disponibilidade de dados para os usuários. Os objetivos da missão Sentinel 2A e 2B são fornecer aquisições sistemáticas de imagens multiespectrais de alta resolução com alta frequência de revisita, para garantir a continuidade das imagens multiespectrais disponíveis e fornecer observações para a próxima geração de produtos operacionais. Conseqüentemente, o Sentinel-2 contribui diretamente para o fornecimento de dados para serviços como gestão de riscos (enchentes e incêndios florestais, subsidência e deslizamentos de terra), estado e mudanças do uso da terra, monitoramento florestal, segurança alimentar/sistemas de alerta precoce, gestão da água e proteção do solo, entre outras informações para pesquisas e estudos (BERTINI *et al.*, 2012)

Para atingir tal propósito, os satélites Sentinel-2 cobrem a superfície terrestre entre as latitudes de 56°S a 83°N, com um campo de visão de 290 km, e fornecem produtos de dados a cada 5 dias. Essa alta resolução temporal é possível devido a presença do satélite Sentinel-2B, de modo que os dois satélites operam em fase orbital de 180° (Figura 7). Assim, os satélites, que estão a uma altitude de 786 km e inclinados em 22° em relação a sua órbita, completam aproximadamente 14,3 órbitas por dia entorno do planeta e, a cada órbita, os satélites passam em média 16,3 minutos observando diretamente a superfície da Terra (BERTINI *et al.*, 2012).

Figura 7 - Configuração da órbita dos satélites Sentinel-2.



Fonte: Bertini *et al.* (2012).

Os produtos gerados pelas missões Sentinel-2 são imagens de alta resolução espectral. Cada imagem é composta por 13 bandas, sendo 4 delas pertencentes as faixas do visível e do infravermelho próximo do espectro eletromagnético, com resolução espacial de 10 m, 6 bandas nas faixas do vermelho médio e do infravermelho de ondas curtas, com pixels de 20 m, e 3 bandas utilizadas para correções atmosféricas, com resolução de 60 m (Tabela 1). As imagens concedidas são ortorretificadas, ou seja, são ajustadas conforme a real posição em relação ao solo, e possuem calibração radiométrica, que faz com a conversão dos valores de radiância em valores de reflectância, possibilitando que a análise das imagens produza resultados confiáveis. Além destas vantagens, as imagens Sentinel-2 possuem política de dados de livre acesso, tornando as imagens disponíveis gratuitamente para o uso em diferentes finalidades (ENGESAT, 2024; PHIRI *et al.*, 2020).

Tabela 1 – Bandas das imagens multiespectrais dos satélites Sentinel 2A e Sentinel 2B.

Resolução	Nº da Banda	Nome da Banda	Comprimento de Onda Central (nanômetro)
10 m	B02	Blue (Azul)	490
	B03	Green (Verde)	560
	B04	Red (Vermelho)	665
	B08	NIR (Infravermelho Próximo)	842
20 m	B05	Red Edge 1	705
	B06	Red Edge 2	740
	B07	Red Edge 3	783
	B08A	Red Edge 4	865
	B11	SWIR 1	1610
	B12	SWIR 2	2190
60 m	B01	Aerosol	443
	B09	Water Vapor	940
	B10	Cirrus	1375

Fonte: ENGESAT (2024). Adaptado.

Diversos estudos foram realizados para avaliar a aplicabilidade das imagens Sentinel-2 em vários ramos da pesquisa científica. Grabska *et al.* (2019) aproveitaram a alta resolução espacial dessas imagens e mapearam as espécies de árvores presentes nas floretas nativas dos Cárpatos poloneses, no sul da Polônia. Já Rotata et al (2019) desenvolveram um novo algoritmo para detecção de áreas queimadas na região da África Subaariana. Os autores utilizaram 11.000 imagens Sentinel-2 para a construção do algoritmo, o que possibilitou a criação de um banco de dados de incêndios na área de estudo, incluindo até mesmo pequenas queimadas (<100 ha). Dentro do setor agrícola, Segarra *et al.* (2020) avaliaram a ampla utilização das imagens Sentinel na agricultura de precisão, que permite a exploração de dados e desenvolvimento de novas tecnologias com base no sensoriamento remoto. Como conclusão, os autores afirmam que as imagens Sentinel-2 podem ser aplicadas no e gerenciamento de estresses bióticos e abióticos, na detecção das necessidades hídricas das culturas, exigências nutricionais de lavoura, e, até mesmo, no monitoramento de safra agrícolas.

2.6. *Machine Learning*

O termo “Aprendizagem de Máquinas” (ou *Machine Learning*, em inglês) se refere ao processo de extrair conhecimento e treinar computadores/software para a automatização de processos e análise da melhor tomada de decisão (MÜLLER; GUIDO, 2016). Isso requer um volumoso conjunto de dados previamente informado ao computador, permitindo que ele

consiga estabelecer padrões, gerar hipóteses e formular modelos a partir desses dados (LUDERMIR, 2021; SOUZA *et al.*, 2020).

A aprendizagem de máquinas surgiu com o desenvolvimento da inteligência artificial e foi moldada por diversas áreas do conhecimento, incluindo probabilidade e estatística, teoria da complexidade computacional, teoria da informação, filosofia, psicologia e neurobiologia (CERRI; CARVALHO, 2017). Nesse sentido, a aprendizagem de máquina está inserida em vários contextos e possui diferentes aplicações, como classificação de tumores cerebrais, previsão de vazões fluviais em rios, identificação de ataques de pragas em lavouras agrícolas, identificação de fitofisionomias em diferentes biomas e estimativa do teor de clorofila em plantas agrícolas, entre outras (ALBA *et al.*, 2022; CORREA *et al.*, 2022; GUMAEI *et al.*, 2019; PEREIRA *et al.*, 2023; YASEEN *et al.*, 2019).

Existem diferentes processos de aprendizagem de máquina aplicados a computadores. Esses processos são divididos em três classes: aprendizagem de máquina Supervisionada, aprendizagem de máquina não Supervisionada e Aprendizagem por Reforço (MARTINS; SILVA, 2020).

A aprendizagem de máquina supervisionada consiste no processo de treinamento onde o algoritmo recebe amostras de dados de um determinado problema e também os resultados obtidos com essas informações, ou seja, são fornecidas as variáveis de entrada (X) e os valores de saída (Y). Essa técnica possibilita treinar um algoritmo para ser capaz de entender a relação entre os conjuntos X e Y de modo que, ao aplicarmos o algoritmo já treinado em um ponto X ao qual ele nunca teve acesso, ele nos retorne o ponto Y esperado (IGNACIO, 2021).

Já para o processo de aprendizagem de máquina não supervisionada, os valores de resposta a um certo conjunto de dados não estão disponíveis e não são utilizados para treinamento da máquina. Nesse caso, o objetivo é que a máquina entenda e apresente os padrões existentes nas estruturas dos dados de entrada (FERNANDES; FILHO, 2019; IGNACIO, 2021).

A aprendizagem por reforço é baseada em interações entre um agente e seu ambiente. O agente é ensinado a interagir com o meio mediante a um conjunto de ações. A escolha da ação será remodelada através de punições ou recompensas para que o algoritmo apresente os melhores resultados ao final do treinamento (IGNACIO, 2021; NEFTCI; AVERBECK, 2019). A aprendizagem de máquina por reforço pode ser aplicada a negociações de ativos futuros para realizar operações de compra e venda e a fim otimizar rotas de trânsito e controle semafórico, por exemplo (BAZZAN, 2021; KINOSHITA; MENEGHETTI; BIANCHI, 2022).

A aprendizagem de máquina supervisionada é subdividida em relação aos problemas a serem solucionados, chamados de classificação e regressão. Na classificação, os dados são separados em classes escolhidas dentre uma lista já pré-definida de possibilidades. Há dois tipos de classificação: a classificação binária, que separa os dados em apenas duas esferas, e a classificação em multiclases, onde há várias possibilidades para ordenar os dados. Já a regressão consiste em prever um valor numérico pertencente a um intervalo contínuo, como prever o valor de uma casa baseado em informações como idade, número de cômodos, tamanho do terreno e outros parâmetros (HOMEM, 2020; LUDERMIR, 2021).

A aprendizagem de máquina por regressão é utilizada principalmente no desenvolvimento de modelos preditivos para diferentes áreas do conhecimento. Corrêdo *et al.*, (2021) usaram a técnica de Regressão de Mínimos Quadrados Parciais (PLSR) para associar vários tipos de amostragem de cana-de-açúcar a parâmetros de qualidade da mesma e, assim, analisar a variabilidade da qualidade da cana em campo. Já Tavares *et al.* (2021) aplicaram a regressão parcial de mínimos quadrados (PLS) para avaliar técnicas de predição do teor argila, matéria orgânica, capacidade de troca catiônica (CTC), pH, saturação por bases e nutrientes extraíveis em solos tropicais.

A ampla aplicabilidade não apenas da regressão, mas da aprendizagem de máquina como um todo, possibilitou o desenvolvimento e aperfeiçoamento de vários algoritmos atuais. Entre eles, os mais encontrados e conhecidos são: Árvores de Decisão (*Decision Tree*), Redes neurais (*Neural Networks*), *Navie Bayes*, Máquinas de Vetores de Suporte (*Support Vector Machine*), *K-Means Clustering*, Aprendizagem por agrupamento (*Ensemble Learning*) e Floresta Aleatória (*Random Forest*), entre outros (MAHESH, 2018).

A seguir, será discutido o funcionamento e aplicabilidade do algoritmo *Random Forest*, utilizado nesta pesquisa.

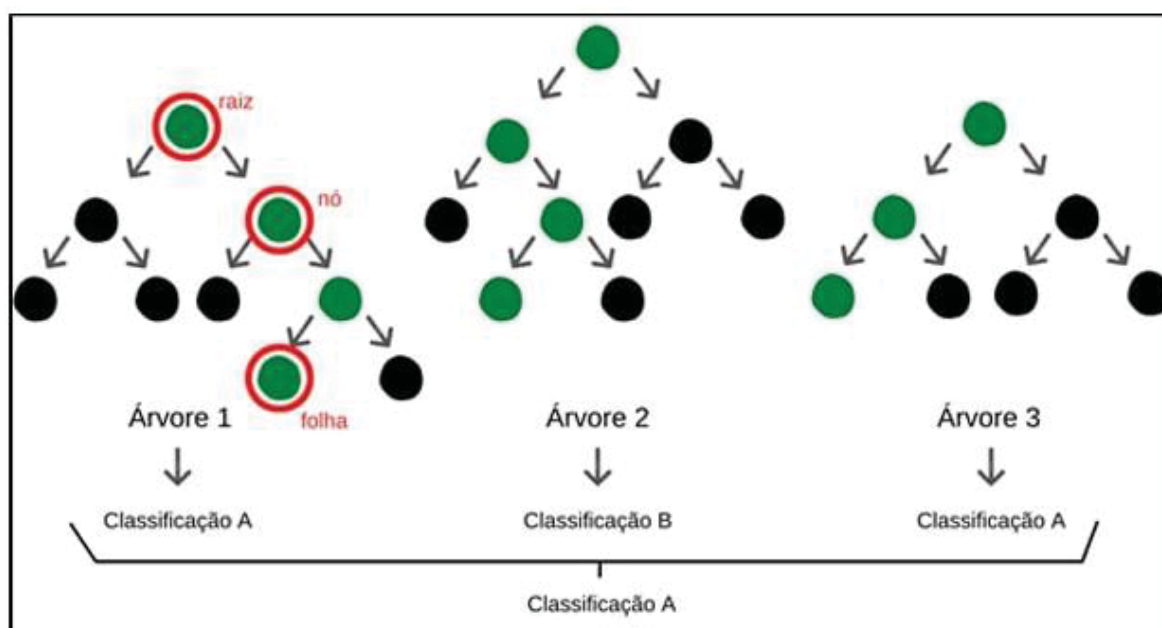
2.6.1. *Random Forest*

O algoritmo *Random Forest*, ou Floresta Aleatória, é denominado assim devido à formação e ao estudo de vários conjuntos de dados aleatórios durante o processo de treinamento do computador. Através de variáveis de entrada informadas ao algoritmo, são formados os nós da raiz da árvore de decisão. A partir de cada nó raiz, as variáveis serão usadas para a realização de diferentes partições, ramificações e formação de subconjuntos de dados descendentes simultaneamente, que são conhecidos como nós ou folhas. As sucessivas ramificações

acontecem com base na seleção do caminho que melhor descreve a variabilidade entre os valores de destino do nó raiz. O resultado final apresenta a média dos valores de previsão entre todas as árvores de decisão geradas (AMARAL; QUEVEDO; PIROLI, 2023; CONCEIÇÃO, 2022).

A figura 8 mostra os nós que representam um determinado teste *Random Forest*. Pode-se observar a construção de vários ramos a partir de um conjunto de decisões hierárquicas sobre os atributos dos dados de entrada. Assim, cada ramo representa os critérios de divisão dos testes e as folhas representam as repartições que não possuem mais descendentes. O algoritmo combina as diversas Árvores de Decisão criadas com seleções aleatórias de seus atributos, convergindo o resultado para a classificação mais realizada pelas diferentes árvores (ARIZA *et al.*, 2022; BREIMAN, 2001).

Figura 8 - Representação esquemática do *Random Forest*.



Fonte: Ariza *et al.* (2022).

A crescente utilização do *Random Forest* é explicada pela capacidade do algoritmo de estudar um grande número de variáveis e construir inúmeros submodelos (árvores de decisão). O *Random Forest* é capaz de apresentar alta robustez aliada à alta performance, mesmo com tamanhos de amostra moderados e com um número limitado de variáveis (ABREU; THOLON; LIMA, 2023).

Apesar do alto desempenho do algoritmo, o *Random Forest* tende ao superajuste das árvores de decisão, sendo necessário a determinação de parâmetros para limitar e finalizar o

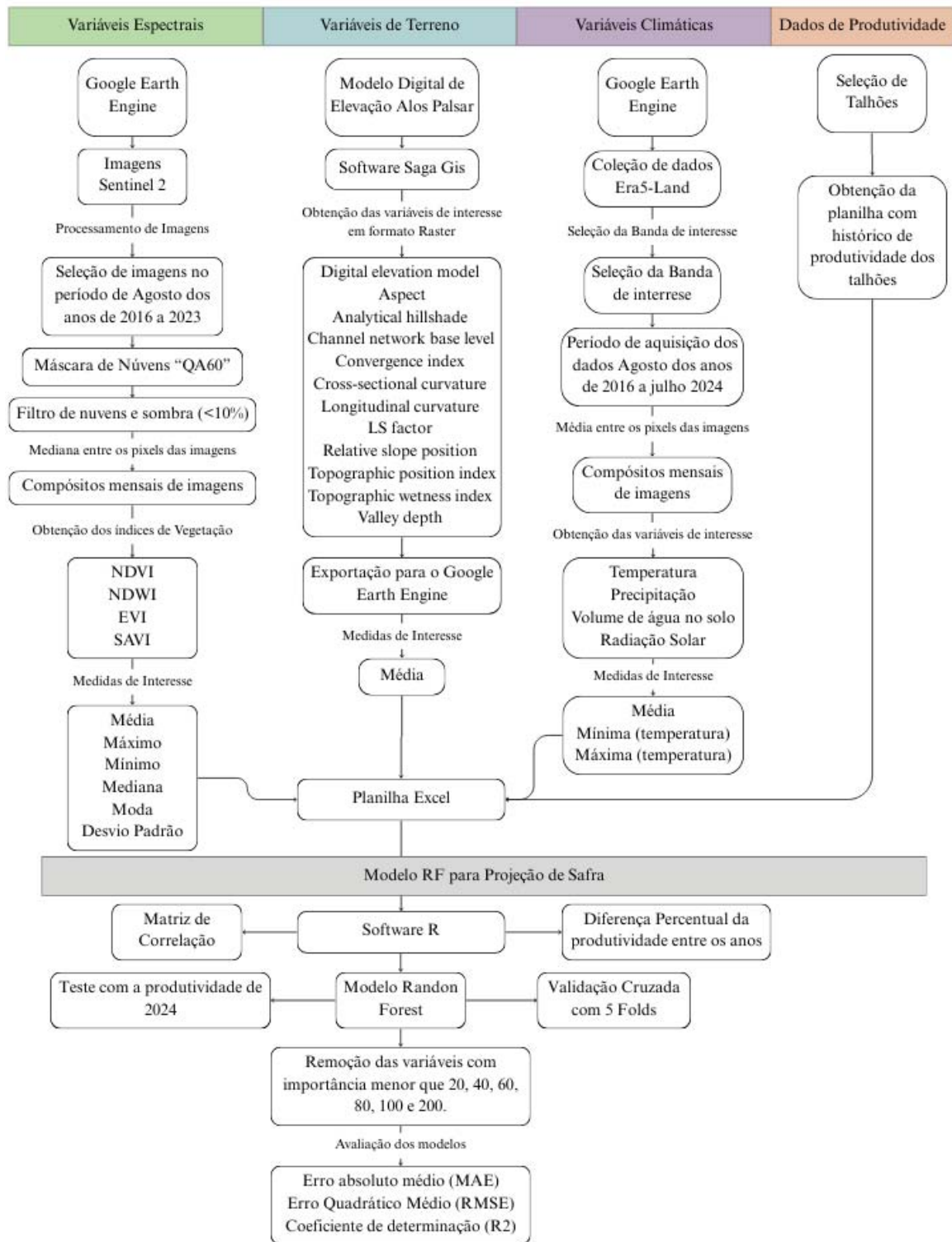
treinamento do modelo. As árvores são ajustadas com base em uma amostra dos dados de entrada, utilizada para o treinamento, e outras amostras são utilizadas para a validação do modelo criado. Para a construção das árvores de decisão, é necessário especificar a complexidade e profundidade das árvores, o número de variáveis em cada árvore, o tamanho de cada amostra de treinamento (geralmente 2/3 dos dados) e o número de árvores a serem construídas, entre outros parâmetros (MULLAINATHAN; SPIESS, 2017).

Devido à complexidade e alta capacidade do algoritmo, foram desenvolvidos métodos para avaliar o desempenho do *Random Forest* em vários setores gerais e os resultados apresentados são, majoritariamente, positivos. Rodrigues e Pereira (2021) compararam a eficiência dos algoritmos CNN (*Convolution Neural Network*), KNN (*K-Nearest Neighbors*), *Random Forest*, SVM (*Support Vector Machine*) e AdaBoost (*Adaptive Boosting*) para estimativa de produtividade na cultura de Cana-de-açúcar e concluíram que o *Random Forest* apresentou maior eficiência na previsão de produtividade das plantas quando comparado aos demais algoritmos. Resultados semelhantes foram discutidos por Silva (2023), que avaliou o desempenho de três algoritmos — CNN, KNN e *Random Forest* — na previsão de produção de café em Rondônia. Segundo o autor, o algoritmo *Random Forest* obteve melhores resultados com base nas avaliações de acurácia, proporcionando um menor Erro Absoluto Médio (MAE) e adequado Erro Percentual Absoluto Médio para o modelo preditivo.

3. MATERIAL E MÉTODOS

A Figura 9 representa o fluxo de atividades realizadas para desenvolver um modelo para projeção de safras de café utilizando técnicas de sensoriamento remoto e de *machine learning*. As etapas do processo contam com a aquisição das variáveis de entrada, incluindo o processamento de imagens, seleção de bandas, definição do período de coleta dos dados, obtenção de medidas de interesse, etc.; a própria modelagem utilizando o algoritmo *Random Forest*; a avaliação de desempenho dos modelos criados e a comparação entre os resultados obtidos e a produtividade real informada pela fazenda.

Figura 9 – Fluxograma de atividades.

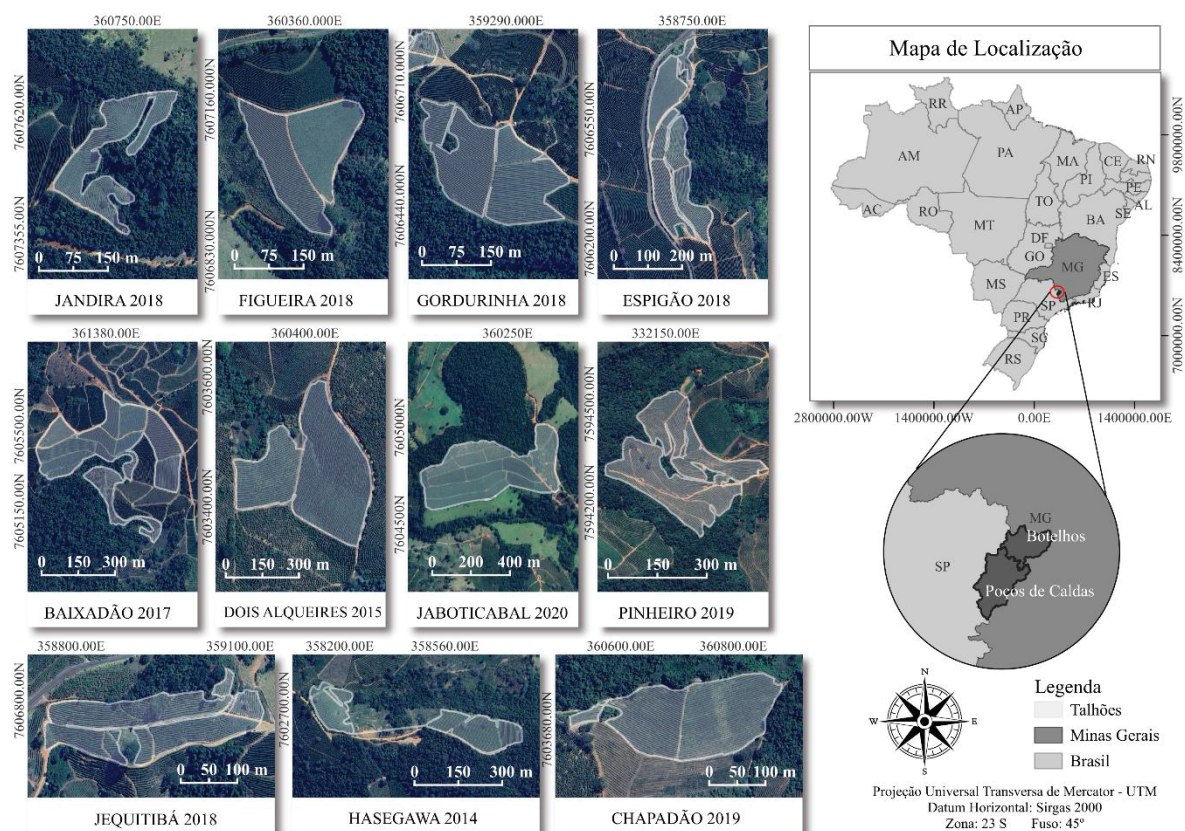


Fonte: Do autor (2025).

3.1. Caracterização da área de estudo

A presente pesquisa foi conduzida nas Fazendas Sertãozinho e Laranjal, localizadas no município de Botelhos, com latitude $21^{\circ}38'50.41''\text{S}$ e longitude $46^{\circ}23'43.14''\text{O}$, e Poços de Caldas, com coordenadas $21^{\circ}46'57.61''\text{S}$ e $46^{\circ}33'31.17''\text{O}$, no estado de Minas Gerais, sendo pertencentes à microrregião de Poços de Caldas e mesorregião Sul/Sudeste de Minas (Figura 10).

Figura 10 – Localização das lavouras de café alvo da pesquisa.



Fonte: Do autor (2025).

A região de ambas as cidades é caracterizada pelo clima CwA, segundo a classificação de Köppen, com clima temperado úmido, inverno seco e verão quente. As temperaturas em Botelhos variam entre a mínima de 8°C na estação de inverno e a máxima de 28°C quando verão. Em Poços de Caldas, a temperatura mínima chega a -6°C e a máxima alcança $31,7^{\circ}\text{C}$. O período chuvoso tem duração de 4 a 5 meses nesses municípios, com início entre outubro e novembro e fim entre março e abril. A média anual de precipitação oscila entre 1450 a 1550 mm em Botelhos e atinge cerca de 1745 mm em Poços de Caldas.

Os municípios são formados predominantemente por relevos ondulados e fortemente ondulados ou montanhosos, com altitudes médias de aproximadamente 984 m em Botelhos e 1200 m em Poços de Caldas. Segundo dados do IDE Sisema, as classes de solos predominantes na região são Cambissolo háplico distrófico, Latossolo vermelho-amarelo distrófico e Latossolo vermelho distrófico.

3.2. Talhões

A fim de construir uma base de dados de produtividade, foram selecionados 11 talhões de lavouras cafeeiras para a realização da pesquisa, sendo 10 localizados na Fazenda Sertãozinho e 1 na Fazenda Laranjal. As informações e características principais de cada talhão estão descritas na Tabela 2 e foram fornecidas pelo gerente de ambas as fazendas.

Tabela 2 - Talhões de estudo e suas respectivas cultivares, área plantada (ha), quantidade de plantas por hectare, número de plantas total e espaçamento entre a linha de plantio e entre plantas.

Talhão	Cultivar	Área (ha)	Plantas / ha	Nº de Plantas	Espaçamento
Baixadão 2017	Acauã	13,51	3.871	58.814	3,5 x 0,6
Chapadão 2019	Arara	3,59	4.050	14.500	3,5 x 0,6
Dois alqueires 2015	Catucai 2sl	2,89	4.590	13.634	3,0 x 0,5
Espigão 2018	Catuai	5,52	7.494	49.240	3,2 x 0,6
Figueira 2018	Catuai 144	4,3	4.213	19.340	3,5 x 0,6
Gordurinha 2018	Mundo novo	4,97	4.222	21.241	3,2 x 0,6
Hasegawa 2014	Arara	4,63	3.747	17.349	3,0 x 0,5
Jaboticabal 2020	Arara	18,15	3.721	67.240	3,5 x 0,6
Jandira 2018	Catucai 24/137	2,8	4.735	12.597	3,5 x 0,6
Jequitibá 2018	Arara	4,97	3.744	11.759	3,2 x 0,6
Pinheiro 2019*	Arara	7,86	3.906	29.377	3,5 x 0,6

*Talhão localizado no município de Poços de Caldas.

Fonte: Do autor (2025).

Os dados de produção foram obtidos ao final da colheita de cada talhão, desde 2017 até 2024, e, posteriormente a secagem e beneficiamento dos grãos, os dados foram convertidos em sacas de café produzidas por hectare. É possível observar que, depois de sua implantação, as lavouras levaram cerca de 2 a 3 anos para a primeira produção. Isso se dá porque, geralmente, o cafeeiro necessita de um longo período de crescimento e desenvolvimento vegetativo antes da produção de gemas reprodutivas e surgimento de flores e frutos. A Tabela 3 apresenta os dados de produtividade dos talhões em seu respectivo ano de colheita.

Tabela 3 - Data de plantio e produtividade (sc.ha⁻¹) anual de cada talhão de estudo.

Nome	Data de Plantio	Produtividade em Sc.ha ⁻¹								Média
		2017	2018	2019	2020	2021	2022	2023	2024	
Baixadão 2017	01/01/2017	-	-	43,20	51,91	37,17	24,82	32,73	68,51	43,06
Chapadão 2019	09/01/2019	-	-	-	-	9,60	0,50	13,00	15,01	9,53
Dois alqueires 2015	01/01/2015	-	26,37	9,20	60,10	10,00	16,41	17,73	34,12	24,85
Espigão 2018	01/01/2018	-	-	-	38,69	9,17	23,47	21,95	58,98	30,45
Figueira 2018	01/01/2018	-	-	-	39,38	15,5	25,61	15,27	40,89	27,33
Gordurinha 2018	01/01/2018	-	-	7,60	38,77	34,26	11,50	29,07	69,90	31,85
Hasegawa 2014	01/01/2014	52,00	24,44	30,50	44,17	21,66	16,99	10,00	22,35	27,76
Jaboticabal 2020	01/01/2020	-	-	-	-	-	11,20	31,81	47,68	30,23
Jandira 2018	01/01/2018	-	-	-	18,89	23,13	20,00	27,74	34,60	24,87
Jequitibá 2018	01/01/2018	-	-	-	20,82	34,84	12,6	44,13	15,25	25,53
Pinheiro 2019	25/02/2019	-	-	-	-	19,29	7,16	38,30	13,03	19,44

Fonte: Do autor (2025)

3.3. Organização do Banco de Dados

As variáveis avaliadas foram obtidas e/ou processadas através da plataforma online Google Earth Engine (GEE). O GEE possui um grande catálogo de dados especializados, na ordem dos multipetabyte. Sua biblioteca pode ser acessada gratuitamente por meio da linguagem de programação computacional JavaScript. Um dos principais benefícios do uso desta plataforma consiste no processamento de dados em seus próprios servidores, garantindo assim dois aspectos que são relevantes para o trabalho em detecção remota: a visualização de dados em grande quantidade e o processamento desses dados com maior rapidez e menor uso de espaço do computador. Ambos são possíveis graças ao processamento paralelo e em servidor (SILVA, 2018a).

O GEE ainda utiliza de uma aplicação de programação (API – Application Programming Interface) para disponibilizar a biblioteca de operadores/funções através de códigos de rotinas, que auxiliam o usuário a gerir, processar e analisar os dados disponibilizados pela biblioteca interna de dados ou de dados externos importados ao GEE (SILVA, 2018a).

3.3.1. Variáveis espectrais

Na presente pesquisa, foram utilizados os IVs NDVI, NDWI, EVI e SAVI, calculados a partir de imagens do satélite Sentinel-2, com resolução espacial de 10 m. Devido à quantidade de nuvens presente na área de estudo entre os meses de maior pluviosidade (outubro a março),

a data escolhida para coletar os dados espectrais foi o mês de agosto, que possui menor porcentagem de nuvens. No primeiro ano fenológico, durante o mês de agosto o cafeeiro se encontra na fase de indução e maturação de gemas florais, após passar por um período de estresse hídrico. Já no segundo ano, em agosto o cafeeiro entra em repouso e senescência dos ramos terciários após alta produção, a chamada auto-poda (CAMARGO, 2001). Estudos avaliando a correlação de índices NDVI, SAVI e NDWI com a produtividade de café gerados por meio de imagem de satélite, mostram que a fase de dormência dos botões florais e início da floração (julho a setembro) é melhor para determinar a produtividade do cafeeiro (NOGUEIRA; MOREIRA; VOLPATO, 2018; RIBEIRO et al., 2024)

No GEE, foram selecionadas as imagens de agosto com porcentagem de nuvens inferior a 10%, e aplicada sobre elas uma máscara de nuvens presente nas imagens Sentinel, denominada “QA60”. Destas imagens, foi calculada a mediana entre seus pixels, a fim de construir um compósito de imagens desse mês com os melhores pixels, sem valores nulos, lacunas, cobertura de nuvens ou sombras de nuvens. Esse procedimento foi repetido para os meses de agosto dos anos de 2016 a 2023. Considerando o ano agrícola, que começa em agosto e finaliza em julho do ano civil, os dados de IVs coletados em agosto estão associados a produtividade do próximo ano, ou seja, a produção de 2017 será estimada com base na biomassa da lavoura em agosto de 2016.

A partir dos compósitos de imagens criados, foram obtidos os IVs desejados para os talhões de estudo. Posteriormente, foram calculadas as medidas descritivas de média, máxima, mínima, mediana, moda e desvio padrão entre os pixels de cada talhão.

3.3.2. Variáveis do terreno

Os aspectos do terreno foram obtidos de um Modelo Digital de Elevação (MDE), fornecido pelo satélite *Advanced Land Observing Satellite* (ALOS), desenvolvido pela Agência Espacial Japonesa. O principal objetivo do lançamento do satélite ALOS, em 2006, foi capturar imagens de alta resolução da superfície terrestre e dados topográficos por meio do radar Palsar (ASF, 2024). No site da *Alaska Satellite Facility* (<https://asf.alaska.edu/asf/about-asf/>), foi realizado o download do MDE, com resolução espacial de 12,5 m, para a região onde estão localizados os talhões.

O MDE foi importado para o software SAGA GIS e explorado por meio da ferramenta Basic Terrain Analysis, encontrada no pacote de ferramentas Terrain Analysis, para gerar as

características do terreno avaliadas (Tabela 4). Dessa forma, foram produzidos 12 rasters, com a mesma resolução espacial de origem, que foram transferidos ao GEE. No GEE, foram calculadas as médias dos pixels de cada raster para os talhões de estudo. Os dados gerados também foram exportados na forma de planilha CSV juntamente com as variáveis espectrais.

Tabela 4 - Variáveis do terreno derivadas do MDE através do software Saga Gis.

Atributo	Descrição
Digital elevation model (DEM)	Representa a altitude (m) de cada pixel da área de estudo
Aspect (ASP)	Orientação da inclinação em relação ao Norte
Analytical hillshade (HILS)	Representa o ângulo entre o sol e a superfície do terreno
Channel network base level (CNBL)	Distância à uma rede de canais
Convergence index (CI)	Área cercada por terrenos mais altos em todas as direções
Cross-sectional curvature (CSC)	Curvatura do plano do terreno
Longitudinal curvature (LC)	Curvatura do perfil do terreno
LS factor (LSF)	Fator de comprimento de inclinação
Relative slope position (RSP)	Posição de cada célula em relação ao topo e base das áreas
Topographic position index (TPI)	Comparação de cada pixel com seu valores vizinhos
Topographic wetness index (TWI)	A viabilidade de acúmulo de água em um espaço
Valley depth (VD)	Distância vertical em relação à vales

Fonte: Adaptado de Silveira *et al.* (2019).

3.3.3. Variáveis climáticas

Os dados meteorológicos de temperatura, precipitação e radiação solar e dados de volume de água presente no solo foram adquiridos da coleção de dados ERA5-Land, disponível para acesso no GEE. Esse conjunto de dados fornece uma série de informações climáticas, como temperatura do solo, temperatura do orvalho, evaporação, além das já citadas e outras, a partir do ano de 1950 até o ano atual para toda a superfície do planeta, na forma matricial, com resolução de 11.132 metros. A Tabela 5 mostra as bandas da coleção ERA5-Land utilizadas na obtenção das variáveis mencionadas.

Apesar da baixa resolução espacial impossibilitar variações de dados entre um talhão e outro, a coleção ERAS-5 Land foi utilizada por apresentar variações climáticas entre os anos

de estudo, possibilitando a associação da produtividade do cafeeiro com os dados climáticos de um período específico do ano.

A partir da coleção de dados, foram selecionadas imagens de cada mês entre o período de agosto de 2016 a julho de 2024 para uma das variáveis. Destas imagens, foi calculada a média da série temporal, formando compósitos mensais de imagens para os municípios de Botelhos e Poços de Caldas. Depois, foi calculada a média entre os pixels de cada talhão alvo (Tabela 5). Esse processo foi repetido para todas as variáveis. No caso da temperatura, além da média, foram calculadas também a temperatura máxima e mínima entre os pixels dos talhões.

Tabela 5 – Variáveis meteorológicas da pesquisa, as bandas da coleção ERA5-Land utilizadas para obtenção destas variáveis, as medidas estatísticas calculadas e o período de obtenção destes dados.

Variável	Medida Calculada	Banda da Coleção ERA5-Land	Período
Precipitação	Média	total_precipitation_sum	Agosto de 2016 a julho 2023
Radiação Solar	Média	surface_solar_radiation_downwards_sum	
Volume de Água no Solo	Média	volumetric_soil_water_layer_3	
Temperatura	Média	temperature_2m	
Temperatura	Máxima	temperature_2m_max	
Temperatura	Mínima	temperature_2m_min	

Fonte: Do autor, 2024.

3.4. Modelo *Random Forest* para Projeção de Safra

Todas as variáveis foram exportadas do GEE em formato tabelar CSV e organizadas em uma planilha Excel juntamente com os dados de produtividade dos talhões. Foram fornecidos para o modelo 47 dados de produtividade, dos anos de 2017 a 2023. O modelo foi construído utilizando 24 variáveis espectrais, 12 variáveis de terreno e 72 variáveis de clima, totalizando 108 variáveis de entrada. O mesmo foi criado no software estatístico RStudio utilizando a linguagem de programação R.

A princípio, foi gerada uma matriz de correlação utilizando coeficiente de correlação de Pearson, a fim de verificar a correlação linear entre as variáveis independentes e a variável dependente.

Posteriormente, foi calculada a diferença percentual de produtividade entre um ano e outro para cada talhão (*produtividade_diff*). Esta informação foi tratada como uma variável no modelo com o objetivo de explicar a dinâmica da produtividade ao longo dos anos, bem como auxiliar o algoritmo a identificar padrões de produção das lavouras e a bienalidade da cultura. Esse método foi empregado a fim de informar ao algoritmo a relação da produtividade do ano anterior com a produtividade atual. Outros estudos sobre modelagem para previsão de produtividades de café também utilizam de séries históricas de produção para calibrar o modelo quanto bienalidade da cultura. No estudo de Carvalho (2003) o objetivo foi avaliar a aplicação da análise harmônica por séries de Fourier de senos e cossenos na previsão de produtividades da cultura do café e, para tanto, foram utilizados dados de produção de 7 município de Minas Gerais dos anos de 1986 a 2000, totalizando uma base de dados de 15 anos. Outros modelos para previsão de safra também utilizam dados da produção do ano anterior para estimar as produções do ano seguinte, como o trabalho de Picini et al. (1999) e de Silva, Lima e Oliveira (2011), onde desenvolveram modelos agrometeorológicos baseados no uso de variáveis climáticas, como déficit hídrico e evapotranspiração da cultura, e o histórico de produções da lavoura como variáveis de entrada.

Para o desenvolvimento do modelo *Random Forest*, foi aplicada a técnica de validação cruzada para treinar o algoritmo usado e potencializar sua capacidade preditiva. A validação cruzada propõe uma partição do conjunto de amostras d , de tamanho n , em duas partes, sendo a primeira amostra de treinamento do algoritmo (dt) e a segunda a amostra de validação do modelo (dv), ambas de tamanho diferentes de acordo com a necessidade do algoritmo para compreensão dos dados. Assim, usa-se a amostra dt para criar uma função preditora e avalia-se o erro esperado submetendo a função à amostra dv (CUNHA, 2019; LAROS; PUENTE-PALACIOS, 2004). Desta forma, foi utilizado a validação cruzada 5-fold cross-validation, onde o conjunto de dados é dividido, de modo aleatório, em 5 subconjuntos (folds). O modelo *Random Forest* foi treinado utilizando 4 folds e o 5º fold foi usado para a validação. Após as 5 interações, o desempenho do modelo é calculado como a média das métricas de avaliação obtidas em cada fold.

No software RStúdio, o algoritmo foi executado utilizando o pacote “randomForest”. Para o *Random Forest*, o parâmetro principal é o número de árvores de decisão a serem construídas (*nTree*). Adotou-se como 100 (PAES, 2022). A projeção para o rendimento do café é a resposta média ajustada de todas as árvores individuais (MARTELLO *et al.*, 2022). O

modelo gerou estimativas para os anos de 2018 a 2023, cujos os valores reais foram informados para treinamento do algoritmo, e projetou a produtividade para o ano de 2024.

Após a criação do modelo, o mesmo foi validado por meio do Erro Absoluto Médio (MAE) (Equação 6), Erro Quadrático Médio (RMSE) (Equação 7) e do Coeficiente de Determinação (R^2) (Equação 8) entre a produtividade medida e a estimada.

$$MAE = \frac{\sum_i^n |\hat{Y}_i - Y_i|}{n} \quad (6)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}}{\frac{\sum_i^n Y_i}{n}} \quad (7)$$

$$R^2 = \frac{\sum (\hat{Y}_i - Y_i)^2}{\sum (Y_i - \bar{Y})^2} \quad (8)$$

onde \hat{Y}_i são os valores previstos, Y_i valores medidos e n número total de observações.

Além disso, foi possível obter as variáveis de maior importância para a construção do modelo. Durante a construção das árvores de decisão a partir dos dados de entrada, cada árvore é treinada com um subconjunto dos dados gerado por amostragem com reposição (*bootstrap*). Em cada nó, algumas variáveis são aleatoriamente selecionadas para direcionar o crescimento da árvore. Destas variáveis, aquela que melhor divide a base de dados é utilizada para realizar a divisão naquele nó (Split). O critério de repartição do nó é RMSE de cada folha. A importância de uma variável é medida pela sua contribuição total na redução da impureza dos nós ao longo de todas as árvores (CARVALHO, 2022; MION, 2019). Na rotina R desenvolvida, a importância das variáveis foi obtida através da função utilizando a função “importance()” do pacote “randomForest”.

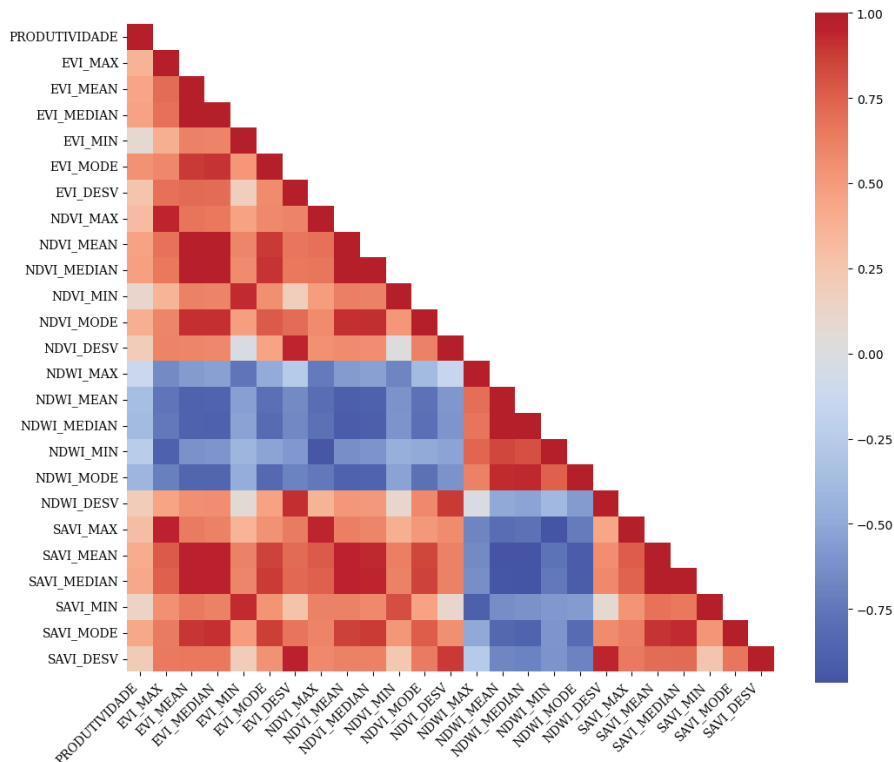
Por fim, para explorar os resultados encontrados, também foram calculados os resíduos gerados pela diferença entre a produtividade real e a produtividade projetada.

4. RESULTADOS E DISCUSSÕES

4.1. Correlação Entre as Variáveis Independentes e a Variável Dependente

A fim de melhorar a visualização dos resultados, a matriz de correlação entre as variáveis independentes e a variável dependente deste estudo foi dividida em três figuras. A Figura 11 mostra a correlação da produtividade com as diferentes as variáveis espectrais. Os valores de correlação oscilaram entre o máximo de 0,545 e mínimo de -0,408. Apesar dos índices NDVI, EVI e SAVI serem todos ligados a biomassa vegetal, as medidas do EVI, em sua maioria, apresentaram altas correlações com a produtividade do café, com valor máximo de 0,545 correspondente a medida de moda deste índice. As medidas do NDVI também apresentaram correlações expressivas, com valor máximo de 0,477 da medida de mediana.

Figura 11 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis espectrais NDVI, SAVI, EVI e NDWI.



Fonte: Do autor (2025).

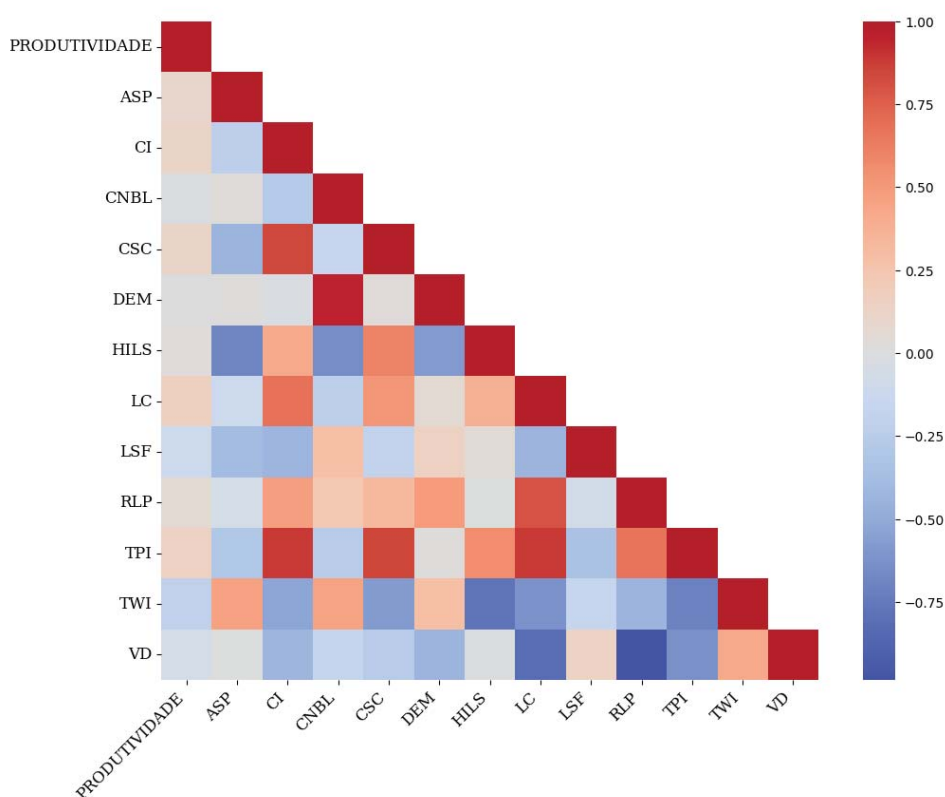
Segundo Huete *et al.* (2002), criadores do EVI, esse índice tem a capacidade de amenizar os efeitos do solo e da atmosfera sobre a resposta espectral da vegetação e, portanto, possui maior sensibilidade às alterações do dossel em áreas de maior biomassa (MOTA *et al.*, 2020). Bernardes *et al.* (2012) investigaram o potencial dos produtos MODIS (MOD 13) para detectar a produtividade bienal do café no período de 2002 a 2009, utilizando como métricas os valores de amplitude, soma, máximo, mínimo e média dos índices NDVI e EVI. Conforme os resultados apresentados, tanto o EVI quanto o NDVI são úteis como indicadores de

bienalidade da produtividade do café, sendo que os valores mínimos de EVI apresentaram maior correlação com a produtividade em todos os anos de sua pesquisa.

O NDWI apresentou significativa correlação inversa com a produtividade. Todas as medidas apresentaram valores entre -0,12 (de máxima) e -0,40 (de moda), exceto o desvio padrão, que teve correlação de 0,2 com a variável de estudo. Na pesquisa de Nogueira, Moreira e Volpato, (2018), o objetivo principal foi avaliar a relação entre produtividade de lavouras de café e os índices de vegetação NDVI, SAVI e NDWI derivada do sensor OLI/Landsat-8 para as safras 2013/2014 e 2014/2015. Segundo os autores, o NDWI obteve maior correlação com a produção do cafeeiro quando comparado ao NDVI e SAVI. Isso ocorre porque a disponibilidade de água é um dos fatores que mais afetam a produtividade desta cultura e, portanto, índices que apontam a quantidade de água no dossel da vegetação podem apresentar melhores resultados de correlação com a produção da lavoura. Ainda no mesmo estudo, o SAVI possuiu menor correlação com a produtividade que os demais índices avaliados, assim como estes resultados.

A Figura 12 demonstra a correlação entre as variáveis de terreno e produtividade do café. Os valores de correlação variam entre -0,2 e 0,15, sendo que nenhuma delas obteve destaque nesta análise. Das correlações entre as variáveis climáticas e a produção (Figura 13), aquelas que obtiveram resultados expressivos foram a radiação solar dos meses de outubro (0,456) e julho (-0,398), a temperatura mínima do mês de março (-0,441) e a precipitação dos meses de dezembro (0,364), fevereiro (0,396) e maio (0,387).

Figura 12 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis de terreno.

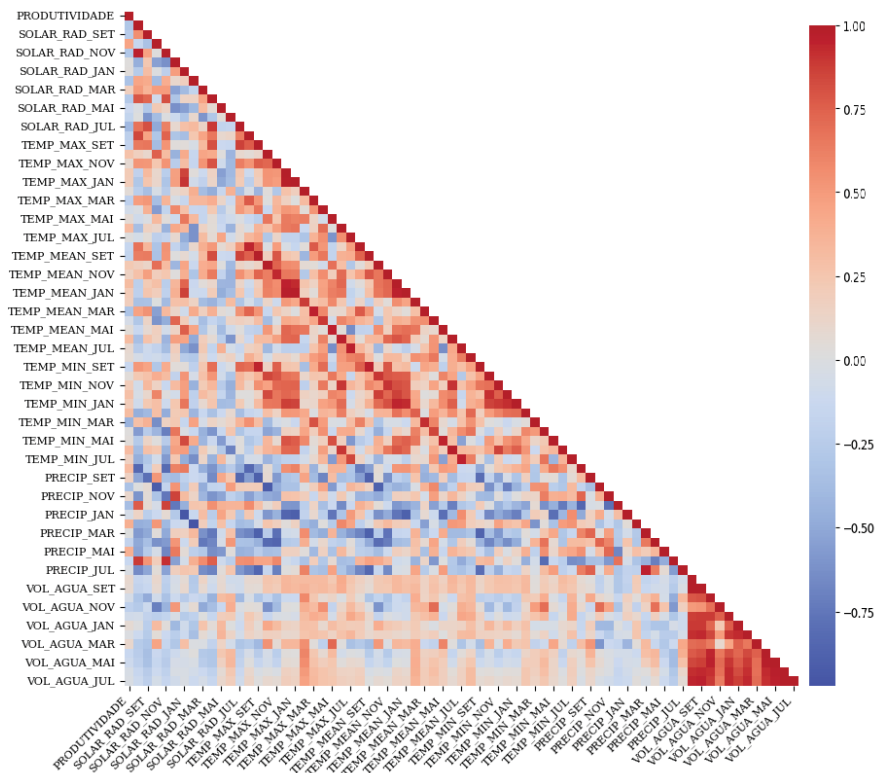


Fonte: Do autor (2025).

As lavouras de café são constantemente expostas a adversidades climáticas que interferem no seu desenvolvimento vegetal, refletindo no potencial produtivo. Dentre estas, a intensidade da radiação solar, as taxas de volume pluviométrico e a sazonalidade das precipitações e da temperatura do ar influenciam na resposta da planta e interferem significativamente na sua capacidade produtiva (MORELLO *et al.*, 2020; RIBEIRO *et al.*, 2017; TRONCHA, 2023).

Já em relação ao volume de água disponível no solo para captação radicular das plantas, na presente pesquisa não houve correlação forte entre esta variável e a variável dependente. Segundo Picini (1998), o déficit hídrico ocupa uma posição de destaque entre os fatores limitantes da produção, pois altera o metabolismo da planta, de modo que a produtividade do cafeeiro é particularmente sensível ao estresse hídrico. Nesse sentido, a variabilidade espacial e temporal dessa variável pode ter diluído possíveis padrões de correlação, exigindo análises mais detalhadas ou o uso de metodologias que capturem melhor sua dinâmica ao longo das fases fenológicas do cafeeiro.

Figura 13 – Matriz de correlação entre a variável produtividade do cafeeiro e as variáveis climáticas.



Fonte: Do autor (2025).

4.2. Importância das Variáveis

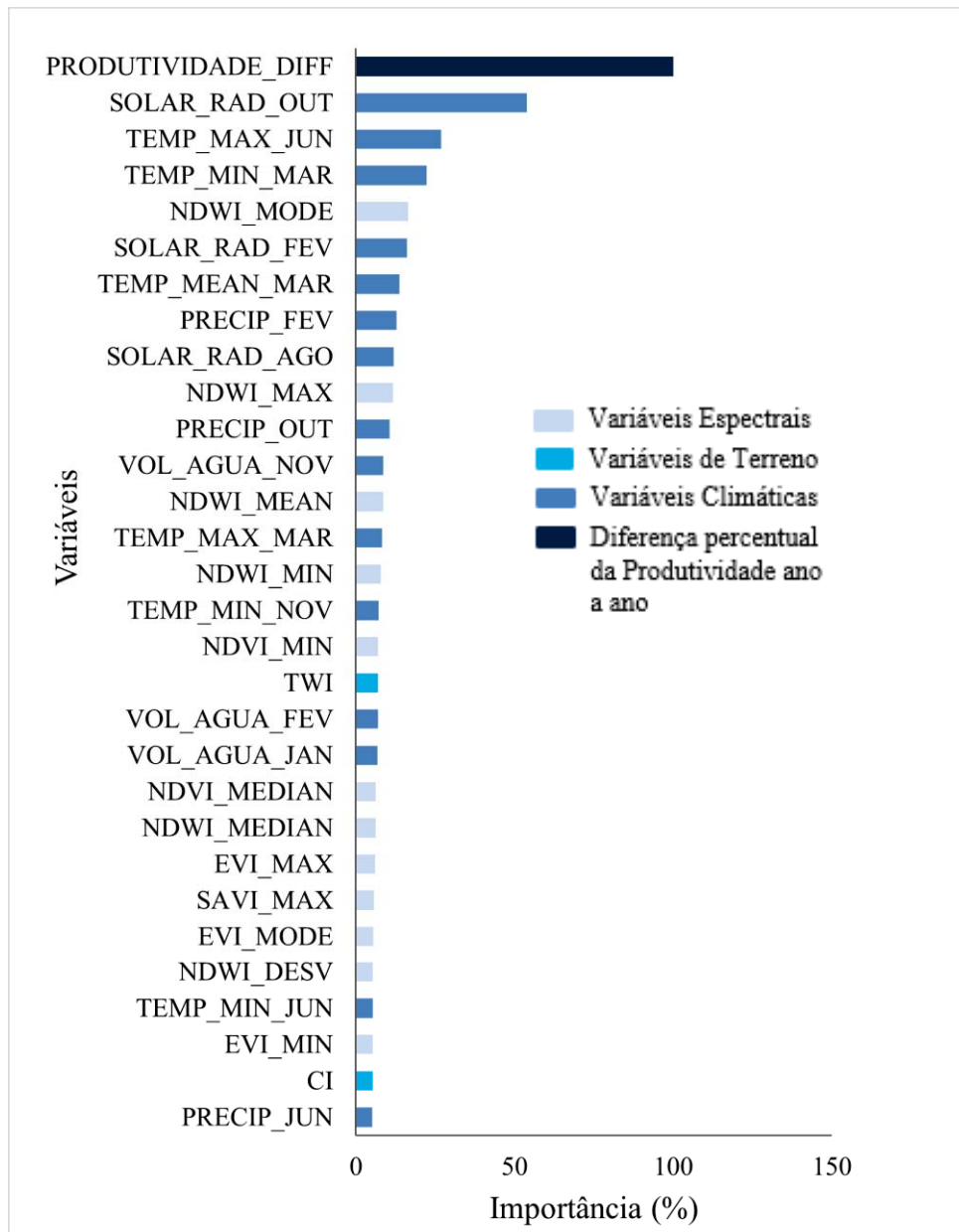
Uma das vantagens dos modelos *Random Forest* é a possibilidade de estimativa da importância relativa das variáveis preditoras para o modelo criado (BHERING *et al.*, 2016; BREIMAN, 2001).

A Figura 14 apresenta as 30 variáveis de maior valor de importância para o modelo desenvolvido. Destas, a variável de maior posição corresponde à diferença de produtividade em porcentagem entre um ano e outro durante o período de estudo. Neste caso, é imprescindível que esta nova abordagem refletiu diretamente na capacidade de predição do modelo *Random Forest*. A diferença percentual entre as produtividades de anos consecutivos está associada a variação temporal do desempenho da cultura. Essa diferença pode decorrer de ações climáticas, adoção de diferentes técnicas de manejo sobre a lavoura, incidência de pragas e doenças em graus de severidade distintos e da bienalidade da cultura. Esta variável pode estar atuando como uma âncora que o modelo utilizou para ajustar a projeção da produtividade futura.

Dentre as 30 variáveis de maior importância identificadas, 15 correspondem as variáveis climáticas, destacando-se o uso majoritariamente da temperatura do ar para a divisão dos nós

das árvores de decisão, independente da medida calculada. Outras 12 são variáveis espectrais extraídas das imagens Sentinel-2, com ênfase no emprego de todas as métricas calculadas para o índice NDWI (média, mínima, máxima, moda, desvio padrão e mediana). Por fim, apenas duas das variáveis de terreno constam entre o ranking apresentado, sendo a Convergence index (CI) e a Topographic wetness index (TWI).

Figura 14 – Ranking das primeiras 30 variáveis mais importantes para o modelo.

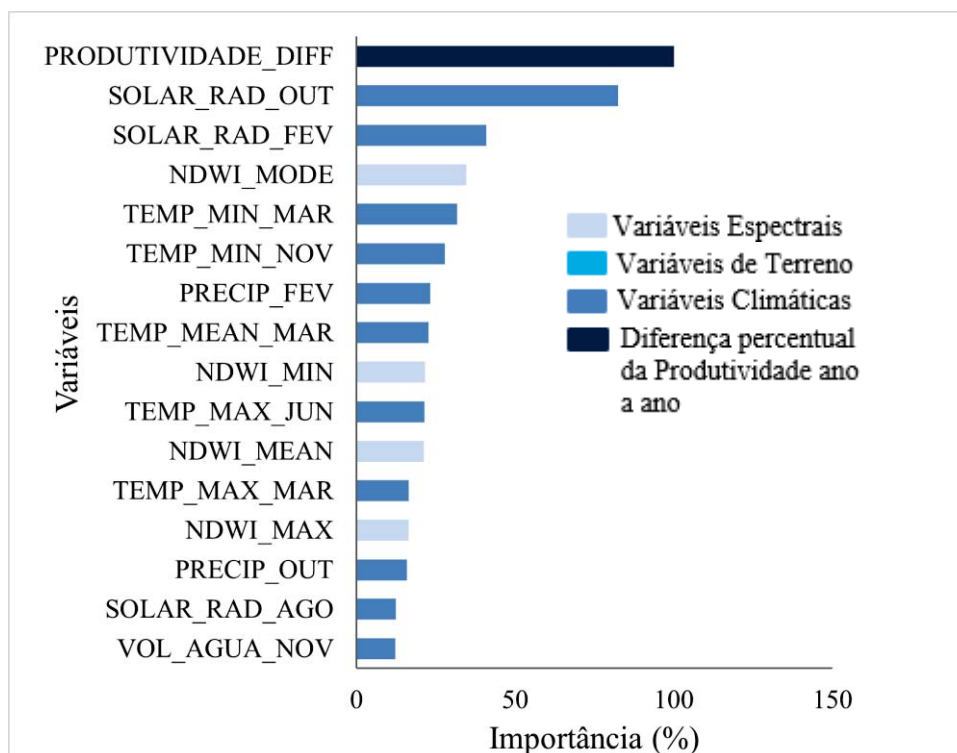


Fonte: Do autor (2025).

Posteriormente, foram retiradas, de forma progressiva, as variáveis com importância inferior a 1, 3, 5, 7, 10 e 15%. O conjunto de variáveis que apresentou menor RMSE e maior

R^2 foi aquele com variáveis cujo o valor de importância era superior a 7%. Este modelo é composto por 16 variáveis, onde 11 delas estão relacionadas ao clima, 4 são medidas do índice NDWI e a última está relacionada a própria produtividade, sendo a diferença percentual da produtividade entre um ano e outro (Figura 15).

Figura 15 - Ranking das variáveis de maior importância quando consideradas apenas aquelas com valores acima de 7%.



Fonte: Do autor (2025).

Nascimento (2019) estudou o desenvolvimento de um modelo *Random Forest* para previsão de safras de café utilizando os índices espectrais SAVI, NDWI e NDVI e os valores de máxima, média e mínima das variáveis de declividade, altitude e aspecto do terreno. Assim como o presente trabalho, as variáveis de menor importância no modelo correspondem aos dados de terreno. Os índices de vegetação obtiveram destaque no ranking de importância das variáveis participantes do estudo, sendo que o NDWI atingiu posições superiores aos demais. Índices de vegetação que incluem bandas SWIR, como o NDWI, auxiliam no monitoramento da produtividade das culturas devido ao seu potencial para detectar propriedades de campo como umidade, vigor da vegetação e solo descoberto (PEÑA-BARRAGÁN *et al.*, 2011).

Já Guimarães (2019) explorou variáveis climáticas, como temperatura, precipitação, evapotranspiração, armazenamento de água no solo, déficit e excedente hídrico, para desenvolver um modelo *Random Forest* de estimativa de produtividade de soja para os

municípios do estado Mato Grosso. O modelo proposto atingiu resultados de 85,88 a 98,48% de acurácia nas previsões de produção realizadas, apontando para o uso significativo e indispensável de fatores meteorológicos para modelos de previsão e de projeção de safra. Dentre as 10 variáveis de maior importância, o algoritmo *Random Forest* utilizou principalmente a temperatura máxima para diminuir a impureza dos nós dentro do modelo, resultado semelhante ao apresentado nesta pesquisa.

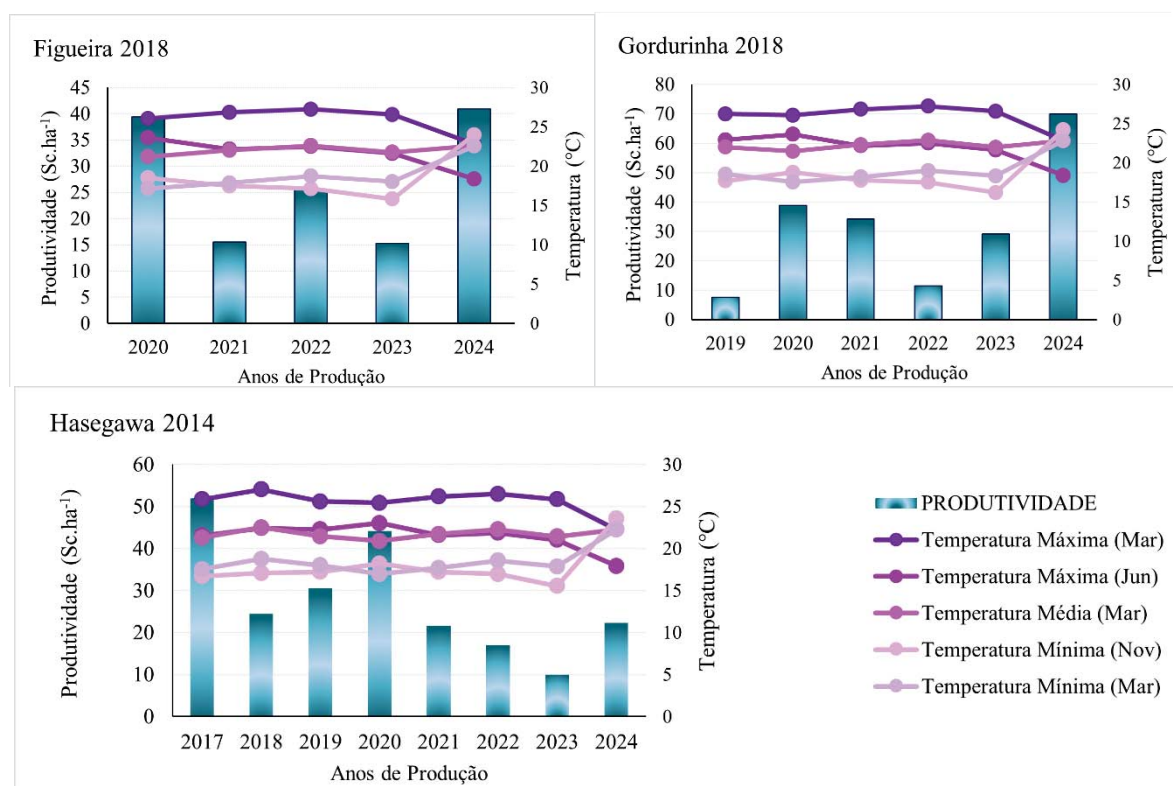
4.2.1. Comparação entre o comportamento das variáveis de maior importância e o comportamento da produtividade nos anos de produção

4.2.1.1. Temperatura do ar

Algumas das variáveis de entrada do modelo estão relacionadas a temperatura ambiente. Foram considerados os valores de temperatura média, mínima e máxima de todos os meses entre o período de agosto de 2016 a julho de 2024. Dentre estas variáveis, cinco obtiveram destaque e estão entre o ranking de variáveis de maior importância para o modelo: a temperatura máxima do mês de março, temperatura máxima do mês de junho, temperatura média de março, a temperatura mínima de novembro e a temperatura mínima de março.

Nos gráficos apresentados na Figura 16 é possível analisar o comportamento das cinco variáveis em relação a produtividade do café. No caso da temperatura mínima, média e máxima do mês de março, é perceptível que sua relação com a produtividade é sempre inversa, de tal forma que, nos anos em que há altas temperaturas nesse mês, a produção de café é menor. De modo contrário, a temperatura máxima do mês de junho e a temperatura mínima do mês de novembro influenciam positivamente no pegamento, formação e amadurecimento dos frutos, uma vez que a produção aumenta quando as temperaturas estão altas e diminui quando estão baixas.

Figura 16 – Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da temperatura média de outubro, da temperatura mínima de novembro e da temperatura mínima de março durante os anos de estudo



Fonte: Do autor (2025).

Segundo Taiz *et al.* (2017), autores de um dos livros fundamentais para área de fisiologia vegetal, a temperatura do ambiente pode influenciar o sistema vegetal de diversas formas. Uma das principais respostas das plantas à temperatura é a abertura e fechamento dos estômatos, também conhecidos como células-guardas, que estão localizadas na fase abaxial da folha e são responsáveis pela respiração celular da planta. Fatores ambientais, como a temperatura do ar, são percebidos pelas células-guarda como sinais e provocam respostas estomáticas diretas. Em casos de ambientes mais quentes, as altas temperaturas reduzem a abertura estomática para conservar água. O fechamento dos estômatos reduz a absorção de CO₂ atmosférico, diminuindo, assim, a taxa de respiração da planta (CID; TEIXEIRA, 2017; TAIZ *et al.*, 2017; VIEIRA *et al.*, 2010).

Em resposta ao fechamento de estômatos e à menor absorção de CO₂, a eficiência da assimilação fotossintética de carbono é limitada, prejudicando a produção de fotoassimilados que garantem a energia da planta consumida nos processos metabólicos necessários para sua sobrevivência (ASSAD *et al.*, 2024; SILVA *et al.*, 2021; TAIZ *et al.*, 2017).

As baixas temperaturas também comprometem o metabolismo e a atividade enzimática da planta. Além disso, a absorção de água decresce quando as raízes são submetidas a baixas temperaturas e há a redução das taxas de respiração do sistema radicular (TAIZ *et al.*, 2017).

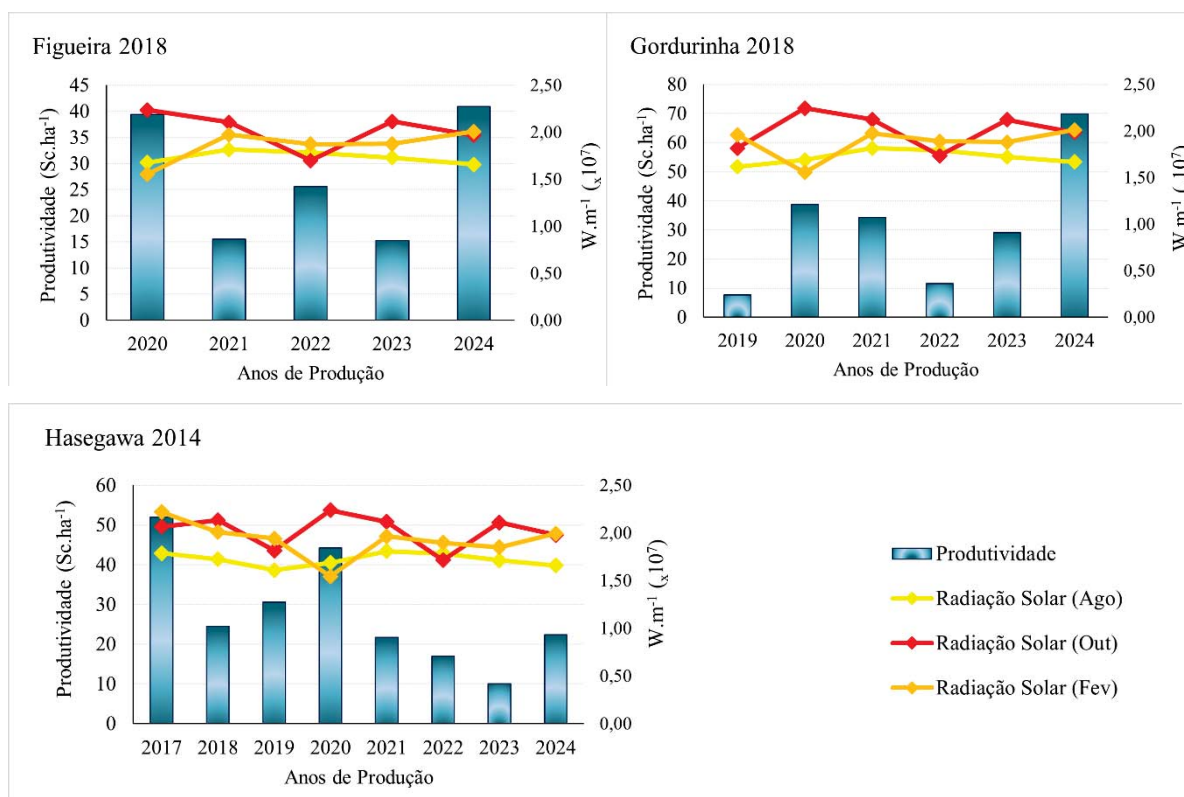
Deste modo, a temperatura ambiente influencia em processos fisiológicos essenciais para a planta, como fotossíntese, respiração e absorção de água e nutrientes. Temperaturas extremas podem comprometer seu desenvolvimento, florescimento e frutificação, interferindo na sua sobrevivência e na perpetuação de sua espécie.

4.2.1.2. Radiação solar na superfície terrestre

A radiação solar é fundamental para o desenvolvimento de qualquer cultura, relacionando-se diretamente com os processos biológicos que regem a vida da planta, quanto a duração (Fotoperíodo), qualidade (comprimento de onda) e intensidade (energia para calor e fotossíntese). Deste modo, os raios solares favorecem condições adequadas para que a planta exteriorize seu potencial produtivo (SOUZA; SMIDERLE, 2024).

Na Figura 17 é possível verificar o comportamento da radiação solar na região de estudo em detrimento a produtividade alcançada nas lavouras. Pode-se notar que a radiação solar do mês de agosto é mais constante que as demais, com poucos picos de variação, ao contrário da radiação que chega à superfície terrestre nos meses de outubro e fevereiro, que demonstraram altas flutuações durante o período estudado. Apesar da radiação solar ser um fator importante para o desenvolvimento da planta e essencial para altas produções, os gráficos dos talhões apontam que a relação entre ela e a produtividade de café não é alta. No entanto, a radiação solar pode estar atuando de forma indireta, interagindo com outros fatores, como a temperatura e a disponibilidade hídrica, que afetam o desenvolvimento do café e refletem na sua produção.

Figura 17 – Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da radiação solar dos meses de agosto, outubro, fevereiro e junho.



Fonte: Do autor, 2024.

O estudo realizado por Troncha (2023) teve como objetivo identificar os efeitos da intensidade luminosa na morfofisiologia e na produtividade de plantas de cafeeiro. O trabalho foi conduzido em campo em plantas de cafeeiro Catuaí com seis anos de idade em Ipameri, de modo que foram coletadas informações de lavouras implantadas nas direções leste e oeste e nas posições superior e inferior na copa das plantas. A pesquisa mostrou que o cafeeiro apresenta elevada plasticidade fenotípica em relação à variação da energia luminosa e tem a capacidade de se adaptar morfológicamente as variações de radiação solar. Porém, a variação na disponibilidade de radiação interferiu decisivamente na produtividade, de modo que lavouras plantadas na direção oeste apresentaram maior produção quando comparadas as lavouras localizadas na fase leste, devido a maior incidência de radiação solar.

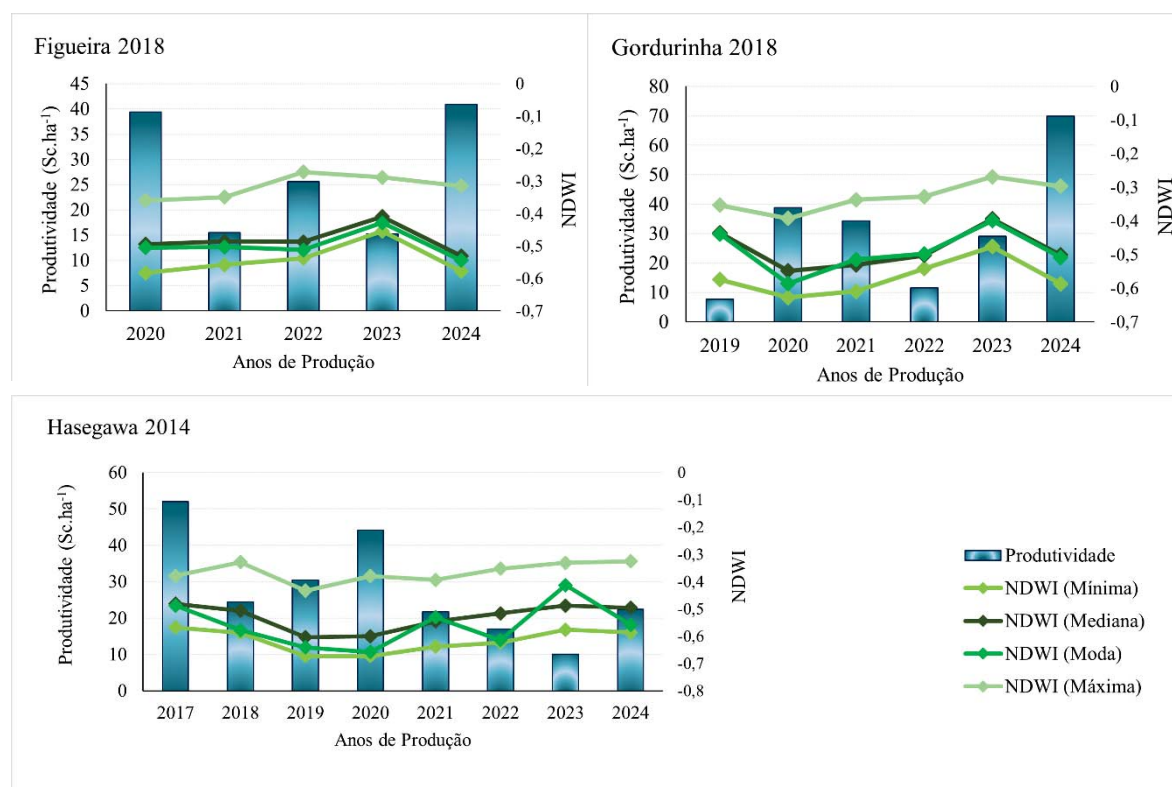
Já a pesquisa desenvolvida por Amorin *et al.* (2024) objetivou avaliar a produtividade e o rendimento do cafeeiro arábica em cultivo sob arborização com mogno brasileiro e a pleno sol. Os autores afirmam que, no cultivo de café em pleno sol, o amadurecimento dos grãos foi acelerado devido ao excesso de radiação solar e temperatura, potenciando uma maior produtividade, porém levando a uma maior incidência de grãos subdesenvolvidos e com baixa

qualidade. Em contraste, o ambiente sombreado proporcionou condições mais favoráveis para o desenvolvimento e maturação tardia dos frutos, resultando em grãos maiores e mais pesados, o que aumenta o volume de café beneficiado. Apesar da adoção de sombreamento em cafezais não resultar em aumento da produtividade, este ambiente proporcionou maior rendimento para a lavoura.

4.2.1.3. NDWI – Presença de água no dossel

O índice NDWI é utilizado para medir a umidade presente no dossel da vegetação, uma vez que é sensível às variações no conteúdo de água na cobertura vegetal. Nos gráficos apresentados na Figura 18 é possível observar que os valores de mínima, moda, máxima e mediana do NDWI possuem uma correlação inversa com a produtividade, de modo que menores taxas de umidade na planta indicam maior produtividade do cafeeiro.

Figura 18 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento dos valores de mínimo, moda e mediana do NDWI.



Fonte: Do autor (2025).

Segundo Nogueira, Moreira e Volpato (2018) o NDWI obtido durante o estágio de dormência das gemas florais do cafeeiro (julho e agosto) apresenta alta correlação com a

produtividade do próximo ano. A capacidade de floração do cafeeiro responde a flutuações na temperatura do ar e ao suprimento de água à planta, de modo que a sincronização do florescimento em cafeeiros tem sido associada a submissão da planta à déficits hídricos nesse estágio fenológico. Mesmo quando irrigado, o cafeeiro possui a capacidade de sujeitar-se a déficits hídricos internos de água para quebrar a dormência das gemas florais, levando ao florescimento após a aplicação da água por irrigação ou chuva (GONZALEZ-ROBAINA; CISNEROS-ZAYAS; MONTILLA, 2017; SILVA et al., 2009). Magalhães e Angelocci (1976) afirmam ser necessário um déficit de até -1,2 Mpa nas folhas do cafeeiro para promover o processo de floração em resposta à irrigação.

Esse período de estresse hídrico durante os meses de julho e agosto, anteriores a florada do café, favorece a sincronização das floradas, permitindo a uniformidade na maturação dos frutos e a alta produção de grãos (SAKAI et al., 2013; SILVA; MAZZAFERA, 2008; SILVA et al., 2009).

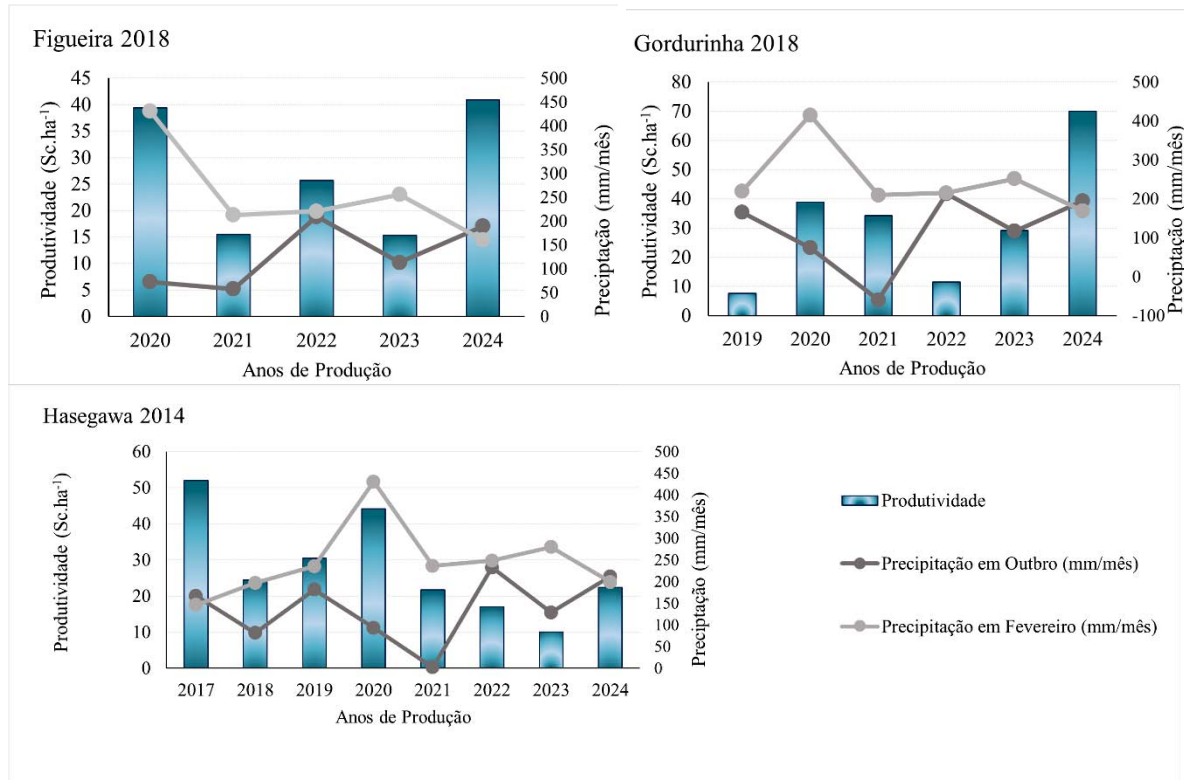
4.2.1.4. Volume de água no Solo e Precipitação

Como toda cultura agrícola, o café depende da incidência de chuvas e de disponibilidade hídrica no solo para completar seu ciclo fenológico (LINHARES, 2022). No modelo desenvolvido, a precipitação do mês de outubro, a precipitação do mês de fevereiro e o volume de água no solo durante o mês de novembro participam das variáveis de maior importância. Pode-se observar que as três variáveis em destaque estão entre o período de maior pluviosidade na região de estudo, que começa em outubro e finaliza em março do ano seguinte, e coincidem com os estágios de vegetação e formação das gemas florais, floração e granação dos frutos, épocas onde a disponibilidade hídrica influencia no crescimento, desenvolvimento e vigor do pé de café, no pegamento da florada e na formação e maturação dos frutos e das sementes de café (CAMARGO; CAMARGO, 2001).

Verifica-se que, em geral, tanto o fator precipitação quanto o volume de água disponível no solo apresentam correlação positiva com a produtividade (Figura 19 e 20), de modo que ambas as variáveis favorecem a produção das lavouras cafeeiras. Segundo Meireles et al. (2009), geralmente, o cafeeiro arábica é afetado, nas suas diversas fases fenológicas, pelas condições ambientais, em especial pela variação fotoperiódica e, principalmente, pela distribuição pluviométrica e temperatura do ar, que interferem não apenas na fenologia, mas também na produtividade e na qualidade da bebida. Logo, adversidades hídricas durante os

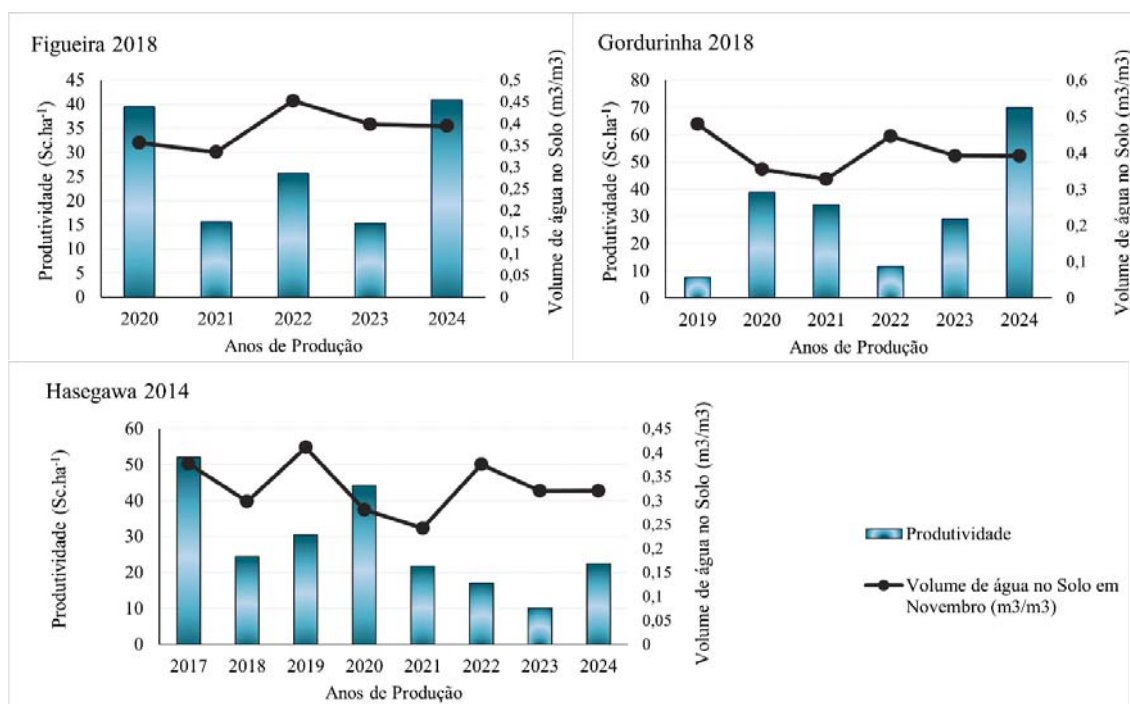
períodos fenológicos citados podem resultar na redução drástica da produção do cafeeiro (SERAFIM et al., 2013).

Figura 19 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento das variáveis precipitação do mês de outubro e precipitação do mês de fevereiro.



Fonte: Do autor (2025).

Figura 20 - Gráficos de produtividade dos talhões Figueira 2018, Gordurinha 2018 e Hasegawa 2014 e do comportamento da variável volume de água disponível no solo.



Fonte: Do autor (2025)

4.3. Análise de desempenho do modelo preditivo

A capacidade preditiva do modelo criado foi avaliada através do MAE, do R^2 e Erro RMSE. A Tabela 6 mostra as medidas de desempenho para os modelos gerados considerando as variáveis com importância acima de 0, 1, 3, 5, 7, 10 e 15%.

Tabela 6 – Medidas de desempenho MAE, R^2 e RMSE para os modelos gerados.

Valor de importância das variáveis (%)	MAE (sc.ha ⁻¹)	R^2 (sc.ha ⁻¹)	RMSE (sc.ha ⁻¹)
>0	2,7188	0,9227	13,5648
>1	2,6215	0,9278	12,6695
>3	2,5662	0,9309	12,1242
>5	2,4789	0,9326	11,8301
>7	2,6423	0,9354	11,3413
>10	2,7807	0,9300	12,2904
>15	2,7495	0,9316	12,0124

Fonte: Do autor (2025).

O maior MAE é encontrado quando são utilizadas apenas as variáveis de importância maior que 15%, de 2,75 sc.ha⁻¹, e o maior RMSE pode ser observado no modelo que considera

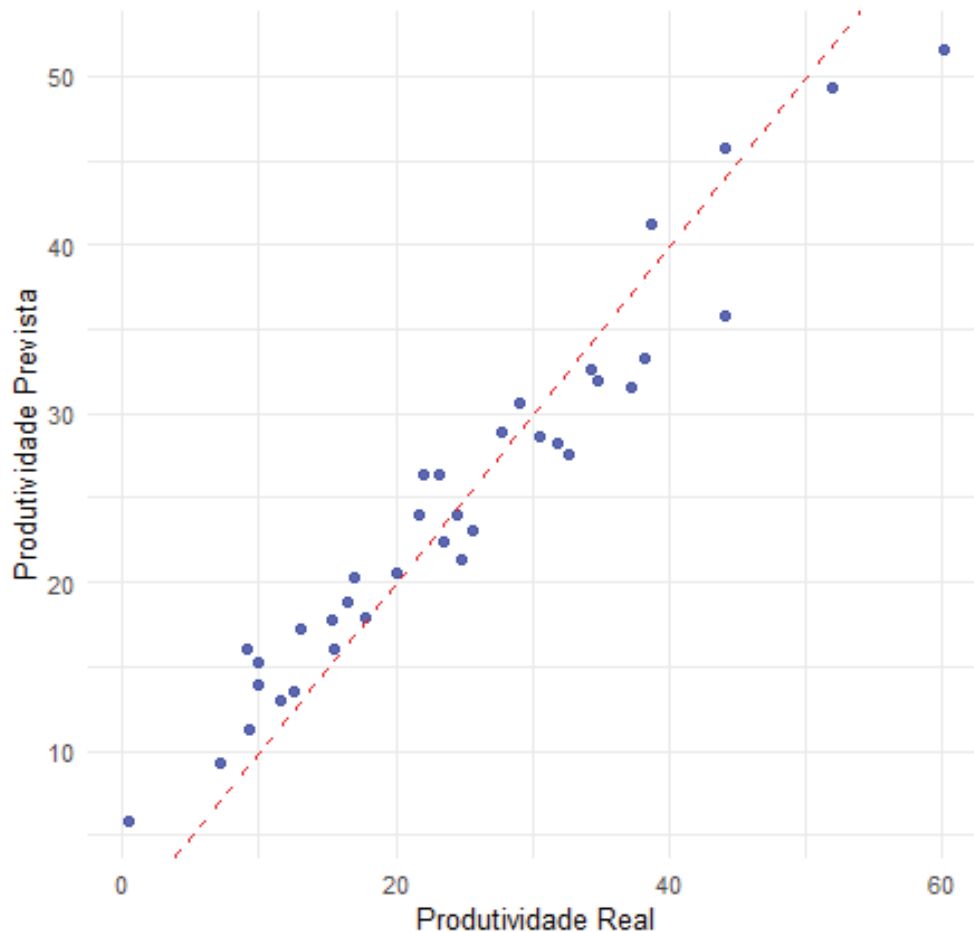
todas as variáveis (importância acima de 0%), de 13,56 sc.ha⁻¹. Os menores valores para ambas as mediadas são de 2,48 sc.ha⁻¹ para o MAE e de 11,34 sc.ha⁻¹ para o RMSE, quando são empregadas as variáveis com importância acima de 7%. O R² aumenta conforme são descartadas as variáveis de menor importância e volta a aumentar quando as variáveis restantes já não são suficientes para estimar a produtividade do ano seguinte com maior precisão, sendo o melhor valor encontrado quando as variáveis com importância acima de 7% constroem o modelo, indicando que as variáveis independentes explicam 93,3% da variável dependente. A falta de informações sobre outros fatores externos, como a ausência ou presença de irrigação artificial, a correção e adubações do solo e adubações foliares, dentre outros fatores, pode ter influenciado no modelo de modo que o mesmo não pode detectar tais cenários, dado que estas variáveis não foram consideradas na análise inicialmente (GUIMARÃES, 2019).

Abreu Junior *et al.* (2022) avaliaram o potencial do uso de imagens multiespectrais de média resolução espacial e modelos baseados em aprendizado de máquina na predição da produtividade da cultura do café. Para tanto, os autores utilizaram imagens Sentinel-2 de onde foram obtidos os índices de vegetação relacionados a clorofila (NDRE, CI-RE, TCARI, CVI e CI-G), biomassa (NDVI, RVI e SAVI), nitrogênio (GNDVI) e índice de área foliar (MCARI). O modelo *Random Forest* foi configurado em 100 iterações e avaliado através das medidas RMSE, MAPE% e R². De acordo com os resultados, o menor valor de RMSE foi de 27% e de R² foi de 81%, valores obtidos na fase de teste do modelo.

A Figura 21 mostra a performance do algoritmo *Random Forest* na projeção da produtividade em relação aos valores reais. Pode-se observar que, quando há produtividades mais elevadas, o modelo apresentou desvios maiores, refletindo limitações na capacidade de estimar valores extremos. O estudo de Sajid *et al.* (2022) objetivou avaliar o desempenho dos modelos de *machine learning* LASSO regression, *Random Forest*, XGBoost, LightGBM e regressão linear e combina-los para comparar a sua eficiência individual e unificada a fim de identificar o modelo de melhor desempenho para previsão do rendimento do milho na região denominada Cinturão do Milho, nos Estados Unidos. Contrapondo os resultados da presente pesquisa, os modelos desenvolvidos apresentaram baixa correlação entre o rendimento previsto e o rendimento real em áreas cuja produção de milho foi baixa, enquanto que houve correlação mais forte entre ambos os rendimentos em regiões onde a produção foi alta. Deste modo, as

previsões foram mais confiáveis para locais com rendimentos normais a altos em comparação com regiões com rendimentos mais baixos.

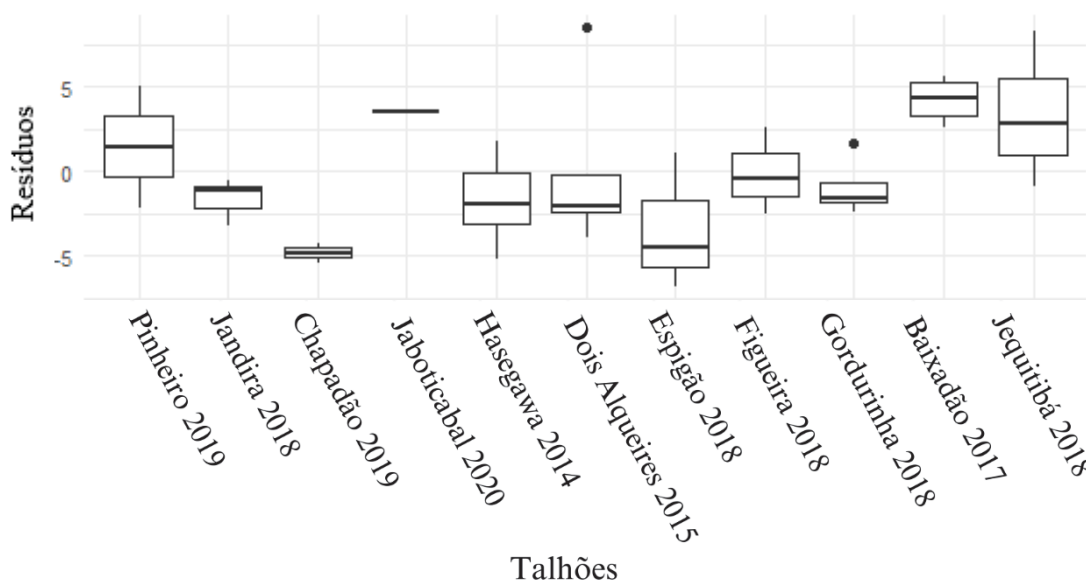
Figura 21 – Relação entre a produtividade real e a produtividade estimada



Fonte: Do autor (2025).

O gráfico boxplot (Figura 22) mostra a distribuição dos resíduos do modelo *Random Forest* em relação à produtividade real e projetada nos diferentes talhões de estudo. Observa-se que os resíduos não são uniformes entre os talhões. Os talhões Jandira 2018 e Jaboticabal 2020 apresentaram menor dispersão entre os resíduos do modelo, enquanto que os talhões como Baixadão 2017 e Jequitibá 2018 mostram uma amplitude maior, sugerindo que as variáveis independentes avaliadas no modelo não foram suficientes para explicar a produtividade nesses casos. O talhão Baixadão 2017 também obteve resíduos predominantemente positivos, indicando a subestimação da variável dependente. Além disso, nos talhões Dois Alqueires 2015, Gordurinha 2018 e Jequitibá 2018 há a presença de valores extremos, conhecidos como outliers, indicando discrepância entre as produtividades projetadas e as reais.

Figura 22 – Resíduos do modelo *Random Forest* para projeção de safra.



Fonte: Do autor (2025)

4.4. Projeção de Safra dos anos de 2018 a 2024

A Tabela 7 apresenta a produtividade real colhida nos talhões de estudo e a produtividade projetada pelo modelo *Random Forest* para os anos de 2018 a 2023. As maiores estimativas obtidas são de 51,60 e 49,34 sc.ha⁻¹, correspondentes aos talhões Dois Alqueires 2015 e Baixadão 2017, respectivamente, ambas do ano de 2020. Já as menores produtividades estimadas são do talhão Chapadão 2019, de 5,89 sc.ha⁻¹ em 2022, e do talhão Pinheiro 2019, de 9,33 sc.ha⁻¹, também em 2022.

Em geral, as projeções do modelo foram satisfatórias. As maiores diferenças entre a produtividade real e a projetada foram de 8,50 sc.ha⁻¹ estimadas a mais que a quantidade colhida pela fazenda, e de 6,85 sc.ha⁻¹ estimadas a menos produtividade real, correspondentes aos talhões Dois Alqueires 2015, no ano de 2020, e Espigão 2018, no ano de 2021. Pode-se observar que, na maioria das projeções, o modelo subestimou a produtividade das lavouras.

O modelo de regressão linear desenvolvido por Aparecido et al. (2022) teve como objetivo desenvolver modelos de regressão linear múltipla em função da precipitação (mm), temperatura máxima e mínima do ar (°C), umidade relativa (%), velocidade do vento (ms⁻¹), radiação solar, balanço hídrico e evapotranspiração de referência para estimar a produtividade média de café em alguns municípios do Paraná. Os modelos criados foram capazes de prever este valor cerca de dois a três meses antes da colheita. Os autores apontam que as estimas variaram entre a máxima de 27,7 e a mínima de 16,1 sc.ha⁻¹, muito próximo dos valores reais

colhidos, que variaram entre 16,9 a 32,5 sc.ha⁻¹. O desvio médio entre a produtividade projetada e a real foi de 3,6 sc.ha⁻¹. Além disso, os melhores valores RMSE e R²_{adj} foram de 0,78 sc.ha⁻¹ e 89%, respectivamente.

Tabela 7 – Produtividade real e prevista pelo modelo (em Sc ha⁻¹) para os talhões e os anos de estudo e a diferença entre ambas.

Talhão	Produtividade (Sc ha ⁻¹)			
	Ano	Real	Projetada	Diferença
Baixadão 2017	2020	51,91	49,34	2,57
	2021	37,17	31,53	5,63
	2022	24,82	21,33	3,48
	2023	32,73	27,52	5,21
Chapadão 2019	2022	0,50	5,89	-5,38
	2023	13,00	17,22	-4,22
Dois Alqueires 2015	2019	9,20	11,27	-2,07
	2020	60,10	51,60	8,50
	2021	10,00	13,97	-3,97
	2022	16,41	18,77	-2,36
Espigão 2018	2023	17,73	17,95	-0,22
	2021	9,17	16,02	-6,84
	2022	23,47	22,41	1,05
Figueira 2018	2023	21,95	26,43	-4,48
	2021	15,50	15,97	-0,47
	2022	25,61	23,00	2,60
Gordurinha 2018	2023	15,27	17,73	-2,47
	2020	38,77	41,21	-2,44
	2021	34,26	32,65	1,61
	2022	11,50	12,98	-1,48
Hasegawa 2014	2023	29,07	30,68	-1,61
	2018	24,44	23,97	0,47
	2019	30,50	28,66	1,84
	2020	44,17	45,75	-1,58
	2021	21,66	24,04	-2,38
Jaboticabal 2020	2022	16,99	20,26	-3,27
	2023	10,00	15,19	-5,19
Jandira 2018	2023	31,81	28,25	3,56
	2021	23,13	26,33	-3,20
	2022	20,00	20,58	-0,58
Jequitibá 2018	2023	27,74	28,84	-1,10
	2021	34,84	32,00	2,84
	2022	12,60	13,50	-0,90
	2023	44,13	35,85	8,28

Continuação...

Talhão	Produtividade (Sc ha ⁻¹)			
	Ano	Real	Projetada	Diferença
Pinheiro 2019	2022	7,16	9,33	-2,17
	2023	38,30	33,21	5,09

Fonte: Do autor (2025).

As projeções de produtividade do ano de 2024 foram marcadas por grandes diferenças entre a produtividade real e os resultados do modelo. As maiores desigualdades podem ser observadas nos talhões Gordurinha 2018, Baixadão 2017 e Espigão 2018 onde as produtividades projetadas foram 42,51 sc.ha⁻¹, 40,10 sc.ha⁻¹ e 34,75 sc.ha⁻¹ a menos que os dados reais (Tabela 8). Estes talhões apresentaram produções surpreendentemente altas em 2024 (de 68,51, 58,98 e 69,90 sc.ha⁻¹, respectivamente), superiores as produções dos últimos anos.

As maiores projeções correspondem aos talhões Figueira 2018 e Dois Alqueires 2015, de 30,70 e 29,16 sc.ha⁻¹, e diferença de 10,19 e 5,44 sc.ha⁻¹ em relação a produtividade real. Já as menores estimativas são encontradas nos talhões Espigão 2018 e Chapadão 2019, de 24,23 e 21,06 sc.ha⁻¹, e diferença de 34,75 a 6,05 sc.ha⁻¹ da produtividade real (Tabela 8).

Tabela 8 – Produtividades reais e projetadas para o ano de 2024 por talhão, em Sc ha⁻¹.

Talhão	Produtividade (Sc ha ⁻¹)		
	Real	Projetada	Diferença
Baixadão 2017	68.51	28,41	40,10
Chapadão 2019	15.01	21,06	-6,05
Dois Alqueires 2015	34.12	28,49	5,63
Espigão 2018	58.98	24,23	34,75
Figueira 2018	40.89	30,70	10,19
Gordurinha 2018	69.90	27,29	42,61
Hasegawa 2014	22.35	25,54	-3,19
Jaboticabal 2020	47.68	25,24	22,44
Jandira 2018	34.60	29,16	5,44
Jequitibá 2018	15.25	25,55	-10,30
Pinheiro 2019	13.03	25,92	-12,89

Fonte: Do autor, 2024.

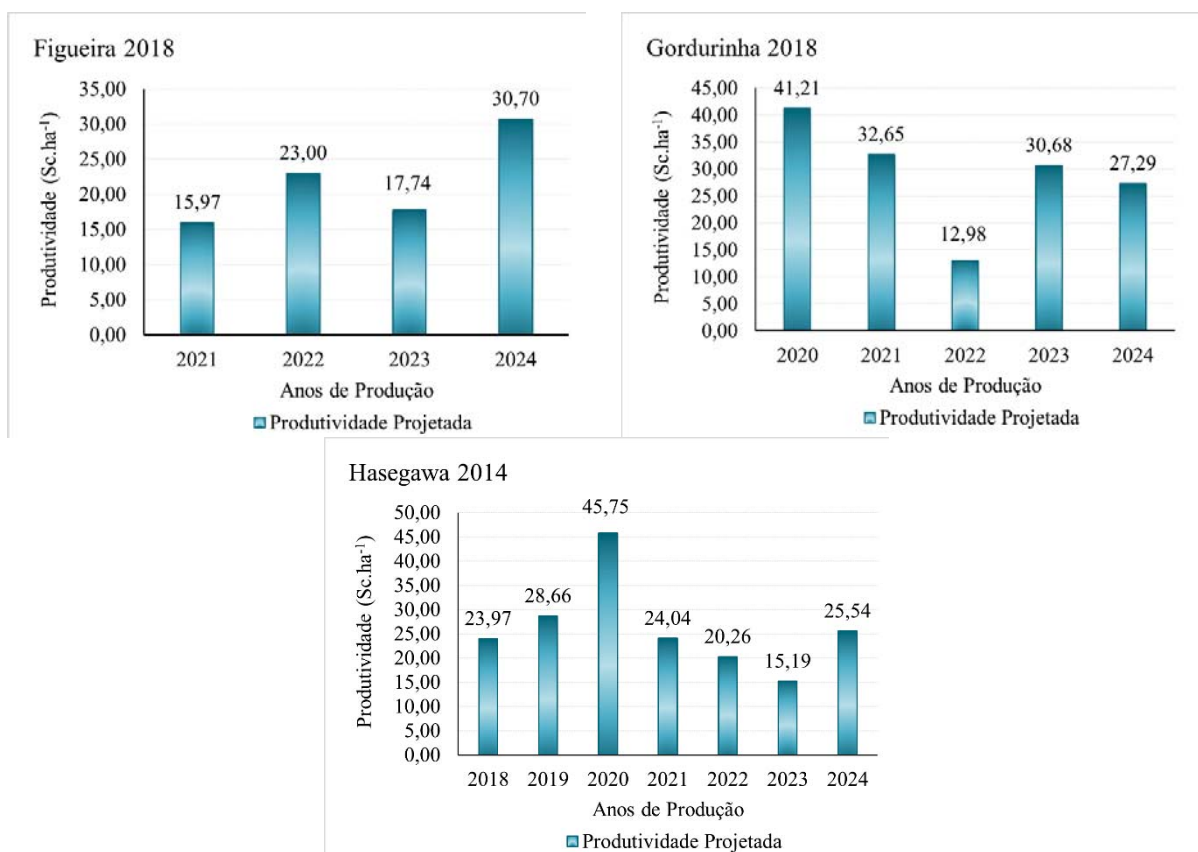
Segundo o último boletim emitido pela CONAB sobre a safra de café de 2024, a produtividade média de café arábica alcançada no estado de Minas Gerais foi de 26 Sc ha⁻¹. Apesar do bom rendimento, a quantidade real de café colhido foi inferior as expectativas da empresa no começo do ano, que não previa uma queda na produção devido, principalmente, às

irregularidades climáticas em fases importantes do ciclo fenológico da cultura. Períodos de estiagem severa, ondas de calor intermitentes em alguns momentos do ciclo, e irregularidade na distribuição e no volume das chuvas são apontados como fatores cruciais na redução, tanto do rendimento médio, como também da qualidade dos grãos obtidos (CONAB, 2024).

O LSPA (Levantamento Sistemático da Produção Agrícola), realizado pelo IBGE, também trouxe dados semelhantes sobre a produtividade do ano agrícola 2023/24, com rendimento médio para o mesmo estado de 26,9 Sc ha⁻¹ (IBGE, 2024b). Neste ano, o único talhão que obteve valores semelhantes às produtividades apresentadas por ambas as organizações foi o Hasegawa 2014 (22.35 Sc ha⁻¹).

Embora a diferença entre a projeção de produtividade do modelo e a produtividade real do ano de 2024 ser alta, os resultados do modelo seguem a lógica das projeções realizadas nos anos anteriores. A Figura 23 mostra o gráfico das produtividades projetadas para 3 talhões de estudo. Pode-se observar que a projeção de 2024 segue a bienalidade da cultura, que foi capturada pelo modelo *Random Forest*, ou seja, quando a projeção de produtividade de 2023 é baixa, a projeção de 2024 é alta, e o inverso também é verdadeiro.

Figura 23 – Projeção de produtividade ao longo dos anos dos talhões Figueira 2018, Gordurinha 2018 e Jequitibá 2018.



Fonte: Do autor (2025).

4.5. Limitações do Modelo

O modelo criado para projeção de produtividade de café enfrentou algumas limitações que refletem diretamente no seu desempenho. A principal delas é a restrita base de dados de produtividade real que foi utilizada nos processos de treinamento e teste do algoritmo *Random Forest*. Apesar do modelo ter reconhecido a bienalidade da cultura e os padrões de produção ao longo dos anos, este foi incapaz de captar e prever produtividades extremas, como os resultados discutidos para o ano de 2024. Series históricas com algumas décadas de dados coletados podem auxiliar o modelo a encontrar valores atípicos e associá-los aos fatores climatológicos, espectrais, de manejo e outros que influenciam na produção de café.

Ao longo do texto científico, foi apresentada a importância do clima para as produções elevadas das lavouras cafeeiras. Neste sentido, a localização próxima entre um talhão e outro, e a baixa resolução espacial da coleção de dados Era5-Land (de 11.132 m) não permitiu que houvessem mudanças significantes de clima entre os talhões. Deste modo, a relação entre a produtividade e os dados meteorológicos ficou limitada entre a pequena variação de valores.

Por fim, pôde-se verificar nesta pesquisa que a produtividade do café apresenta altas variações de um ano a outro, ignorando, em alguns casos, a bienalidade genética da cultura. Essa limitação pode dificultar o reconhecimento do padrão bianual de produção e confundir o modelo, prejudicando a previsão final da produtividade nas próximas safras.

Em estudos futuros, pode-se verificar os resultados do modelo ao adicionar aspectos da cultura e do manejo da lavoura como variáveis de entrada, como o material genético implantado nas lavouras, a incidência de pragas, doenças e plantas daninhas, a forma de adubação, a presença ou ausência de irrigação artificial e o manejo geral da propriedade rural, que são fatores suma importância para a produção de cafés (MOREIRA *et al.*, 2019) e podem aumentar a eficiência do modelo na predição de produtividades mais próximas ao real.

CONCLUSÃO

Dentre as variáveis climáticas, de terreno e espectrais que foram utilizadas como dados de entrada para a construção do modelo, aquelas que obtiveram maiores correlações com a produtividade de café foram os IVs EVI, NDWI e NDVI, a radiação solar, temperatura mínima e precipitação. As variáveis de terreno obtiveram baixa correlação com a variável dependente.

As variáveis de maior importância para a divisão dos nós das árvores de decisão construídas pelo algoritmo *Random Forest* foram a diferença percentual da produtividade entre um ano e outro, a temperatura mínima dos meses de março e novembro, a temperatura média do mês de outubro, a radiação solar média dos meses de agosto, outubro, fevereiro e junho, o índice Convergence index (CI) e as medidas de moda, mediana e mínima do índice NDWI do mês de agosto.

O modelo obteve melhor desempenho quando foram consideradas apenas as variáveis de importância acima de 200 (citadas), apresentando R^2 de 93,74, MAE de 2,49 e RMSE de 10,99. Em geral, as estimativas de produtividade geradas pelo modelo de projeção de safra para os anos de 2018 a 2023 foram satisfatórias. Para o ano de 2024, é possível observar que o modelo seguiu a tendência de bienalidade da cultura, porém não conseguiu prever produtividades extremas.

Uma das principais limitações do modelo foi a pequena base de dados de produtividades utilizadas para treinamento e teste do modelo de projeção de safras criado. Series históricas com algumas décadas de dados coletados podem auxiliar o modelo a encontrar valores atípicos e estimar, com maior precisão, valores de produtividade extremos.

REFERÊNCIAS

ABREU JUNIOR, C. A. M de. *et al.*. Estimating Coffee Plant Yield Based on Multispectral Images and Machine Learning Models. **Agronomy**, v. 12, n. 12, 2022.

ABREU, U. G. P.; THOLON, P.; LIMA, H. P. de. Classificação do desempenho dos rebanhos de seleção nelore por meio de aprendizado de máquina. *In: Congresso Brasileiro de Agroinformática, 14., 2023, Natal. Anais [...]*. Natal: Universidade Federal do Rio Grande do Norte, 2023.

ALASCA SATELLITE FACILITY – ASF. Alos Palsar. Disponível em: <<https://asf.alaska.edu/datasets/daac/alos-palsar/>>. Acesso em: 15 abr. 2024.

ALBA, E. *et al.* Comparação entre algoritmos de aprendizado de máquina para a identificação de floresta tropical sazonalmente seca. **Anuário do Instituto de Geociências**, vol. 45, p. 1-10, 2022.

ALMEIDA, G. S. de. *et al.* Monitoramento de interações solo-planta e produtividade de milho por índices de vegetação satélite, condutividade elétrica do solo e zonas de manejo. *In: XX Simpósio Brasileiro de Sensoriamento Remoto, 2023, Florianópolis. Anais [...]*. São José dos Campos: INPE, 2023.

ALVES, M. H. D. *et al.* Aplicação dos índices de vegetação NDVI, SAVI e NDWI na caracterização da cobertura vegetativa do plantio de açaí (*Euterpe oleracea Mart.*) em regiões Amazônicas. *In: XLIX Congresso Brasileiro de Engenharia Agrícola, 2020. Anais [...].* CONBEA, SBEA, 2020.

AMARAL, F. H. F. do; QUEVEDO, R. P.; PIROLI, E. L. Floresta aleatória aplicada ao mapeamento de suscetibilidade a movimento de massa na bacia hidrográfica do Rio Palena, Chile. *In: XX Simpósio Brasileiro de Sensoriamento Remoto, 2023, Florianópolis. Anais [...].* São José dos Campos: INPE, 2023.

AMARO, R. P. **Estimativa de produtividade da cana de açúcar a partir de imagens do satélite Sentinel-2A e o algoritmo de aprendizagem de máquina Random Forest.** 2023. 42 f. (Dissertação em Engenharia de Sistemas Agrícolas) - Programa de Pós-Graduação, Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2023.

AMORIN, M. A. *et al.*. Análise da produtividade e rendimento do cafeeiro arábica com e sem arborização. *In: Encontro nacional do café, 15., 2024, Planalto da Conquista. Anais [...].* Planalto da Conquista: Fazenda Vidigal, 2024.

ANDRADE, T. G. *et al.* Soybean yield prediction using remote sensing in southwestern Piauí state, Brazil. **Revista Caatinga**, Mossoró, v. 35, n. 1, p. 105 – 116, 2022.

ANTHONY, F. **Les ressources génétiques des caféiers: Collecte, gestion d'un conservatoire et evaluation de la diversité génétique.** Paris: Orstom Editions, 1992.

APARECIDO, L. E. de O. *et al.*. Predicting coffee yield based on agroclimatic data and machine learning. *Theoretical and Applied Climatology*, v. 148, p. 899–914, 2022.

ARIZA, V. M. P. *et al.*. Uso do algoritmo "Floresta Aleatória" na identificação do comportamento da população na busca por serviços de saúde a pós o início da pandemia do novo coronavírus. **AtoZ: novas práticas em informação e conhecimento**, v. 11, p. 1-15, 2022.

ASSAD, E. D. Impacto das mudanças climáticas no zoneamento agroclimático do café no Brasil. **Pesquisa agropecuária brasileira**, v. 39, n. 11, 2024.

BACH, C. L. *et al.* **Metodologia para estimativa de produtividade em lavouras de milho, trigo, soja e feijão.** Florianópolis: Epagri, 2020.

BASSOI, L. H. *et al.* Agricultura de precisão e agricultura digital. **Revista Digital de Tecnologias Cognitivas**, n. 20, p. 17-36, 2019.

BARROS, A. S.; FARIAS, L. M. de; MARINHO, J. L. A. aplicação do índice de vegetação por diferença normalizada (ndvi) na caracterização da cobertura vegetativa de Juazeiro Do Norte – CE. **Revista Brasileira de Geografia Física**, v. 13, n. 06, p. 2885-2895, 2020.

BARROS, P. P. da S. *et al.* Monitoramento fitossanitário utilizando sensoriamento remoto: avanços e desafios. **Revista Brasileira de Cartografia**, vol. 73, n. 2, 2021.

- BATISTELLA, D. *et al.* Comparative analysis of orbital sensors in soybean yield estimation by the random forest algorithm. **Ciência e Agrotecnologia**, v. 47, 2023
- BAZZAN, A. L. C. Contribuições de aprendizado por reforço em escolha de rota e controle semaforico. **Estudos Avançados**, v. 35, n. 101, 2021.
- BERNARDES, T. *et al.*. Monitoring biennial bearing effect on coffee yield using MODIS remote sensing imagery. **Remote Sensing**, v.4, p. 2492-2509, 2012. Doi:10.3390/rs4092492
- BERNARDES, T. *et al.*. Variáveis e modelos para estimativa da produtividade do cafeeiro a partir de índices de vegetação derivados de imagens Landsat. *In: XVI SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 2013, Foz do Iguaçu. **Anais...** Foz do Iguaçu: INPE, 2013. p. 0720-0727.
- BERTINI, F. B. *et al.*. **Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services**. ESA Communications, 2012.
- BIAU, G.; SCORNET, E. A random forest guided tour. **Test**, v. 25, p. 197–227, 2016
- BHERING, S. B. *et al.*. Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa agropecuária brasileira**, Brasília, v. 51, n. 9, p. 1359-1370, set. 2016.
- BOLFE, E. L. *et al.* Challenges, trends and opportunities in digital agriculture in Brazil. *In: MASSRUHÁ, S. M. F. S. et al. Digital agriculture: research, development and innovation in production chains*. Brasília, DF: Embrapa, 2023. cap. 16, p. 281-299.
- BRAINER, M. S. de C. P. Análise de aspectos da produção e mercado do café. **Escritório Técnico de Estudos Econômicos do Nordeste – ETENE**, n. 106, 2019.
- BRASIL. **Zoneamento Agrícola de Risco Climático – ZARC para a cultura do café arábica no Estado de Minas Gerais**. Mapa, 2021.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001.
- BUSKE, T. C. *et al.* Modelos agrometeorológicos para estimativa e comparativo de quebra de produtividade da cultura do milho. **Irriga**, Botucatu, v. 24, n. 4, p. 693-703, 2019.
- CAMARGO, A. P. de.; CAMARGO, M. B. P. de. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, Campinas, v. 60, n.1, p. 65-68, 2001.
- CARRIJO, W. R.; BORGES, M. T.; SOUZA, F. G. de. Resposta espectral do índice NDVI em cada fase fenológica do café. *In: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA – CONBEA*, 2021, Congresso On-Line. **Anais...** Congresso On-line: SBEA, 2021.
- CARVALHO, L. G. de. Modelos prognósticos de produtividade da cultura do café no estado de Minas Gerais. 2003. 168 f. Tese (Doutorado em Engenharia Agrícola) – Programa de Pós-Graduação, Universidade Federal de Viçosa, Viçosa, 2003.

CARVALHO, M. C. Data mining applied to feature selection methods for aboveground carbon stock modelling. **Pesquisa agropecuária brasileira**, Brasília, v.57, 2022.

CARVALHO, C. H. S. de. **Cultivares de Café**. Brasília, DF: Embrapa, 2007.

CARVALHO, C. S. de *et al.* **Catálogo de cultivares de café arábica**. Brasília, DF: Embrapa Café, 2022.

CASTRO, B. de O. S. *et al.* Correlação espacial dos índices de vegetação (NDVI, NDWI E EVI) de imagem landsat-8 com atributos do solo. *In: TULLIO, L. Aplicações e Princípios do Sensoriamento Remoto 2*. Ponta Grossa: Atena Editora, 2018.

CASTRO, R. Remote monitoring of coffee cultivation through computational processing of satellite images. *In: 7th International Engineering, Sciences and Technology Conference (IESTEC)*, 2019, Panama. **Anais [...]**. Panama, 2019, p. 13-18.

CERRI, R.; CARVALHO, A. C. P de L. F. de. Aprendizado de máquina: breve introdução e aplicações. **Cadernos de Ciência & Tecnologia**, Brasília, v. 34, n. 3, p. 297-313, set./dez. 2017.

CID, L. R. B.; TEIXEIRA, J. B. **Fisiologia vegetal: definições e conceitos**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2017.

CLARINDO, W. da R. *et al.* Estimativa de produtividade da cultura do milho safrinha por imagens de satélite Landsat 8. **Enciclopédia Biosfera**, Jandaia, v.19 n.40; p. 201-215, 2022.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento da safra brasileira de grãos: Safra 2020/21 – 4º Levantamento – v.8, n.4**. Brasília: Conab, 2021.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento da safra brasileira de café: 3º Levantamento (2024) – v. 11, n. 3**. Brasília: Conab, 2024.

CONCEIÇÃO, L. S. M. da. **A utilização do random forest para prever a inflação**. 2022. Monografia – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2022.

CORREA, D. V. *et al.* Aprendizagem de máquina para identificação de plantas de soja sob ataque de insetos usando dados hiperespectrais. **Colloquium Exactarum**, v.14, p.146–153, 2022.

CORRÊDO, L. de P. *et al.* Evaluation of minimum preparation sampling strategies for sugarcane quality prediction by vis-nir spectroscopy. **Sensors**, v. 21, n. 2195, 2021.

COSTA, J. P. de S.; RABELO, M. W. de O. Avaliação de modelos de regressão linear múltipla para estimativa da produtividade de cana-de-açúcar a partir de parâmetros biométricos e imagens do sensor MSI/Sentinel-2. *In: XX Simpósio Brasileiro de Sensoriamento Remoto*, 2023, Florianópolis. **Anais [...]**. São José dos Campos: INPE, 2023.

CRUSIOL, L. G. T. *et al.* **Conceitos e definições de sensoriamento remoto para monitoramento da soja**. Londrina: Embrapa Soja, 2024.

CUI, B. *et al.* A new integrated vegetation index for the estimation of winter wheat leaf chlorophyll content. **Remote Sensing**, v. 11, n. 974, 2019.

CUNHA, J. P. Z. **Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos**. 2019. 59 f. Dissertação (Mestrado em Estatística) – Programa de Pós-Graduação do Instituto de Matemática e Estatística, Universidade Federal de São Paulo, São Paulo, 2019.

CUSTÓDIO, A. A. de P. *et al.* Qualidade do café sob manejos de irrigação, faces de exposição solar e posições na planta. **Irriga**, Botucatu, v. 20, n. 1, p. 177-192, jan/mar, 2015.

DAVIDSON, C. *et al.* NDVI/NDRE prediction from standard RGB aerial imagery using deep learning. **Computers and Electronics in Agriculture**, v. 203, 2022.

DELAVARPOUR, N. *et al.* A Technical Study on UAV Characteristics for Precision Agriculture Applications and Associated Practical Challenges. **Remote Sensing**, v. 13, n. 1204, p. 1-25, 2021.

DONG, T. *et al.* Assessment of red-edge vegetation indices for crop leaf area index estimation. **Remote Sensing of Environment**, v. 222, p. 133-143, 2019.

DOORENBOS, J.; KASSAM, A. H. Yield response to water. Rome: **FAO**, 1979.

DORNELLES, K. A. **Absortância solar de superfícies opacas: métodos de determinação e base de dados para tintas látex acrílica e PVA**. 2008. Tese (Doutorado em Engenharia Civil, Arquitetura e Urbanismo) – Faculdade Engenharia Civil, Arquitetura e Urbanismo, Campinas, 2008.

ENGESAT. Sentinel 2. Disponível em: < <https://www.engesat.com.br/sentinel-2/> >. Acesso em 06 nov. 2024.

ESA - *European Space Agency*. **SentiWiki**.. Disponível em: < <https://sentiwiki.copernicus.eu/web/copernicus-programme> >. Acesso em: 06 nov. 2024

FERNANDES, F. T.; FILHO, A. D. P. C. Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. **Revista Brasileira de Saúde Ocupacional**, v. 44, n. 13, 2019.

FERRÃO, M. A. G. *et al.* **Indicação de cultivares de café arábica para o estado do Espírito Santo e avaliação comparativa com o conilon em altitude elevada**. Brasília, DF: Embrapa, 2021.

FERRÃO, R. G. *et al.* Coffea canphora. In: FERRÃO, R. G. *et al.* **Café conilon**. 2 ed. Vitória: Incaper, 2017.

FERREIRA, L. G.; FERREIRA, N. C.; FERREIRA, M. E. Sensoriamento remoto da vegetação: evolução e estado da arte. Acta Scientiarum. Biological Sciences, Maringá, v. 30, n. 4, p. 379-390, 2008.

FERREIRA, W. P. M. *et al.* **Estado nutricional dos solos e cafeeiros da região das Matas de Minas**. Brasília, DF: Embrapa Café, 2021.

FILHO, J. G. Ecofisiologia do cafeeiro. Cap. 6. *In*: DIAS, J. P. T. **Ecofisiologia de culturas agrícolas**. Belo Horizonte: EdUEMG, 2018.

FONSECA, A. C. P. da *et al.* Teores de macronutrientes do consórcio café com capim Braquiária no Cerrado. *In*: Congresso Brasileiro de Agrometeorologia, 22., 2023, Natal. **Anais [...]**. Natal: Sociedade Brasileira de Agrometeorologia, 2023.

FREITAS, A. F. de *et al.* Productivity and beverage sensory quality of arabica coffee intercropped with timber species. **Pesquisa agropecuária brasileira**, Brasília, v.55, p. 1-10, 2020.

FREITAS, R. A. S. de M. *et al.*. Hyperspectral characterization and estimation models for agronomic parameters of coffee cultivars after pruning. **Precision Agriculture**, v. 24, p. 2374-2394, 2023.

GIOVOS R. *et al.* Remote sensing vegetation indices in viticulture: a critical review. **Agriculture**, v. 11, n. 457, p. 1-20, 2021.

GOMES, J. V. P.; CUBAS, M. G. Fundamentos do sensoriamento remoto. Editora Intersaberes, 2021.

GONÇALVES, M. **Processamento digital de imagens de sensoriamento remoto para análise ambiental e geográfica**. Curitiba: Intersaberes, 2022.

GONZALEZ-ROBAINA, C. F.; CISNEROS-ZAYAS, C. H.; MONTILLA, C. E. Resposta ao déficit hídrico em plantas de café (*Coffea arabica* L.) em diferentes estágios de desenvolvimento. **Rev Cie Tech Agr**, San Jose de las Lajas, v. 26, n. 3, 2017.

GRABSKA, E. *et al.*. Forest Stand Species Mapping Using the Sentinel-2 Time Series. **Remote Sens.**, v. 11, n. 10, 2019.

GUEDES, P. T. *et al.* Avaliação de viabilidade econômica em um projeto de irrigação por gotejamento na cultura de café no nordeste paulista. *In*: XI Sintagro, 2019, Ourinhos-SP. **Anais [...]**. Ourinhos: Fatec, 2019.

GUIMARÃES, E. da S. **Aprendizado de máquina aplicado à predição da produtividade da cultura da soja utilizando dados de clima e solo**. 2019. 78 f. Dissertação (Mestrado Profissional em matemática, estatística e computação aplicadas à indústria) – Programa de Pós-Graduação do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos, 2019.

GUMAEI, A. *et al.* A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification. **IEEE Access**, vol. 7, p. 36266-36273, 2019.

GUTIERRES, M. I. de A.; NEVES, E. das. A importância do monitoramento da umidade do solo através de sensores para otimizar a irrigação nas culturas. **Enciclopédia Biosfera, Centro Científico Conhecer**, Jandaia, v.18, n.35, p. 1-16, 2021.

HOMEM, W. L. **Apostila de machine learning**. Vitória: Universidade Federal do Espírito Santo, 2020.

HUETE, A. *et al.*. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, n. 83, p. 195-213, 2002.

IBGE. **Levantamento sistemático da produção agrícola: conceitos e métodos**. 2024. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistemico-da-producao-agricola.html?=&t=conceitos-e-metodos>>. Acesso em: 28 abr. 2024.a.

IBGE - Instituto Brasileiro de Geografia e Estatística. LSPA (Levantamento Sistemático Da Produção Agrícola). 2024. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistemico-da-producao-agricola.html>>. Acesso em: 24 dez. 2024.b.

IGNACIO, L. F. F. **Aprendizado de máquina: da teoria à aplicação**. 2021. Trabalho de Conclusão de Curso (Graduação em Matemática) - Universidade Federal Fluminense, Instituto de Ciências Exatas, Volta Redonda, 2021.

INÁCIO, A. da S. **Análise qualitativa do índice de vegetação por diferença normalizada (NDVI) extraído a partir de imagens Rapideye e utilizando-se de software livre – estudo de caso município de São Miguel dos Campos/AL**. 2016. 89 f. Monografia (Graduação em Engenharia de Agrimensura) – Universidade Federal de Alagoas, Maceió, 2016.

INDEX DATABASE. **List of available Indices**. 2024. Disponível em: <<https://www.indexdatabase.de/>>. Acesso em: 08 maio 2024.

JARA, Q. R. *et al.*. Compostos de esterco de galinha e de ovino melhoram a produtividade de plantas de cafeeiro sob um sistema de produção orgânica. **Arnaldoa**, v. 26, n. 2, p. 735-750, 2019.

KAMAL, B. K. *et al.*. Effect of different altitudes in qualitative and quantitative attributes of green coffee beans (*Coffea arabica*) in Nepal. **International journal of Horticulture, Agriculture and Food Science**, v. 5, n. 3, p. 1-7, maio/jun 2021.

KINOSHITA, J. K.; MENEGHETTI, D. de R.; BIANCHI, R. A. da C. Aprendizado por Reforço Profundo com Redes Recorrentes Aplicado à Negociação do Minicontrato Futuro de Dólar. In: BRAZILIAN WORKSHOP ON ARTIFICIAL INTELLIGENCE IN FINANCE (BWAIF), 2022, Niterói. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 13-24.

LAROS, J. A.; PUENTE-PALACIOS, K. E. Validação cruzada de uma escala de clima organizacional. **Estudos de Psicologia**, v. 9, n. 1, p. 113-119, 2004.

LEDA, V. C.; GOLÇALVES, A. K.; LIMA, N. da S. Sensoriamento remoto aplicado a modelagem de produtividade da cultura da cana-de-açúcar. **Energia na Agricultura**, Botucatu, v. 34, n. 2, p. 263-270, 2019.

LEIVAS, J. F. *et al.* Monitoramento agrícola baseado em imagens de satélite de diferentes resoluções espaciais em áreas do oeste da Bahia. *In: IV Simpósio da Bacia Hidrográfica do Rio São Francisco*, 2022, Belo Horizonte. **Anais [...]**. Belo Horizonte SBHSF, 2022.a.

LEIVAS, J. F. *et al.* Análise da dinâmica temporal de índices de vegetação NDVI, EVI, SAVI e IRECI através de imagens Sentinel-2a e MODIS. *In: MELO, J. O. F. Ciências agrárias: o avanço da ciência no Brasil*. São Paulo: Editora Científica Digital, 2022.b.

LINHARES, A. L. de C. **Impactos da precipitação na variação dos preços do café entre 2012 E 2020**. 2022. 42 f. Trabalho de Conclusão de Curso (Graduação em Economia) - Programa De Graduação Em Ciências Econômicas Da Faculdade De Economia, Administração, Atuária E Contabilidade Departamento De Economia, Universidade Federal Do Ceará, 2022

LOURENÇO, R. W.; LANDIM, P. M. B. Estudo da variabilidade do “Índice de Vegetação por Diferença Normalizada/NDVI” utilizando krigagem indicativa. **HOLOS Environment**, v. 4, n. 1, p. 38-55, 2004.

LUDERMIR, T.B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Inteligência Artificial**, v. 35, n. 101, 2021.

LUNZ, A. M. P.; BERNARDES, M. S.; RIGHI, C. A. Crescimento e produtividade do cafeeiro sob diferentes níveis de sombreamento com seringueira. *In: GONÇALVES, R. C.; OLIVEIRA, L. C. de. Embrapa Acre: Ciência e tecnologia para o desenvolvimento sustentável do sudoeste da Amazônia*. Rio Branco: Embrapa Acre, 2009.

MACHADO, A. H. R. *et al.* A cultura do café (*Coffea arabica*) em sistema agroflorestal. **Brazilian Journal of Animal and Environmental Research**, Curitiba, v. 3, n. 3, p. 1357-1369, 2020.

MAHESH, B. Machine Learning Algorithms - A Review. **International Journal of Science and Research**, v. 9, 2018.

MAPA, MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Projeções do agronegócio, Brasil – 2021/22 a 2031/32. SPA/Mapa, 2022.

MARTELLO, M. *et al.*. **Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning**. *AgriEngineering*, v. 4, n. 4, p. 888-902, 2022.

MARTINS, G. D.; SILVA, D. Comparação de algoritmos de aprendizado de máquina na previsão do churn em uma empresa de tecnologia na área de educação continuada a distância. **Inteligencia Artificial**, v. 20, n. 59, p. 123-137, 2020.

MEIRELES, E. J. L. *et al.* **Fenologia do cafeeiro**: condições agrometeorológicas e balanço hídrico do ano agrícola 2004–2005. Brasília, DF: Embrapa Informação Tecnológica, 2009.

MELO, B. de.; SOUSA, L. B. de. Biologia da reprodução de *Coffea arabica*. L. e *Coffea canephora* Pierre. **Revista Verde**, Mossoró, v.6, n.2, p. 01-07, abr/ jun. 2011.

MENESES, P. R.; ALMEIDA, T. de; BAPTISTA, G. M. de M. **Reflectância dos materiais terrestres: análise e interpretação**. São Paulo: Oficina de Textos, 2019.

MION, R. M. R. **Previsão de tendência de preços do boi gordo no Estado de São Paulo utilizando o Random Forest**. 2019. 72 f. Dissertação (Mestrado Profissional em Economia) – Programa de Pós Graduação, Fundação Getúlio Escola de Economia de São Paulo, São Paulo, 2019.

MIQUELLUTI, D. L. **Métodos alternativos de previsão de safras agrícolas**. 2015. 87 f. (Dissertação em Estatística e Experimentação Agrícola) - Programa de Pós-Graduação, Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2015

MORAIS, G. L.; BREDEMEIER, C. **Índices de vegetação para estimativa da biomassa da parte aérea e produtividade do trigo**. In: SIC - XXXI SALÃO DE INICIAÇÃO, 2019.

MOREIRA, M. A.; ADAMI, M.; RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa agropecuária brasileira**, Brasília, v. 39, n. 3, p. 223-231, mar. 2004.

MOREIRA, N. J. **Breves considerações sobre a história e cultura do cafeeiro e consumo de seu produto**. Rio de Janeiro: Typ. Do Imperial Instituto Artístico, 1873.

MOREIRA, P. C. *et al.* Produtividade e economia de fatores de produção na cafeicultura brasileira. **Revista de Política Agrícola**, v. 28, n. 2, p. 6-21, 2019.

MOREIRA, V.R.R. **Consórcio de café com bananeira**. Caderno tecnológico. 2 ed. Fortaleza: Ministério da Ciência e Tecnologia, 2004.

MORELLO, O. F. *et al.* Agronomic performance, postharvest and indirect selection of *Coffea arabica* L. cultivars for high-temperature regions. **Revista Brasileira de Ciências Agrárias**, Recife, v. 15, n. 3, 2020.

MOTA, F. M. *et al.* Relação entre variáveis meteorológicas e Índices de Vegetação (IV's) em lavouras cafeeiras. **Brazilian Journal of Animal and Environmental Research**, Curitiba, v. 3, n. 4, p. 4287-4209, 2020.

MULLAINATHAN, S.; SPIESS, J. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, v. 31, n. 2, p 87–106, 2017.

MÜLLER, A.C.; GUIDO, S. **Introduction to Machine Learning with Python: a guide for data scientists**. O'Reilly Media, October 2016.

NASCIMENTO, A. L. do. **Estimativa de produtividade de café por meio de métodos de machine learning**. 2019. Tese (Doutorado em Engenharia Agrícola) - Universidade Federal de Viçosa, Viçosa, 2019.

NDUKU, L. *et al.* Global Research Trends for Unmanned Aerial Vehicle Remote Sensing Application in Wheat Crop Monitoring. **Geomatics**, v. 3, p. 115–136, 2023.

NEFTCI, E. O.; AVERBECK, B. B. Reinforcement learning in artificial and biological systems. **Nature Machine Intelligence**, v. 1, p. 133–143, 2019.

NOGUEIRA, S. M. C.; MOREIRA, M. A.; VOLPATO, M. M. L. Relationship between coffee crop productivity and vegetation indexes derived from OLI / Landsat-8 sensor data with and without topographic correction. **Engenharia Agrícola**, Jaboticabal, v.38, n.3, p.387-394, 2018.

NÓIA JUNIOR, R. de S. *et al.* Sistema de monitoramento de produtividade de milho 2ª safra no Mato Grosso. **Agrometeoros**, Passo Fundo, v.30, 2022.

OLIVEIRA, A. D. de *et al.* **Monitoramento das variáveis ambientais em café cultivado com e sem braquiária no Planalto Central Brasileiro**. Boletim de Pesquisa e Desenvolvimento. Planaltina: Embrapa Cerrados, 2024.

PAES, A. P. B. **Uso de séries temporais espectrais para estimar a produtividade do café**. 2022. Trabalho de Conclusão de Curso (Graduação em Engenharia de Agrimensura e Cartográfica) - Universidade Federal de Uberlândia campus Monte Carmelo, Monte Carmelo, 2022.

PARAJARA, G. do C. *et al.* Avaliação de espécie leguminosa em consórcio com café no controle de plantas invasoras. *In: XI Congresso Brasileiro de Agroecologia*, 2020, São Cristóvão. **Anais [...]**. São Cristóvão: UFS, 2020.

PEREIRA, E. M.; BEKER, J. P. dos S.; SCHMOELLER, R. Aplicação de Machine Learning na Previsão da Produtividade da Soja. **Revista Pleiade**, v. 18, n. 45, p. 25-38, 2024.

PEREIRA, F. V. *et al.* estimativa da clorofila do cafeeiro utilizando dados radiométricos e aprendizado de máquina. *In: X Simpósio Brasileiro de Sensoriamento Remoto*, 2023, Florianópolis. **Anais [...]**. São José dos Campos: INPE, 2023.

PEÑA-BARRAGÁN, J. M. *et al.* Object-based crop identification using multiple vegetation indices, textural features and crop phenology. **Remote Sensing of Environment**, v. 115, n. 6, p. 1301–1316, 2011.

PHIRI, D. *et al.*. Sentinel-2 Data for Land Cover/Use Mapping: A Review. **Remote Sens.**, v. 12, n. 14, 2020.

PICINI, A. G. **Desenvolvimento e teste de modelos agrometeorológicos para a estimativa de produtividade do cafeeiro (*Coffea arabica* L.) a partir do monitoramento da disponibilidade hídrica do solo**. 1998. 132 f. Dissertação (Mestrado em Agronomia) – Programa de Pós-Graduação, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 1998.

PIRES, I. F. *et al.* Diagnóstico nutricional em plantação de café Conilon na região norte do estado do Espírito Santo. **Revista Agrogeoambiental**, v.14, p. 1-9, 2022.

QUEIROZ, D. M. de. *et al.* Sensors applied to Digital Agriculture: A review. **Revista Ciência Agrônômica**, v. 51, 2020.

- RIBEIRO, A. F. F. *et al.*. Paclobutrazol e restrição hídrica no crescimento e desenvolvimento de plantas de café. **Coffee Science**, Lavras, v. 12, n. 4, p. 534 - 543, 2017.
- RIBEIRO, G. M. *et al.*. Imagens RGB para estimativa e avaliação da variabilidade espacial da produtividade do cafeeiro. In: 11º SIMPÓSIO DE CIÊNCIAS AGRÁRIAS E AMBIENTAIS – SICAA, 2024, Monte Carmelo. Anais... Monte Carmelo: UFU, 2024.
- RIBEIRO, M. de F.; PEDROSA, A. W.; FERREIRA, W. P. M. Introdução. In: FERREIRA, W. P. M. **Boas práticas agrícolas aplicadas à lavoura cafeeira para o estado de Minas Gerais**. Brasília, DF: Embrapa Café, 2022.
- RODRIGUES FILHO, S. V. *et al.* Desenvolvimento de um sistema de monitoramento ambiental e agrícola utilizando Arduino e Xbee. In: Congresso Brasileiro de Agroinformática (SBIAGRO), 2023, Natal. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2023.
- RODRIGUES, L. de S.; Pereira, D. R. Aprendizado de máquina aplicado em imagem NDVI para previsão da produtividade da cana-de-açúcar. **Colloquium Exactarum**, v. 13, n. 4, p.82–98, 2021.
- ROSA, V. G. C. da. *et al.* Estimativa da produtividade de café com base em um modelo agrometeorológico-espectral. **Pesquisa agropecuária brasileira**, Brasília, v.45, n.12, p.1478-1488, dez. 2010.
- ROTATA, E. *et al.*, Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. **Remote Sensing of Environment**, v. 222, n. 1, p. 1-17, 2019.
- SAJID, S. S. *et al.*. County-scale crop yield prediction by integrating crop simulation with machine learning models. **Frontiers in Plant Science**, v. 13, p. 01-16, 2022.
- SAKAI, E. *et al.*. Coffea arabica (cv Catuaí) production and bean size under different population arrangements and soil water availability. **Scientific Papers**, v. 33, n. 1, 2013.
- SANDRINI, F. de O. *et al.* Índices de vegetação na estimativa da produtividade do feijoeiro cultivado sob doses de nitrogênio. **Revista Brasileira de Ciências Agrárias**, Recife, v.14, n.4, 2019.
- SANTOS, E. de O. *et al.* Controlador Fuzzy para Monitoramento Agrícola Contextualizado. In: Escola Regional de Computação Bahia, Alagoas e Sergipe (ERBASE), 2020, Arapiraca. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020.
- SANTOS, R. P. dos; BEKO, M.; LEITHARDT, V. R. Q. Modelo de machine learning em tempo real para agricultura de precisão. In: Escola regional de alto desempenho da região sul (ERAD-RS), 2022, Curitiba. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2022.
- SANTOS, W. B. dos. **Análise espectro-temporal de produtos do sensor Modis como diagnóstico para a cafeicultura de precisão**. 2015. 71 f. Dissertação (Mestrado em

Engenharia Agrícola) – Programa de Pós Graduação da Universidade Federal de Lavras, Lavras, 2015.

SEGARRA, J. *et al.*. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. **Agronomy**, v. 10, n. 5, 2020.

SERAFIM, M. E. *et al.*. Disponibilidade hídrica e distinção de ambientes para cultivo de cafeeiros. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, n. 4, 2013.

SILVA, A. L. F. **Estudo de seca na Península Ibérica usando o Google Earth Engine**. 2018. Dissertação (Mestrado em Engenharia Geográfica) – Universidade do Porto, 2018.a.

SILVA, E. A. da. *et al.*. Influência de déficits hídricos controlados na uniformização do florescimento e produção do cafeeiro em três diferentes condições edafoclimáticas do Estado de São Paulo. **Irrigação**, Bragantia, v. 68, n. 2, 2009.

SILVA, E. A. da; MAZZAFERA, P. Influence of Temperature and Water on Coffee Culture. **The Americas Journal of Plant Science and Biotechnology**, v. 2, n. 2, p; 32-41, 2008.

SILVA, E. P. da; ARAÚJO, V. S. M. Matemática na cultura da soja. **Revista Multidebates**, Palmas, v.4, n.5, p. 156-169, 2020.

SILVA, F. S. *et al.* Correlações da produtividade do algodoeiro com índices de vegetação obtidos por séries temporais de imagens de satélite. *In*: Simpósio Nacional De Instrumentação Agropecuária, 4., 2019, São Carlos, SP. **Anais [...]**. São Carlos: Embrapa Instrumentação, 2019.

SILVA, G. H. N. da. **Utilizando as técnicas de Random Forest e Redes Neurais para previsão de níveis na cidade de Piaçabuçu**. 2018. 89 f. Monografia (Graduação em Engenharia de Computação) - Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, 2018b.

SILVA, M. B. da. *et al.* Nutritional balance and its relationship to yield in a coffee field: Inferences from geospatial analysis. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.24, n.12, p.834-839, 2020.

SILVA, P. H. V. V. da. **Previsão da produção de café de Rondônia utilizando algoritmos de aprendizado de máquina**. 2023. Monografia (Graduação em Engenharia de Produção) - Universidade Federal de Rondônia, Cacoal, 2023.

SILVA, S. de A.; LIMA, J. S. de.; OLIVEIRA, R. B. de. Modelo agrometeorológico na estimativa da produtividade de duas variedades de café arábica considerando a variabilidade espacial. **Irriga**, Botucatu, v. 16, n. 1, p. 1-10, 2011.

SILVEIRA, E. M. de O. *et al.* Modelling aboveground biomass in forest remnants of the Brazilian Atlantic Forest using remote sensing, environmental and terrain-related data. **Geocarto International**, 2019. DOI: 10.1080/10106049.2019.1594394.

- SIQUEIRA, H. M. de *et al.* Análise econômica de consórcios do cafeeiro conilon com espécies perenes e florestais no sul do Espírito Santo. **Revista Brasileira de Agroecologia**, v. 15, n. 5, p. 222-235, 2020.
- SOARES, W. O. *et al.* Influência das mudanças climáticas na produção cafeeira segundo a percepção das cafeicultoras. **Revista Formação**, v. 27, n. 52, p. 77-100, 2020.
- SOUZA, A. das G.; SMIDERLE, O. J. Cultura da soja: radiação solar, acúmulo de matéria seca e suas interações para alta produtividade. *In*: MELO, J. O. F. *et al.*. **Ciências agrárias: tecnologia, sustentabilidade e inovação** – v. 1. Guarujá-SP: Científica Digital, 2024.
- SOUZA, K. X. S. de *et al.* Agricultura digital: Definições e tecnologia. *In*: MASSRUHÁ, S. M. F. S. *et al.* (Ed.). **Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas**. Brasília: Embrapa, 2020. cap. 2, p. 46-66.
- TAIZ, L. *et al.*. **Fisiologia e desenvolvimento vegetal** – 6. ed. Porto Alegre: Artmed, 2017.
- TAVARES, T. R. *et al.* Multi-sensor approach for tropical soil fertility analysis: comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies. **Agronomy**, v. 11, n. 1028, 2021.
- TSOUROS, D. C.; BIBI, S.; SARIGIANNIDIS, P. G. A Review on UAV-Based applications for precision agriculture. **Information**, v. 10, n. 349, 2019.
- TRONCHA, M. F. A. C. **Importância da radiação solar na morfologia, fisiologia e produtividade de plantas de cafeeiro**. 2023. 11 f. Trabalho de Conclusão de Curso (Graduação em Agronomia) – Programa de Graduação, Universidade Estadual de Goiás, Ipameri, 2023.
- UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. **Coffee: World Markets and Trade**. 2024. Disponível em: <https://www.fas.usda.gov/sites/default/files/2024-12/coffee.pdf> >. Acesso 14 jan. 2025.
- VEIGA, A. D. *et al.* **Clones de cafés arábicas no cerrado central: desempenho agrônomico e resistência ao bicho-mineiro (*Leucoptera coffeella*)**. Planaltina: Embrapa Cerrados, 2023.
- VERDIN FILHO, A. C. *et al.* Implicações do espaçamento e do número de ramos ortotrópicos sobre o crescimento e a produção do cafeeiro conilon. *In*: X Simpósio de Pesquisa dos Cafés do Brasil, 2019, Vitória. **Anais [...]**. Vitória: Consórcio Pesquisa Café, 2019.
- VIANA, M. J. do N. *et al.* Morfologia de cafeeiros cultivados com diferentes fontes de adubos fosfatados. *In*: X Simpósio de Pesquisa dos Cafés do Brasil, 2019, Vitória. **Anais [...]**. Vitória: Consórcio Pesquisa Café, 2019.
- VIEIRA, E. L. *et al.*. **Manual de Fisiologia Vegetal**. Brasil: EDUFMA, 2010.
- VIEIRA, H. D. **Café Rural: Noções da cultura**. Rio de Janeiro: Interciência, Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro, 2017.

WINTGENS, J. N. The coffee plant. *In*: WINTGENS, J. N. **Coffee: Growing, Processing, Sustainable Production**. Suíça: Wiley-VCH, 2004.

YASEEN, Z. M. *et al.* An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research Direction. **Journal of Hydrology**, v. 569, p. 387–40, 2019.

YIN, L. *et al.*. Research on Stock Price Rise and Fall Prediction Based on Optimization Random Forest. *CAAI Transactions on Intelligence Technology*, v. 8, n. 7, p. 61-67, 2021.

ZACARIAS, A. J. *et al.* Custo / benefício da adubação verde do consórcio no café conilon. **Research, Society and Development**, v. 9, n. 11, p. 1-22, 2020.