



**STHÉFANY AIRANE DOS SANTOS SILVA**

**INTEGRAÇÃO DE TECNOLOGIAS DE AGRICULTURA DE  
PRECISÃO E DIGITAL APLICADAS NA GESTÃO DA  
CAFEICULTURA**

**LAVRAS - MG  
2025**

**STHÉFANY AIRANE DOS SANTOS SILVA**

**INTEGRAÇÃO DE TECNOLOGIAS DE AGRICULTURA DE  
PRECISÃO E DIGITAL APLICADAS NA GESTÃO DA  
CAFEICULTURA**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós  
Graduação em Engenharia Agrícola, área de  
concentração em Máquinas e Mecanização Agrícola,  
para obtenção do título de Doutora.

Prof. Dr. Gabriel Araújo e Silva Ferraz  
Orientador  
Dra. Vanessa Castro Figueiredo  
Coorientadora

**LAVRAS - MG**

**2025**

**Ficha de identificação da obra elaborada pelo(a) autor(a) através do Sistema de Geração Automática de Ficha Catalográfica da Biblioteca Universitária da UFLA.**

Silva, Sthéfany Airane dos Santos.

Integração de tecnologias de agricultura de precisão e digital aplicadas na gestão da cafeicultura / Sthéfany Airane dos Santos Silva. 2025.  
95 p.: il.

Orientador: Gabriel Araújo e Silva Ferraz  
Coorientadora: Vanessa Castro Figueiredo

Tese (Doutorado) - Universidade Federal de Lavras, 2025.  
Bibliografia.

1. cafeicultura de precisão. 2. variabilidade. 3. aeronaves remotamente pilotadas. 4. índices de vegetação. 5. aprendizado de máquina. I. Ferraz, Gabriel Araújo e Silva. II. Figueiredo, Vanessa Castro. III. Universidade Federal de Lavras. IV. Título.

**STHÉFANY AIRANE DOS SANTOS SILVA**

**INTEGRAÇÃO DE TECNOLOGIAS DE AGRICULTURA DE  
PRECISÃO E DIGITAL APLICADAS NA GESTÃO DA  
CAFEICULTURA**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de Pós  
Graduação em Engenharia Agrícola, área de  
concentração em Máquinas e Mecanização Agrícola,  
para obtenção do título de Doutora.

APROVADA em 24 de abril de 2025.

Dra. Vanessa Castro Figueiredo	EPAMIG
Dra. Margarete Marin Lordelo Volpato	EPAMIG
Dr. Lucas dos Santos Santana	UFVJM
Dr. Danton Diego Ferreira	UFLA

Orientador  
Prof. Dr. Gabriel Araújo e Silva Ferraz

Coorientadora  
Dra. Vanessa Castro Figueiredo

**LAVRAS - MG**

**2025**

À Deus, minha fortaleza, força e fonte de sabedoria,  
Aquele que nunca me desampara,  
mesmo nos momentos mais difíceis.  
Foi Sua presença constante  
que me sustentou, iluminou meu caminho e  
me deu forças para seguir em frente.  
Esta conquista é, acima de tudo, resultado da fé,  
da graça e do amor d'Ele por mim.

**DEDICO.**

## **Agradecimentos**

Primeiramente, agradeço a Deus por me guiar e ser sempre o meu amparo; e à Nossa Senhora Aparecida, por interceder constantemente em minhas orações.

Agradeço aos meus pais, Mário e Kênia, pelo amor incondicional e por nunca medirem esforços ao longo da minha trajetória. Cheguei até aqui porque houve muito trabalho, dedicação e entrega por parte deles.

Sou especialmente grata ao meu esposo, Adilson, por sempre confiar em mim e apoiar todas as minhas escolhas, por acreditar que meu sonho era possível e por ser minha fortaleza quando tudo parecia difícil, principalmente pelo seu amor e compreensão.

Agradeço à minha família, em especial à minha irmã Raiany e aos meus avós José e Lourdes, por todo o incentivo e apoio incondicional ao longo dessa jornada.

Ao meu orientador, Professor Dr. Gabriel Araújo e Silva Ferraz, agradeço profundamente pelo aprendizado, orientação e amizade ao longo desses dez anos. Sua confiança em meu trabalho e seus ensinamentos foram fundamentais para que eu chegasse até aqui.

À minha coorientadora Dra. Vanessa e também aos pesquisadores Dra. Margarete, Dr. Marley e Dra. Vânia, agradeço por todo o apoio ao longo dessa caminhada, pelo incentivo, pela orientação e, principalmente, pela amizade. Vocês foram peças essenciais na construção deste trabalho.

Agradeço também aos membros da EPAMIG Lavras, EPAMIG Três Pontas e EPAMIG Sede, por serem grandes apoios na realização desta pesquisa.

Aos amigos do NESAs, minha gratidão pela amizade, ajuda e troca de conhecimento ao longo do curso.

À Universidade Federal de Lavras (UFLA) e ao Programa de Pós-Graduação em Engenharia Agrícola, agradeço pela oportunidade de realizar este curso e pela contribuição para minha formação.

À EMBRAPA e ao Projeto Consórcio Pesquisa Café, agradeço pelo financiamento deste trabalho, por meio dos projetos (10.18.20.023.00.00 e 10.18.20.041.00.00), que incentivaram e apoiaram esta pesquisa.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (projeto 310186/2023-4) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG (projeto APQ-00661-22), agradeço pelo apoio financeiro e pela concessão da bolsa de estudos.

Meus sinceros agradecimentos a todos que fizeram parte não apenas da construção deste trabalho, mas de toda a minha trajetória.

## RESUMO GERAL

Os fatores que mais influenciam o rendimento das lavouras estão interligados e abrangem aspectos climáticos, agronômicos e tecnológicos. A avaliação e monitoramento desses fatores tornam-se necessários, uma vez que, quando bem investigados, podem trazer retornos como uma maior produtividade e qualidade dos frutos, além de tornarem o ciclo produtivo mais sustentável. Este trabalho tem como objetivo investigar o uso de tecnologias de Agricultura de Precisão (AP) e Agricultura Digital (AD) na cafeicultura, aplicadas em um campo experimental da EPAMIG, situado em Três Pontas, MG. A área de estudo possui 1,2 ha com cafeeiros da espécie *Coffea arabica*, cultivar Topázio MG1190. No período compreendido entre 2020 e 2022, foram coletados dados de produtividade, umidade do solo, potencial hídrico foliar, fertilidade do solo, nutrição das folhas, além da realização de voos com Aeronave Remotamente Pilotada (ARP). A partir desse conjunto de dados, esta pesquisa foi dividida em três partes. A primeira parte oferece uma revisão de literatura sobre a importância da cafeicultura no Brasil, os desafios da produção e o papel das tecnologias emergentes, como a AP e a AD, para otimizar a produtividade e sustentabilidade. Na segunda parte, são apresentados três artigos. O primeiro artigo avaliou a variabilidade espacial e temporal da produtividade de cafeeiros ao longo de três safras, utilizando métodos geoestatísticos. Os resultados revelaram uma alta heterogeneidade na produtividade, com variações significativas tanto no espaço quanto no tempo, mesmo em uma área relativamente reduzida. Essa variabilidade foi atribuída a fatores como bienalidade e ao manejo convencional adotado na área. O segundo artigo analisou a variabilidade espacial do índice de vegetação NDVI e da clorofila, obtidos por sensores ativos e passivos do sensoriamento remoto. Pela análise geoestatística foi possível constatar a variabilidade espacial dos atributos avaliados. Através da análise do coeficiente de Pearson (R) foi possível quantificar o nível de correlação entre os dados obtidos pelos diferentes sensores e através do teste t foi possível encontrar correlações significativas entre eles. No terceiro artigo, foram testados os algoritmos de aprendizado de máquina Random Forest (RF), Gradient Boosting (GB), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN) e Decision Tree (DT) para a previsão da produtividade dos cafeeiros. Os modelos preditivos foram construídos utilizando como variáveis de entrada dados de umidade do solo, potencial hídrico foliar, fertilidade do solo, nutrição foliar, índices de vegetação, diâmetro e altura das plantas, além do índice de área foliar. A variável de saída corresponde à produtividade observada para os anos 2021 e 2022. A implementação e manipulação dos modelos foram realizadas na linguagem Python, e a avaliação do desempenho dos algoritmos foi conduzida por meio das métricas  $R^2$  (coeficiente de determinação) e MAPE (erro percentual absoluto médio), permitindo analisar a precisão das previsões geradas. Na terceira parte deste trabalho, foi apresentada uma consideração final sobre os estudos desenvolvidos, destacando a importância da Agricultura de Precisão e Digital na cafeicultura.

**Palavras-chave:** cafeicultura de precisão; variabilidade; aeronaves remotamente pilotadas; índices de vegetação; aprendizado de máquina; previsão de produtividade.

## GERAL ABSTRACT

The factors that most influence crop yield is interconnected and encompass climatic, agronomic, and technological aspects. The evaluation and monitoring of these factors become necessary, as, when well investigated, they can lead to increased productivity and fruit quality, in addition to making the production cycle more sustainable. This study aims to investigate the use of Precision Agriculture (PA) and Digital Agriculture (DA) technologies in coffee cultivation, applied in an experimental field of EPAMIG, located in Três Pontas, MG, Brazil. The study area comprises 1.2 hectares of *Coffea arabica* plants, cultivar Topázio MG1190. Between 2020 and 2022, data were collected on yield, soil moisture, leaf water potential, soil fertility, leaf nutrition, along with data obtained through flights with a Remotely Piloted Aircraft (RPA). Based on this dataset, the research was divided into three parts. The first part provides a literature review on the importance of coffee cultivation in Brazil, the production challenges, and the role of emerging technologies such as PA and DA in optimizing productivity and sustainability. In the second part, three scientific articles are presented. The first article evaluated the spatial and temporal variability of coffee yield over three harvests using geostatistical methods. The results revealed high heterogeneity in yield, with significant variations in both space and time, even within a relatively small area. The second article analyzed the spatial variability of the NDVI vegetation index and chlorophyll content obtained through active and passive remote sensing sensors. Through geostatistical analysis, spatial variability of the evaluated attributes was observed. Pearson's correlation coefficient (R) was used to quantify the correlation level between the data from different sensors, and the t-test revealed significant correlations among them. In the third article, the machine learning algorithms Random Forest (RF), Gradient Boosting (GB), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), and Decision Tree (DT) were tested for predicting coffee yield. The predictive models were built using input variables such as soil moisture, leaf water potential, soil fertility, leaf nutrition, vegetation indices, plant diameter and height, and leaf area index. The output variable corresponds to the observed yield for the years 2021 and 2022. Model implementation and manipulation were conducted in Python, and performance evaluation was carried out using  $R^2$  (coefficient of determination) and MAPE (mean absolute percentage error) metrics, allowing an analysis of prediction accuracy. The third part of this work presents final considerations on the studies carried out, highlighting the importance of Precision and Digital Agriculture in coffee farming.

**Keywords:** precision coffee farming; variability; remotely piloted aircraft; vegetation indices; machine learning; yield forecasting.

## INDICADORES DE IMPACTO

Os resultados deste trabalho apresentam impactos significativos em diferentes esferas. No âmbito tecnológico, destacam-se o uso integrado de sensores, aeronaves remotamente pilotadas (ARPs) e métodos geoestatísticos para mapear e compreender a variabilidade espacial e temporal da produtividade dos cafeeiros, mesmo em uma área relativamente pequena. A análise aprofundada dos dados de NDVI e clorofila permitiu avaliar o estado nutricional e o vigor das plantas com precisão, possibilitando diagnósticos mais rápidos e eficientes no campo. Essas ferramentas, aliadas à aplicação de algoritmos de aprendizado de máquina na previsão da produtividade, demonstram o potencial das tecnologias emergentes para transformar a cafeicultura por meio da Agricultura de Precisão (AP) e da Agricultura Digital (AD). No aspecto econômico, a identificação da variabilidade produtiva ao longo das safras e o uso de modelos preditivos possibilitam um planejamento agrícola mais eficiente, otimizando o uso de insumos, mão de obra e recursos naturais. Essa abordagem favorece a redução de custos operacionais e melhora o retorno financeiro dos produtores, especialmente em regiões como o Sul de Minas, onde a cafeicultura tem grande importância econômica. Em termos sociais, a disseminação dessas tecnologias e conhecimentos por meio da Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG) fortalece a conexão entre pesquisa e extensão rural, promovendo práticas sustentáveis e baseadas em evidências entre técnicos, produtores e estudantes. A adoção dessas estratégias contribui para o empoderamento do produtor, promove o uso racional dos recursos e reforça o compromisso com a sustentabilidade ambiental e produtiva. Dessa forma, os impactos deste estudo extrapolam o campo experimental, com potencial de aplicação em diversas regiões produtoras, agregando valor à cadeia do café. A pesquisa, portanto, gera impactos sociais, tecnológicos e econômicos alinhados com os Objetivos de Desenvolvimento Sustentável (ODS) da ONU, com destaque para os ODS 2 (fome zero e agricultura sustentável), 9 (indústria, inovação e infraestrutura), 12 (consumo e produção responsáveis).

## **IMPACT INDICATOR**

The results of this work demonstrate significant impacts across different spheres. In the technological domain, the integrated use of sensors, remotely piloted aircraft (RPAs), and geostatistical methods stands out for mapping and understanding the spatial and temporal variability of coffee crop productivity, even in relatively small areas. The in-depth analysis of NDVI and chlorophyll data enabled accurate assessment of the nutritional status and vigor of the plants, allowing for faster and more efficient diagnostics in the field. These tools, combined with the application of machine learning algorithms for yield prediction, highlight the potential of emerging technologies to transform coffee farming through Precision Agriculture (PA) and Digital Agriculture (DA). From an economic perspective, identifying productivity variability over multiple harvests and using predictive models allows for more efficient agricultural planning, optimizing the use of inputs, labor, and natural resources. This approach contributes to reducing operational costs and increasing farmers' financial returns, especially in regions such as southern Minas Gerais, where coffee production holds substantial economic importance. On the social front, the dissemination of these technologies and insights through EPAMIG (Empresa de Pesquisa Agropecuária de Minas Gerais – Agricultural Research Company of Minas Gerais) strengthens the link between research and rural extension, fostering sustainable and evidence-based practices among technicians, farmers, and students. The adoption of such strategies contributes to farmer empowerment, encourages the rational use of resources, and reinforces the commitment to environmental and productive sustainability. Thus, the impacts of this study go beyond the experimental field, with the potential for application in various coffee-producing regions, adding value to the coffee production chain. The research, therefore, generates social, technological, and economic impacts aligned with the United Nations Sustainable Development Goals (SDGs), especially SDG 2 (Zero Hunger and Sustainable Agriculture), SDG 9 (Industry, Innovation and Infrastructure), SDG 12 (Responsible Consumption and Production).

## SUMÁRIO

<b>PRIMEIRA PARTE.....</b>	<b>12</b>
1 INTRODUÇÃO GERAL.....	12
2 OBJETIVO GERAL.....	14
2.1 Objetivos específicos.....	14
3 REFERENCIAL TEÓRICO.....	15
3.1 Importância da cafeicultura e desafios em sua produção.....	15
3.2 Produtividade dos cafeeiros.....	16
3.3 Cafeicultura de precisão.....	17
3.3.1 Geoestatística empregada na cafeicultura de precisão.....	18
3.3.2 Sistema de informação geográfica e sensoriamento remoto.....	20
3.3.3 Aeronaves remotamente pilotada.....	22
3.3.3.1 Aeronaves Remotamente pilotadas na cafeicultura.....	23
3.3.3.2 Índices de vegetação.....	24
3.4 Agricultura digital.....	25
3.4.1 Machine learning.....	26
3.4.2 Cafeicultura digital.....	27
REFERÊNCIAS .....	30
<b>SEGUNDA PARTE – ARTIGOS .....</b>	<b>37</b>
ARTIGO 1 - SPATIAL AND TEMPORAL VARIABILITY OF PRODUCTIVITY OF COFFEE PLANTS GROWN IN AN EXPERIMENTAL FIELD LOCATED IN TRÊS PONTAS, BRAZIL.....	37
INTRODUCTION.....	34
MATERIALS AND METHOD.....	38
RESULTS AND DISCUSSION.....	42
CONCLUSION.....	47
REFERENCES.....	47
ARTIGO 2 - SPATIAL VARIABILITY OF CHLOROPHYLL AND NDVI OBTAINED BY DIFFERENT SENSORS IN AN EXPERIMENTAL COFFEE FIELD.....	50
INTRODUCTION.....	51
MATERIALS AND METHOD.....	52
RESULTS AND DISCUSSION.....	56
CONCLUSION.....	61
REFERENCES.....	61
ARTIGO 3 – DESEMPENHO DE ALGORITMOS DE MACHINE LEARNING PARA A PREDIÇÃO DE PRODUTIVIDADE DE UMA LAVOURA CAFEEIRA.....	64
1 INTRODUÇÃO.....	65
2 MATERIAL E MÉTODOS.....	66
3 RESULTADOS E DISCUSSÃO.....	78
4 CONCLUSÃO.....	91
REFERÊNCIAS.....	92
<b>TERCEIRA PARTE.....</b>	<b>95</b>
CONSIDERAÇÕES FINAIS.....	95

## PRIMEIRA PARTE

### 1 INTRODUÇÃO GERAL

O Brasil destaca mundialmente na produção de café, sendo o maior produtor e exportador mundial e o segundo maior consumidor. A cafeicultura é uma atividade de grande importância, gerando um impacto positivo no agronegócio brasileiro. A segunda estimativa para a safra brasileira de café em 2024 realizada pela Companhia Nacional de Abastecimento (CONAB) apontam uma produção total de 58,81 milhões de sacas de café beneficiado, considerando as espécies arábica e conilon, um acréscimo de 6,8% em relação ao ano de 2023 (CONAB 2024).

Os fatores que mais influenciam o rendimento das lavouras estão interligados e abrangem aspectos climáticos, agronômicos e tecnológicos. As condições climáticas desempenham um papel crucial, uma vez que temperatura, precipitação e incidência de luz solar impactam diretamente o crescimento e o desenvolvimento das plantas. O solo e sua fertilidade também são determinantes, pois fornecem os nutrientes essenciais para a saúde dos cafeeiros. Além disso, a utilização da mecanização e da automação exerce impacto significativo, aumentando a eficiência dos processos e reduzindo os custos de produção.

A cafeicultura no Brasil ocorre dentro de uma diversidade de fatores que podem influenciar fortemente a produtividade dos cafeeiros (Ferraz et al., 2017). A variabilidade espacial é uma das premissas para a aplicação da Cafeicultura de Precisão (CP), que surge como uma abordagem para aplicar, tecnologias, técnicas e ferramentas inovadoras para melhorar o sistema cafeeiro de maneira sustentável, aumentando a produtividade e a rentabilidade para o produtor (Santana et al., 2021). O termo Cafeicultura de Precisão (CP) foi definido por Ferraz et al. (2012), sendo um conjunto de técnicas e tecnologias que visam apoiar o manejo da cultura cafeeira.

A Agricultura de Precisão concentra seu foco na avaliação da variabilidade espacial das características do solo e das plantas, buscando otimizar a rentabilidade, a eficiência da fertilização, dos tratamentos de pulverização e da colheita, o que leva a uma maior produtividade e qualidade dos grãos. Atualmente muitas pesquisas estão sendo realizadas com foco na aplicação de agricultura de precisão (AP) na cafeicultura. A aplicação de tecnologias baseada em Sistemas de Informação Geográfica (SIG) tem se mostrado eficiente para entender a distribuição espacial do rendimento em um campo.

Atualmente, a Agricultura Digital, que incorpora as técnicas da Agricultura de Precisão, é apresentada pela EMBRAPA (Empresa Brasileira de Pesquisa e Agropecuária) como uma solução para desafios agrícolas. Entre eles, destacam-se a otimização da produção e do planejamento agrícola, promovendo maior sustentabilidade e decisões mais assertivas. Essa abordagem também integra máquinas inteligentes com tecnologias digitais emergentes, como Big Data e aprendizado de máquina, permitindo o processamento e a análise de grandes volumes de dados em tempo real. Além disso, o uso de sensores, satélites, drones e robôs aprimora o monitoramento das lavouras, tornando mais precisa a identificação de padrões, a previsão de produtividade e a otimização dos recursos agrícolas (Massruhá et al., 2020; Rosalen, 2022).

A utilização de técnicas e ferramentas da CP associadas a AD, podem otimizar o uso de recursos como água, fertilizantes e defensivos agrícolas, resultando em uma redução de custos e minimização de impactos ambientais, e consequentemente, trazendo um aumento de produtividade e qualidade dos grãos de café. Além disso, a implementação da AD na cafeicultura pode proporcionar aos produtores de café dados detalhados e em tempo real sobre suas lavouras, permitindo assim uma gestão eficiente, o que resulta em melhorias significativas de produção e lucratividade para os produtores.

Observa-se uma tendência notável para estudar as aplicações da agricultura digital em vários cultivos, dessa forma há uma demanda na comunidade científica por pesquisas que relacionem a cultura do café a estas novas tecnologias, avançando assim para novas tendências.

Considerando que a produtividade do cafeeiro é influenciada por diversos fatores, relacionados ao solo, as condições de saúde da planta, o clima e também pelo manejo realizado dentro do ciclo produtivo, torna-se imprescindível aplicar técnicas e ferramentas que identifiquem e monitorem as condições das lavouras de maneira rápida e eficiente, bem como avaliar as causas da variabilidade da produtividade, e que através da aplicação de técnicas da agricultura digital seja possível criar um modelo que descreva o atributo produtividade.

Neste sentido, este estudo foi dividido em três partes: a) avaliar a variabilidade espacial da produtividade através de métodos geoestatísticos, b) avaliação do NDVI (normalized difference vegetation index) e clorofila obtidos por diferentes sensores, com o intuito de verificar a correlação entre eles e c) criar um modelo de previsão de produtividade através de dados de campo e índices de vegetação obtidos por Aeronaves Remotamente Pilotada (RPA).

## **2 OBJETIVO GERAL**

Avaliar o desempenho de técnicas e ferramentas associadas à cafeicultura de precisão e agricultura digital em lavoura cafeeira situada em campo experimental.

### **2.1 Objetivos específicos**

- Avaliar a variabilidade espacial da produtividade no período de três safras por meio de análise geoestatística;
- Avaliar a variabilidade espacial do NDVI e da clorofila obtidos por diferentes sensores, bem como suas correlações;
- Avaliar o desempenho de diferentes algoritmos de aprendizado de máquina para construção de modelos de previsão de produtividade dos cafeeiros.

### 3 REFERENCIAL TEÓRICO

#### 3.1 Importância da cafeicultura e desafios em sua produção

O café é um produto de grande importância no comércio agrícola internacional, especialmente em países que produzem e dependem dela para compensar a balança de pagamentos, impostos e subsistência (Marin, 2022). Depois do petróleo bruto, o café é a commodity mais comercializada no mundo (Esposito et al., 2020), sendo produzido em mais de 70 países, onde a maior parte desta produção concentra-se em países em desenvolvimento.

A cafeicultura tem importância histórica no Brasil e se mantém como importante cultura geradora de divisas e empregos. Ainda que seja uma cultura tradicional, o padrão produtivo do café segue em transformação (Moreira et al., 2019). A safra de 2023, de bialidade negativa, a maior safra do ciclo de baixa, sucedeu duas safras, 2021 e 2022, com eventos climáticos adversos como, estiagens e baixas temperaturas e, em 2023, com as condições climáticas favoráveis, deu-se início à fase de recuperação das produtividades (CONAB 2024).

O cultivo de cafeeiros no Brasil ocorre em ambientes com grande diversidade de clima, solo, relevo, características fisiológicas (cultivares, idade, diâmetro de saia e copa, cor do fruto, etc.), infestações fitossanitárias, manejo (sistema de plantio, sistemas de colheita, sistemas de irrigação, mecanização, sistemas de adubação e pulverização, etc.), dentre muitas outras características (Ferraz et al., 2012). O entendimento da variabilidade espacial de atributos relacionados ao solo e a planta dentro de uma lavoura exige uma maior quantidade de informações que podem ser obtidas por técnicas e ferramentas associadas a Agricultura de Precisão (AP).

A cafeicultura utiliza tecnologia intensiva em trabalho, esta cultura é uma atividade geradora de emprego e de renda, sobretudo quando se consideram as demais atividades ao longo de toda a cadeia do produto, bem como o superávit da balança comercial, fator que favorece o desenvolvimento econômico (Saraiva; Fernandes, 2018). Segundo Faria et al. (2019) a produção de café está cada dia mais tecnológica em busca da otimização do uso dos insumos e também do uso sustentável dos recursos naturais.

Pesquisas têm sido realizadas no intuito de determinar a dependência espacial de variáveis do solo e das plantas, em separado ou correlacionadas na tentativa de interpretar e otimizar a aplicação de insumos.

A utilização conjunta de diferentes variáveis do sistema solo-planta pode conduzir a resultados mais confiáveis, no levantamento de dados, oferecendo informações mais precisas

para o gerenciamento do agronegócio cafeeiro (Silva et al., 2014), identificando e interpretando, por exemplo, zonas de diferentes produtividades.

### **3.2 Produtividade dos cafeeiros**

Ainda que a cafeicultura seja uma cultura tradicional e com importância histórica para o Brasil, a dinâmica produtiva do café segue em transformação (Moreira et al., 2019). Modelos recentes preveem que a produção de café será afetada por mudanças climáticas relacionadas ao aumento de temperatura e da variabilidade da precipitação, induzindo a ocorrência de pragas e doenças (Jawo; Kyereh; Lojka, 2023) e consequentemente o fornecimento limitado de nutrientes do solo.

Apesar da importância econômica do café para países de baixa renda ao redor do mundo, a indústria tem lutado por décadas com vários estressores de longa data e choques repentinos, incluindo reformas institucionais, volatilidades de preços de mercado, eventos climáticos extremos e a infestação de pragas e doenças. Os produtores enfrentam um cenário desafiador, onde precisam adaptar para garantir a qualidade do café e a sustentabilidade de suas práticas. É crucial que eles se mantenham informados sobre melhores práticas e inovações na produção (Rhiney; Knudson; Guido, 2020; Bacon et al., 2017; Avelino et al., 2015).

Altas taxas de produtividade de café resultam da aplicação de práticas tecnológicas durante as etapas de produção e processamento. Os aumentos na produtividade devem ser atribuídos a uma combinação de fatores, que vão desde a mecanização e automação do plantio e da colheita, passando pelo aprimoramento das técnicas de manejo e pela aplicação de defensivos com equipamentos mais avançados e tecnológicos, até o desenvolvimento de variedades resistentes a doenças por meio do melhoramento genético, o que ajuda a reduzir as perdas de produção e a preservar a qualidade do café (Moreira et al., 2019).

A produtividade da cultura sofre oscilações devido a ocorrência de infestações fitossanitárias (Chalfoun et al., 1978), fatores climáticos (Carvalho et al., 2004), características fisiológicas da cultura (Rena et al., 1996), sistema de plantio adotado, densidade e população de planta (Toledo; Barros, 1999), declividade e posição topográfica (Souza et al., 2004) dentre diversos outros fatores ainda não bem elucidados (Carvalho et al., 2006).

A agricultura moderna é caracterizada pela rápida expansão das tecnologias de informação decorrentes do monitoramento e controle do armazenamento, organização e atividades agrícolas (Oliveira et al., 2021). O uso de técnicas e tecnologias voltadas para altos níveis de produtividade aliadas à sustentabilidade é conhecido como agricultura de precisão

(Oliveira et al., 2020). A prática agrícola de precisão pode maximizar o potencial de cada região, tornando a cultura mais produtiva e favorecendo a redução de custos (Murugan et al., 2017).

Na literatura muitas pesquisas foram desenvolvidas para estimar a produtividade do cafeeiro através de: modelos agrometeorológicos (Picini et al., 1999; Santos; Camargo, 2006; Rosa et al., 2010; Victorino; Carvalho; Ferreira, 2016; Jayakumar; Rajavel, 2017; Oliveira et al., 2017; Aparecido; Rolim, 2018; Valeriano et al., 2018), modelos matemáticos (Miranda; Reinato; Silva, 2014, Bolaños; Corrales; Campos, 2023), técnicas de cokrigagem (Lima et al., 2016).

Apesar de ser uma técnica ainda recente, o aprendizado de máquinas baseado em dados climáticos, condições do solo e imagens multiespectrais também têm sido utilizado para gerar modelos de previsão de safras dos cafeeiros como encontrados nos trabalhos de (Kouadio et al., 2018; Barbosa et al., 2021; Rodríguez; Griol Barres; Callejas Carrión, 2021; Abreu Júnior et al., 2022; Carvalho Alves et al., 2022; Kittichotsatsawat; Tippayawong; Tippayawong, 2022; Martello et al., 2022; Oliveira et al., 2022; Santhosh; Umesh, 2022; Yen; Khoi, 2022; Zanella et al., 2024)

O ponto inicial para a aplicação de AP nas lavouras, demanda a identificação da variabilidade nas lavouras, sendo assim, os mapas de produtividade são tidos como a informação mais completa e verdadeira para visualizar a variabilidade nos cultivos (Molin et al., 2015).

### **3.3 Cafeicultura de precisão**

Dentre vários conceitos acerca de AP, um deles destaca-se como sendo uma forma de gestão de todo o processo agrícola, levando em consideração a variabilidade existente nas lavouras, sendo necessário avaliar, quantificar e mapear, afim de conduzir adequadamente e de forma eficiente (Molin et al., 2015). Os avanços tecnológicos na agricultura de precisão contribuem para a obtenção de medições precisas e confiáveis em uma lavoura, isso, pode facilitar o monitoramento de aspectos agronômicos em uma escala mais precisa (Paccioretti; Córdoba; Balzarini, 2020).

Considerando que são grandes os desafios na produção de café, principalmente se tratando da grande variabilidade de atributos relacionados ao solo, as plantas, ao clima e também ao manejo, surge o conceito de Cafeicultura de Precisão (CP). Este termo foi designado por Alves et al. (2006), e segundo Santana et al. (2022) pode ser definido como o uso de técnicas e tecnologias atualizadas que visam maximizar a lucratividade das culturas, aumentar a

eficiência das operações, buscar a sustentabilidade do negócio, uma produção ambientalmente sustentável, a maximização da produtividade e melhoria da qualidade final do produto.

Sistemas de posicionamento global via satélite (GNSS), sistema de informação geográfica (SIG), sensores de campo ou embarcados em satélites e aeronaves remotamente pilotadas, são alguns exemplos de tecnologias da AP que auxiliam o monitoramento das lavouras cafeeiras. Por meio dessas tecnologias é possível criar mapas de rendimentos, avaliar e monitorar atributos relacionados ao solo e as plantas, bem como todos os aspectos que envolvam o ciclo produtivo, como clima, manejo e saúde das plantas, permitindo uma melhor compreensão do ambiente e das operações, gerando assim, um processo mais rápido e eficiente para a tomada de decisão (Liakos et al., 2018; Cadenas et al., 2020).

A aplicação de técnicas de AP na cafeicultura vem passando por transformações expressivas, a partir do uso de novas ferramentas e tecnologias. Geoestatística, sistema de informação geográfica, sensoriamento remoto, aeronaves remotamente pilotadas, são exemplos de tecnologias empregadas na AP e quando aliadas à agricultura digital promovem maior eficiência e precisão na gestão agrícola. A partir desse ponto, os tópicos apresentados nesta revisão de literatura vão apresentar definições e como essas ferramentas vêm sendo aplicadas na cafeicultura.

### **3.3.1 Geoestatística empregada na cafeicultura de precisão**

Desde o início do século XX, existe uma preocupação entre pesquisadores em relação a variabilidade do solo e conseqüentemente das suas propriedades físico-hídricas, por proporcionar diferenças no desenvolvimento, produtividade e manejo da irrigação, alterando resultados de pesquisas, mesmo com o solo considerado homogêneo em parcelas experimentais. Por esta razão, a caracterização da variabilidade espacial é essencial para um melhor entendimento das inter-relações entre atributos do solo e manejo de irrigação (Andrade et al., 2005).

O comportamento dos atributos do solo pode ser determinado tanto por estatística clássica quanto por geoestatística, entretanto se os dados não revelam uma variação completamente aleatória, a geoestatística torna-se uma ferramenta imprescindível para descrever o comportamento desses atributos na área (Almeida; Guimarães, 2016). A geoestatística é uma ciência consolidada e uma ramificação da estatística aplicada que aborda questões relacionadas a variáveis regionalizadas, as quais apresentam um comportamento espacial com características que se situam entre as variáveis puramente aleatórias e as

completamente determinísticas (Marasca et al., 2017).

Segundo Mazione & Zimback (2011), a geoestatística é uma ferramenta imprescindível para trabalhar com dados de atributos que apresentam dependência espacial, permitindo compreender o comportamento dessa dependência a partir da modelagem dos semivariogramas experimentais. Com um modelo de dependência espacial de variáveis do solo, a geoestatística possibilita a estimativa em pontos não amostrados, viabilizando o mapeamento e o zoneamento das variáveis (Andrade et al., 2005).

A geoestatística foi uma das primeiras ferramentas empregadas para a aplicação de técnicas de AP, considerando que um dos seus maiores objetivos é a identificação e interpretação da variabilidade dos atributos que envolvem todo o ciclo produtivo. Manzione et al. (2001) afirmam que na agricultura, a aplicação da análise geoestatística tem sido útil na caracterização e mapeamento da variação espacial dos atributos das culturas.

Para a cultura do café, o conhecimento da variabilidade espacial dos atributos da planta, é de suma importância para a determinação de estratégias para atender as necessidades da cultura e conseqüentemente alcançar maior produtividade (Silva et al., 2010).

Muitas pesquisas utilizando a geoestatística foram empregadas na cafeicultura com o intuito de avaliar a variabilidade espacial de atributos químicos (Sanchez et al., 2005; Silva et al., 2007; Silva et al., 2008; Silva et al., 2010; Silva; Lima; Souza, 2010; Silva; Lima, 2012; Ferraz et al., 2012; Almeida; Guimarães, 2016, Cezana et al., 2024) e físicos (Sanchez et al., 2005; Valente et al., 2014; Ferraz et al., 2012; Ferraz et al., 2017) do solo e das plantas bem como a avaliação da variabilidade espacial da produtividade (Sanchez et al., 2005; Silva et al., 2007; Silva et al., 2008; Silva; Lima, 2012, Ferraz et al., 2012, Ferraz et al., 2017).

Um dos problemas chave do uso da geoestatística na AP é a necessidade de uma grade de amostragem densa, afim de se detectar a variabilidade espacial dos atributos em questão, possibilitando a geração de mapas que representem de maneira real o comportamento espaço/temporal de cada atributo, porém quanto maior o número de amostras, maior se torna o ônus financeiro para o produtor.

Dessa forma, surge neste contexto, o uso de sensores advindos do Sensoriamento Remoto (SR), que possibilitam a avaliação da variabilidade espacial dos atributos de maneira rápida e eficiente, sem ter contato direto com o alvo, aumentando o número de amostragens e reduzindo as amostragens intensivas com custos elevados.

### 3.3.2 Sistema de informação geográfica e sensoriamento remoto

A dinâmica e a complexidade das operações no setor agrícola demandam um controle rigoroso e a atualização contínua das informações (Souza, Vieira e Alves, 2004). O conjunto de geoprocessamento, ou geotecnologias, que incluem as tecnologias para coleta, processamento, análise e disponibilização de informação com referência geográfica, tem grande potencial para gestão da produção agrícola e pecuária (Batistella et al., 2011; Filippini Alba, 2014).

A modelagem via sistema de informações geográficas (SIG) possibilita a fusão dessas camadas de informações, ampliando a capacidade de interpretação dos dados e auxiliando na tomada de decisão para a gestão do sistema de produção (Filippini Alba, 2014). Uma das definições mais atuais utilizadas de um Sistema de Informação Geográfica (SIG) é de um sistema baseado em hardware, software, informação espacial e procedimentos computacionais que permitem e facilite a análise, gestão ou representação do espaço e dos fenômenos que nele ocorrem (Gonçalves; Sousa, 2017).

O SIG vêm sendo amplamente utilizado na cafeicultura para otimizar a produção e a gestão das lavouras. Neles, são integrados dados geoespaciais que auxiliam na tomada de decisão (Alvarenga et al., 2022), permitindo o mapeamento detalhado de áreas de cultivo, identificando variáveis como altitude, declividade, tipo de solo e padrões climáticos bem como fatores críticos para o crescimento do café. Com o SIG, é possível monitorar a saúde das plantas, gerenciar pragas e doenças e identificar áreas com necessidade de intervenções específicas, como irrigação ou adubação.

A incorporação de tecnologias SIG na agricultura ampliou as possibilidades de desenvolver estruturas espaciais ainda mais avançadas, que apoiam a criação de bancos de dados agrícolas dinâmicos e sistemas interativos (Nie et al., 2010). Esses sistemas permitem que os usuários interajam em tempo real com dados agrícolas georreferenciados, oferecendo informações posicionais precisas e criando bases mais robustas para a tomada de decisões.

Uma ferramenta importante, presente em ambientes SIG, é a análise de dados advindos do Sensoriamento Remoto (SR), uma técnica que obtém dados da superfície terrestre por meio da captação e registro em imagem da energia refletida/emitida pela superfície, sem que haja contato físico entre o sensor e o alvo estudado (Pereira et al. 2016).

Informações de sensoriamento remoto como crescimento, vigor e a dinâmica da vegetação terrestre podem fornecer informações extremamente úteis para monitoramento das culturas. Especificamente, esses tipos de informações aplicadas à agricultura fornecem não

apenas uma base objetiva para macro e micro região agrícola, mas também, em muitas ocasiões, as informações necessárias para estimativa do rendimento das culturas (Mulla, 2013).

Brandão (2009) afirma que são várias as aplicações obtidas pelo SR, dentre elas, se destacam: a estimativa da biomassa e produtividade da cultura, monitoramento de estresse hídrico e do vigor nas plantas e avaliação do estágio fenológico. Os dados de SR têm sido propostos como uma boa solução para a medição de parâmetros da vegetação.

Em relação a aplicação de SR na cafeicultura, trabalhos como os de Velloso (1974, 1976) e Velloso e Souza (1978) são um dos primeiros registros de aplicação dessa ferramenta nas lavouras cafeeiras. Os autores avaliaram a capacidade do sensor (MSS/Landsat) no monitoramento dos cafeeiros. Porém, em nenhum dos casos obteve-se bons resultados, devido à baixa resolução espacial do sensor.

A partir dos lançamentos dos satélites Landsat 5 (Thematic Mapper) e Landsat 7 (Enhancement Thematic Mapper Plus) de média resolução espacial, diversos trabalhos de mapeamento da cafeicultura foram desenvolvidos com imagens desses satélites (Cordero-Sancho; Sader 2007; Machado et al. (2001,2010); Martínez-Verduzco et al., 2012; Moreira et al. (2008, 2010); Sarmiento et al., 2014; Trabaquini et al., 2011). Os trabalhos apresentaram bons resultados, porém houve confusão espectral entre as classes de café e áreas florestais, e outro problema identificado foi a presença de nuvens em boa parte das imagens, dificultando a discriminação das classes.

Informações espectrais são fundamentais para obter uma acurácia mais elevada na identificação de respostas da vegetação. O sensoriamento remoto tem evoluído significativamente ao longo das últimas décadas, impulsionado por avanços tecnológicos em sensores, plataformas orbitais e Aeronaves Remotamente Pilotadas (ARPs), os quais têm ampliado a capacidade de coleta, resolução e aplicação de dados em diversas escalas (Zhang et al., 2021; Shi; Yin, 2018; Salcedo-Sanz et al., 2020). Essa evolução permite o monitoramento cada vez mais preciso e detalhado da superfície terrestre, possibilitando a obtenção de dados em alta resolução sobre uma ampla gama de fenômenos naturais e atividades humanas.

O uso de dados de SR tem sido amplamente utilizado no monitoramento de lavouras considerando o avanço tecnológico dessa ferramenta. Na lavoura de café, o uso desses dados tem se mostrado muito promissor (Barbosa et al., 2021). Apesar da aplicabilidade das imagens orbitais (satélites), estas possuem algumas desvantagens, seu processamento pode ser complexo, levando em consideração o tipo de sensor, às configurações que podem variar de acordo com as características de cada cultura, como espaçamento, sombreamento, idade e espécies (Jiménez-Brenes et al., 2019). Dentro das tecnologias e ferramentas de SR, quando se

compara o uso de uma ARP em relação as imagens advindas de plataformas orbitais, o primeiro apresenta vantagens devido à sua capacidade de voar em baixas altitudes, permitindo uma maior resolução do alvo, além do uso irrestrito sob demanda e em momentos críticos (Jiménez-Brenes et al., 2019), além da facilidade de operação (Oliveira et al., 2016).

Os dados adquiridos por sensores orbitais ou em plataformas aéreas está passando por uma transição com o surgimento e desenvolvimento de sensores de alta resolução (Benediktsson; Chanussot; Moon, 2012). Uma alternativa para os sensores multiespectrais orbitais é o uso de Aeronaves Remotamente Pilotadas (ARPs). Este equipamento no setor agrícola, surgiu como importante ferramenta da Agricultura de Precisão na missão de reconhecimento da exploração agropecuária por meio de tomadas de imagens (Viana et al. 2018).

### **3.3.3 Aeronaves remotamente pilotadas**

Dentro da agricultura, a inserção das aeronaves remotamente pilotadas (ARPs) tem se mostrado promissora, e vêm crescendo em ritmo acelerado, devido ao baixo custo de operação, pela sua resolução espacial e temporal e também, pela sua flexibilidade na aquisição de imagens (Puri et al., 2017; Reger et al., 2018). As ARPs são utilizadas para imageamento de pequenas áreas com alta taxa de revisita e altíssima resolução espacial (Eisenbeiss, 2004; Esposito et al., 2006; Nebiker et al., 2008).

O uso de ARPs ainda é uma tecnologia emergente, que ganhou popularidade dentro do SR nos últimos anos, devido principalmente à sua capacidade de obter imagens de alta resolução e de maneira contínua (Rango et al., 2006), operando em baixas altitudes, o que as torna menos limitadas pelas condições climáticas. As operações com ARP envolvem menor custo e complexidade, além da flexibilidade para missões repetidas (Hunt et al., 2010; Laliberte et al., 2011).

Dados SIG obtidos de tecnologias como ARP por exemplo e de outros dispositivos de informação georreferenciada (Ahmad et al., 2020; Manlove et al., 2021), combinados com o uso de imagens multiespectrais, têm o potencial de aumentar a frequência de monitoramento e a resolução espacial dos dados de imagem (Wan et al., 2020).

Na cafeicultura de precisão, uma das inovações mais relevantes é o uso de Aeronaves Remotamente Pilotadas (ARPs) para a captura de imagens aéreas, facilitando o monitoramento de áreas extensas de cultivo.

### 3.3.3.1 Aeronaves remotamente pilotadas na cafeicultura

Os sistemas de aeronaves remotamente pilotadas surgiram como ferramentas valiosas na agricultura de precisão, particularmente na cafeicultura. Esses sistemas permitem processos eficientes de monitoramento, coleta de dados e tomada de decisão de culturas (Bento et al., 2023a; Santos et al., 2019a). As ARPs podem ser usados para várias aplicações, incluindo mapeamento, extração de índice de vegetação e medição da altura da planta (Santos et al., 2019a).

Nas lavouras de café, as APRs têm sido empregadas para classificação supervisionada do uso e cobertura da terra, bem como para o cálculo do Índice de Vegetação por Diferença Normalizada (NDVI) para avaliar o vigor das plantas (Santos et al., 2021b). Além disso, imagens adquiridas por ARP têm sido utilizadas para mapear parâmetros agrônômicos, como teor de clorofila e área foliar em lavouras de café (Assis et al., 2023). A integração da tecnologia RPA na cafeicultura contribui para melhorar a amostragem, as medições e a eficiência operacional, levando ao aumento da lucratividade e práticas de manejo sustentável (Bento et al., 2023a; Santos et al., 2019a).

Oliveira et al. (2018), por sua vez, aplicaram técnicas de processamento de imagens RGB obtidas por ARP, com operadores morfológicos, para identificar e quantificar falhas no plantio do cafeeiro.

Santos et al. (2019b) trabalharam na correção de erros geométricos em imagens capturadas por ARP, com foco em melhorar a qualidade de mapeamentos em lavouras de café.

Cunha et al. (2019) avaliaram a utilização de sensores embarcados em ARPs para estimar o volume da vegetação em cafezais, obtendo resultados equivalentes aos métodos convencionais de levantamento.

Santana et al. (2019) utilizaram dados obtidos por RPAs para avaliar a qualidade de diferentes técnicas de plantio em terrenos de alta declividade, aplicando controle estatístico de processo (CEP) no município de Bom Sucesso, Sul de Minas Gerais.

Santos et al. (2020) realizaram o monitoramento de parâmetros biofísicos, altura e diâmetro de copa de cafeeiros durante um período de doze meses, demonstrando a eficiência dos produtos fotogramétricos gerados por ARPs para estimativas precisas, reduzindo a necessidade de coletas em campo.

Bento et al. (2022) monitoraram o desenvolvimento de cafeeiros recém-transplantados por meio de imagens aéreas obtidas por RPAS, realizando estimativas de altura e diâmetro da copa em diferentes períodos e validando os dados com medições de campo. Em outro estudo, Bento et

al. (2023b) aplicaram imagens multiespectrais de alta resolução para mapear áreas com infestação de plantas daninhas, indicando estratégias de controle localizado e obtendo uma economia de 92,68% em relação ao tratamento total da área.

Barata et al. (2024) avaliaram a utilização de imagens multiespectrais de alta resolução capturadas por ARP para acompanhar a resposta de cafeeiros após o transplante de mudas, demonstrando a eficácia do sensoriamento remoto no monitoramento do desenvolvimento inicial da cultura.

Santos et al. (2024) detectaram a presença do bicho-mineiro (*Leucoptera coffeella*) em lavouras cafeeiras, utilizando índices de vegetação derivados de imagens aéreas e algoritmos de floresta aleatória para classificar plantas saudáveis e infestadas.

Esses estudos demonstram a ampla aplicabilidade das ARPs na cafeicultura de precisão, abrangendo desde a detecção de pragas e falhas no plantio até o monitoramento de desenvolvimento vegetativo, com ganhos em agilidade, precisão e redução de custos operacionais.

Entre as várias aplicações das ARPs na agricultura, destaca-se a utilização de índices de vegetação, que permitem avaliar o estado fisiológico das plantas por meio da análise de informações espectrais. O cálculo de índices de vegetação (IV) a partir de imagens multiespectrais tornou-se indispensável no desenvolvimento de modelos eficazes para prever a produtividade das culturas ou monitorar vários parâmetros agrícolas (Goigochea-Pinchi et al., 2024).

### **3.3.3.2 Índices de vegetação**

A utilização de ARP em conjunto com a AP tem sido direcionada para a aplicação de sensores multiespectrais acoplados às ARPs e que utilizam a espectroscopia de reflectância. Isso envolve a medição da reflexão da radiação eletromagnética (REM) após sua interação com várias superfícies em diferentes comprimentos de onda, abrangendo as regiões do visível, infravermelho próximo e infravermelho de ondas curtas.

Os índices de vegetação são calculados por meio de combinações matemáticas dos valores de reflectância, que são então relacionados às características fisiográficas das plantas (Shiratsuchi et al., 2014; Pezzopane et al., 2019). Vários índices de vegetação podem ser determinados a partir dessas imagens de alta resolução.

Pesquisas têm sido direcionadas com o propósito de demonstrar a aplicabilidade de IV's, explorando as propriedades espectrais da vegetação, especialmente nas regiões do visível e do infravermelho próximo (Ponzoni; Shimabukuro, 2010). Na cafeicultura, algumas pesquisas são

encontradas usando índices de vegetação para monitoramento: índice de área foliar (Barbosa et al., 2021), cinzas residuais e desenvolvimento de plantas de café (Santana et al., 2024), ervas daninhas (Bento et al., 2023); condições hídricas (Santos et al., 2022a); clorofila (Santos et al., 2022b); bicho mineiro (Santos et al., 2022c; Santos et al., 2024), nitrogênio foliar (Marin et al., 2021), maturação do café (Nogueira Martins et al., 2021), caracterização de cultivares (Bento et al., 2022); atributos químicos das folhas (Bento et al., 2022), produtividade (Bento et al., 2022), estado nutricional (Rivera; Bonilla; Obando-Vidal, 2021).

### **3.4 Agricultura digital**

Até o presente momento, este referencial teórico abordou vários aspectos, pesquisas e ferramentas utilizadas na agricultura de precisão com foco na cafeicultura. Levando em consideração que a partir da inserção de novas práticas e tecnologias, pode-se afirmar que nos recentes últimos anos, a AP passou por uma transição, sendo então nomeada para Agricultura Digital/Inteligente ou Agricultura 4.0 e até mesmo em relatos recentes denominando-a de Agricultura 5.0, representando uma evolução significativa no setor agrícola.

Enquanto a agricultura de precisão focava na utilização de tecnologias para otimizar o uso de insumos e maximizar a produtividade, a agricultura digital vai além, integrando dados de diversas fontes em tempo real para permitir uma tomada de decisão ainda mais informada e automatizada. A agricultura digital é crucial para maximizar o rendimento e as receitas das colheitas e preservar os recursos naturais (Yost et al., 2019). Queiroz et al. (2021) afirmam que um dos maiores benefícios da Agricultura Digital (AD) é o suporte ao agrônomo na definição de suas recomendações, em tempo real, de forma precisa e à luz e informações relevantes.

A modernização do setor agropecuário tem um grande potencial para garantir sustentabilidade, produtividade máxima e um ambiente mais seguro (Lampridi; Sørensen; Bochtis, 2019). Nesse contexto, a agricultura digital surge como um novo campo científico que utiliza abordagens baseadas em dados massivos (big data) para impulsionar a produtividade agrícola e minimizar os impactos ambientais (Liakos et al., 2018).

A AD é uma agricultura cada vez mais conectada e remota que vai atuar no levantamento e no processamento dessa grande quantidade de dados coletados em todos os elos das cadeias produtivas, incluindo as fases de pré-produção, produção e pós-produção. Ela envolve, portanto, diferentes tipos de tecnologias digitais: sensores embarcados em plataformas orbitais, suborbitais, aerotransportados ou sistemas autônomos (drones, máquinas agrícolas), instalados diretamente no campo ou em diferentes ‘coisas’ (IoT) ao longo das cadeias

produtivas, sistemas de telecomunicação, posicionamento global, software de controle, gestão e análise (data analytics) e atuadores (Massruhá et al., 2020).

A evolução da transformação digital na agricultura tem modificado diversos aspectos da gestão agrícola, promovendo a transição para sistemas baseados em inteligência artificial, cujo objetivo é agregar valor aos dados gerados por diferentes fontes do campo (Benos et al., 2021). A agricultura inteligente, nesse sentido, fundamenta-se em quatro pilares essenciais para atender às demandas crescentes do setor: (a) gestão eficiente dos recursos naturais, (b) conservação do ecossistema, (c) desenvolvimento de serviços adequados e (d) utilização de tecnologias modernas (Zecca, 2019).

No entanto, o grande volume de dados gerado por tecnologias digitais na agricultura exige recursos avançados de armazenamento, edição, análise e interpretação. As técnicas convencionais de processamento de dados não são suficientes para atender às demandas dessa nova era da agricultura inteligente, tornando-se um obstáculo para a extração de informações valiosas do campo (Evstatiev; Gabrovska-Evstatieva, 2020). Para superar esse desafio, o Machine Learning (ML) emergiu como uma solução viável, sendo um subconjunto da inteligência artificial que permite aproveitar o crescimento exponencial da capacidade computacional para aprimorar as análises agrícolas e a tomada de decisão baseada em dados (Helm et al., 2020).

### **3.4.1 Machine learning**

A utilização do Machine Learning (ML) constitui uma das metodologias empregadas no campo da inteligência artificial, utilizando algoritmos baseados em matemática e estatística para realizar tarefas de aprendizado de máquina. Esse processo visa extrair informações significativas dos dados, representando-as por meio de um modelo matemático que pode ser empregado para inferências ou previsões com base em novos conjuntos de dados (Machado; Mundim, 2023).

O ML tem se destacado como uma ferramenta essencial no aprimoramento da Agricultura 4.0. Ele pode ser definido como um sistema computacional capaz de aprender e executar tarefas específicas sem a necessidade de programação explícita para cada situação (Ahmed; Pathan, 2018; Liakos et al., 2018; Mahesh, 2020). Esse processo envolve a análise de múltiplas entradas de dados, permitindo que o computador identifique padrões e tome decisões de forma autônoma (Liakos et al., 2018).

Quando adequadamente modeladas, as técnicas de ML fornecem uma melhor

capacidade de processamento e análise de dados, facilitando previsões e estimativas (Santos, 2020). A previsão difere da estimação na medida em que a estimação se refere ao cálculo de uma condição atual de várias variáveis, enquanto a previsão busca quantificar um provável evento futuro (Santos; Santos; Rolim, 2021b).

Nos últimos anos, os modelos de ML têm sido amplamente utilizados para prever a produtividade dos cultivos. O modelo Random Forest, por exemplo, é uma técnica de ML que combina vários algoritmos de árvore de decisão e tem sido aplicado com sucesso na previsão de safras. Já os modelos de regressão linear múltipla utilizam várias variáveis independentes, aumentando a precisão na previsão da produtividade dos cultivos (Meneses, 2020). Outro exemplo são as Redes Neurais Artificiais (RNAs), modelos de aprendizagem profunda que podem ser treinados para analisar e interpretar dados complexos, incluindo segurança alimentar e previsões de características físicas e químicas (Silva et al., 2022).

O desempenho do ML está diretamente relacionado à quantidade e qualidade dos dados disponíveis, uma vez que a precisão dos modelos aumenta com bases de dados robustas e diversificadas. No entanto, no contexto agrícola, a obtenção de grandes volumes de dados confiáveis pode ser um desafio, sendo um fator crucial para garantir a eficácia das previsões e otimizações geradas pelos algoritmos de ML.

Com a evolução da inteligência artificial e do Machine Learning, novas abordagens tecnológicas têm sido incorporadas a culturas agrícolas específicas. No setor cafeeiro, essas inovações vêm sendo aplicadas dentro do conceito de Cafeicultura Digital, que integra tecnologias como Agricultura de Precisão, sensoriamento remoto, Big Data e Internet das Coisas (IoT) para aprimorar a tomada de decisão no cultivo do café. Essa modernização possibilita uma gestão mais eficiente dos recursos e contribui para o aumento da produtividade e qualidade dos grãos.

### **3.4.2 Cafeicultura digital**

As tecnologias digitais foram retratadas como uma solução promissora para enfrentar vários desses desafios. Foi proposto que essas tecnologias podem permitir que os produtores usem os recursos de forma mais eficiente, aumentem a produtividade e reduzam a pegada ambiental por meio da gestão baseada em dados (Sharma et al., 2020).

Nas literaturas encontradas, poucas pesquisas são relacionadas a aplicação de técnicas de agricultura digital na cafeicultura (Barbosa et al., 2021; Santhosh; Umesh, (2022 e 2023); Bolaños; Corrales; Campo, 2023; Kittichotsawat; Tippayawong; Tippayawong, 2022;

Zanella et al., 2024; Faria et al., 2024), além disso são estudos muito recentes e datados a partir do ano 2022 até o presente momento.

No contexto da aplicação de tecnologias digitais na cafeicultura, um estudo relevante foi conduzido por Barbosa et al. (2021), que avaliaram o uso de RPA equipada com câmera RGB e algoritmos de visão computacional para estimar a previsão de rendimento do café. Os autores coletaram dados de índice de área foliar (LAI), altura da planta, diâmetro da copa e valores individuais extraídos das bandas RGB. Esses dados foram utilizados para desenvolver modelos de previsão de rendimento por meio de cinco algoritmos. Com base nos resultados, os autores afirmam que o algoritmo neuro evolução de topologias de aumento (NEAT) apresentou o melhor desempenho (MAPE de 31,75%), sendo o LAI e o diâmetro da copa as variáveis mais relevantes para a construção do modelo preditivo.

Com o intuito de explorar abordagens preditivas baseadas em dados históricos, Santhosh e Umesh (2022) analisaram informações coletadas entre os anos de 2008 e 2019, referentes a cafeeiros cultivados no estado de Karnataka. Os dados incluíam variáveis agrônômicas como idade, nutrientes do solo e rendimento. A partir desse conjunto, os autores avaliaram o desempenho de três algoritmos de regressão na construção de modelos preditivos de rendimento. Os resultados indicaram que a melhor estimativa foi obtida por meio do modelo de regressão de rede elástica (ENET).

No ano seguinte, Santhosh e Umesh (2023) propuseram uma estrutura baseada em aprendizado de máquina para prever o rendimento do café, considerando a influência de fatores agrônômicos nas safras. Para isso, foram utilizados dados de idade, nutrientes do solo (carbono orgânico, fósforo, potássio, pH), zona e o respectivo rendimento obtido na região de Chikkamagaluru, Karnataka, Índia. Segundo os autores, os modelos desenvolvidos com os algoritmos Extra Tree Classifier e Random Forest apresentaram precisão de 91% segundo as métricas de validação.

Conforme descreveram Bolanõs, Corrales & Campo (2023), modelos de previsão de rendimento nas fases iniciais do ciclo produtivo do café foram construídos a partir de imagens multiespectrais obtidas por um veículo aéreo não tripulado. Para esse estudo, foram utilizadas variáveis como altura, diâmetro, área lateral, índice de área foliar e diversos índices de vegetação. Os autores identificaram que os atributos de volume do cafeeiro e os índices NDVI e CRI apresentaram as maiores correlações com o rendimento. Utilizando esses preditores e os algoritmos de Regressão Linear e Regressão de Gradiente Descendente Estocástica, obtiveram  $R^2$  de 56% e 55%, respectivamente.

Em um estudo semelhante, Kittichotsawat, Tippayawong & Tippayawong (2022)

investigaram o uso de Redes Neurais Artificiais (RNA) e Regressão Linear Múltipla (RLM) para prever o rendimento do café arábica. Foram utilizadas informações coletadas ao longo de 180 meses (2004–2018) sobre áreas, zonas de produtividade, precipitação, umidade relativa e temperaturas mínima e máxima. Os resultados indicaram que o algoritmo RNA obteve  $R^2$  de 95,24% e RMSE de 0,0784 toneladas, evidenciando seu potencial para estimar os rendimentos.

Já Zanella et al. (2024) exploraram a combinação de dados de bandas espectrais, índices de vegetação e nutrientes foliares para desenvolver modelos preditivos da produtividade do café. Os dados foram obtidos por meio do satélite RapidEye, e as análises foram realizadas com técnicas de regressão linear múltipla. Os modelos apresentaram valores de  $R^2$  entre 0,57 e 0,72, e RMSE entre 23,78 e 19,39 sacos  $ha^{-1}$ .

Por fim, Faria et al. (2024) buscaram prever a produtividade em cafezais utilizando aprendizado de máquina com base em atributos do solo e da planta. Foram considerados dados de taxas variadas de fertilização e mapas de análise foliar de duas safras consecutivas. Os autores destacam que o modelo com melhor desempenho alcançou uma correlação de Pearson de 0,86, e que os atributos químicos do solo não influenciaram significativamente os resultados preditivos.

Considerando a importância da previsão de produtividade dos cafeeiros, e como está ligada a diversas características, o que gera um grande volume de dados e informações, o emprego de técnicas de aprendizado de máquinas torna-se imprescindível para a avaliação da produtividade dos cafeeiros, viabilizando o manejo das lavouras de maneira adequada e minimizar os impactos gerados pela produção de café.

## REFERÊNCIAS

- AHMAD, A., et al. Remotely Piloted Aircraft (RPA) in Agriculture: A Pursuit of Sustainability. **Agronomy**, v.11, n.7, 2020.
- ALMEIDA, L. S.; GUIMARÃES, E. C. Geoestatística e análise fatorial exploratória para representação espacial de atributos químicos do solo, na cafeicultura. **Coffee Science**, v. 11, n. 2, p. 195-203, 2016.
- ANDRADE, A. R. S. D., et al. Variabilidade espacial da densidade do solo sob manejo da irrigação. **Ciência e Agrotecnologia**, v.29, 322-329, 2005.
- ASSIS, P. H. S., MARTINS, G.D., ORLANDO, V.S.W. Mapeamento de parâmetros agronômicos do cafeeiro a partir de imagens tomadas por aeronave remotamente pilotada: mapping coffee agronomic parameters based on remotely piloted aircraft images. **Boletim Goiano de Geografia**, v. 43, n. 01, 2023.
- AVELINO, J., et al., The coffee rust crises in Colombia and Central America (2008–2013): Impacts, plausible causes and proposed solutions. **Food Secur.** v.7, 303–321, 2015.
- BACON, C. M., et al. Vulnerability to cumulative hazards: Coping with the coffee leaf rust outbreak, drought, and food insecurity in Nicaragua. **World Dev.** v.93, 136–152, 2017.
- BARATA, R. A. P. et al. UAV-based vegetation indices to evaluate coffee crop response after transplanting seedlings grown in different containers. **Agriculture**, v. 14, n. 3, p. 356, 2024.
- BARBOSA, B. D. S., et al. Application of rgb images obtained by uav in coffee farming. **Remote Sensing**, v.13, n.12, 2397, 2021.
- BATISTELLA, M., et al. Geotecnologias e gestão territorial da bovinocultura no Brasil. **Revista Brasileira de Zootecnia**, v. 40, p. 251-260, 2011.
- BENTO, N. L., et al. Evaluation of coffee plant attributes by field collection and remotely piloted aircraft system images. **Spanish Journal of Agricultural Research**, v.20, n.3, e0205-e0205, 2022.
- BENTO, N. L. et al. Coffee Growing with Remotely Piloted Aircraft System: Bibliometric Review. **AgriEngineering**, v. 5, n. 4, p. 2458-2477, 2023a.
- BENTO, N. L., et al. Weed detection and mapping of a coffee farm by a remotely piloted aircraft system. **Agronomy**, v.13, n.3, 830, 2023b.
- BOLAÑOS, J., CORRALES, J. C., CAMPO, L. V. Feasibility of Early Yield Prediction per Coffee Tree Based on Multispectral Aerial Imagery: Case of Arabica Coffee Crops in Cauca-Colombia. **Remote Sensing**, v.15, n.1, 282, 2023.
- BONNAIRE RIVERA, L., MONTOYA BONILLA, B., OBANDO-VIDAL, F. Processing multispectral imaging captured by drones to evaluate the normalized difference vegetation index of Castillo coffee plantations. **Ciencia y Tecnología Agropecuaria**, v.22, n.1, 2021.

- BRANDÃO, Z. N. Estimativa da produtividade e estado nutricional da cultura do algodão irrigado via técnicas de sensoriamento remoto. 2009. 152 f. Tese (Doutorado em Recursos Naturais) -Universidade Federal de Campina Grande, Campina Grande, 2009.
- CADENAS, J. M., et al. Making decisions for frost prediction in agricultural crops in a soft computing framework. **Computers and Electronics in Agriculture**, v. 175, p. 105587, 2020.
- CARVALHO, L. G. et al. A regression model to predict Coffee productivity in Southern Minas Gerais, Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 8, n. 2/3, p. 204-211, 2004.
- CARVALHO, G. R. et al. Avaliação de produtividade de progênies de cafeeiro em dois sistemas de plantio. **Ciência e Agrotecnologia**, v. 30, n. 5, p. 838-843, 2006.
- CEZANA, D. C., et al. Spatio-temporal variability of leaf macronutrients in a conilon coffee crop. **Brazilian Journal of Production Engineering**, v.10, n.1, 178-187.2024.
- CHALFOUN, S. M. et al. Relação entre diferentes níveis de infecção de ferrugem (*Hemileia vastatrix* Berk. & Br.) e produção dos cafeeiros (*Coffea arabica* L.) em algumas localidades de Minas Gerais. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEEIRAS, 6., 1978, Ribeirão Preto. Resumos... Rio de Janeiro, 1978. p. 392-394.
- CUNHA, J. P. A. R.; SIRQUEIRA NETO, M. A.; HURTADO, S. M. C. Estimating vegetation volume of coffee crops using images from unmanned aerial vehicles. **Engenharia Agrícola**, v. 39, n. SPE, p. 41-47, 2019.
- FARIA, R. O. de., et al. Models for predicting coffee yield from chemical characteristics of soil and leaves using machine learning. **Journal of the Science of Food and Agriculture**, v.104, n.9, 5197-5206, 2024.
- FERRAZ, G. A. S., Silva, F. M. D., Carvalho, L. C., Alves, M. D. C., & Franco, B. C. (2012). Spatial and temporal variability of phosphorus, potassium and of the yield of a coffee field. **Engenharia Agrícola**, v.32, 140-150, 2012.
- FERRAZ, G. A. E. S., et al. Variabilidade espacial dos atributos da planta de uma lavoura cafeeira. **Revista Ciência Agronômica**, v.48, n.1, 81-91, 2017
- FERRAZ, G. A. e S. Plant sampling grid determination in precision agriculture in coffee field. **Coffee Science**, v. 13, n. 1, p. 112–121, 2018.
- FILIPPINI ALBA, J. M. Modelagem SIG em agricultura de precisão: conceitos, revisão e aplicações. In: BERNARDI, A. C. C. et al. (org.). *Agricultura de precisão: resultados de um novo olhar*. Brasília, DF: Embrapa, 2014. p. 84-95.
- GOIGOCHEA-PINCHI, D., et al. Yield prediction models for rice varieties using UAV multispectral imagery in the Amazon lowlands of Peru. **AgriEngineering**, v. 6, n. 3, p. 2955-2969, 2024.
- GONÇALVES, A. C., SOUSA, A. M. O. **Tecnologias de Informação Geográfica. ARCGIS**

10. Textos didáticos. Universidade de Évora. 67 p, 2017.

JAWO, T. O., KYEREH, D., & LOJKA, B. The impact of climate change on coffee production of small farmers and their adaptation strategies: a review. **Climate and Development**, v.15, n.2, 93-109, 2023.

JIMÉNEZ-BRENES, F.M. et al. Automatic UAV-based detection of *Cynodon dactylon* for site-specific vineyard management. **PLoS ONE**, v.14, e0218132, 2019.

KITTICHOTSATSAWAT, Y., TIPPAYAWONG, N., TIPPAYAWONG, K. Y. Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. **Scientific Reports**, v.12, n.1, 14488, 2022.

MACHADO, H. G.; MUNDIM, K. Revisão das Tecnologias de Inteligência Artificial e Machine/Deep Learning: Restrições, Oportunidades, Estado da Arte e Desafios. **Revista Processos Químicos**, v. 16, n. 32, p. 9-22, 2022.

MANZIONE, R. L., RODRIGUES, J. B. T., ZIMBACK, C. R. L. Análise espacial multivariada na avaliação de parâmetros químicos do solo. In: BALASTREIRE, L. A. Avanços na Agricultura de Precisão no Brasil no período de 1999-2001. Piracicaba, 2002. 347 p.

MANLOVE, J.L., SHEW, A.M., OBEMBE, O.S. Arkansas Producers Value Upload Speed More than Download Speed for Precision Agriculture Applications. **Comput. Electron. Agric.**, 190, 106432, 2021.

MANZIONE, R. L., ZIMBACK, C. R. L. Análise espacial multivariada aplicada na avaliação da fertilidade do solo. **Engenharia na Agricultura**, v. 19, n. 3, p. 227-235, 2011.

MARASCA, I., et al. Geoestatística aplicada na determinação da compactação do solo no cultivo da cana-de-açúcar. **Revista de Agricultura Neotropical**, v. 4, n. 2, p. 49-55, 2017.

MARIN, D. B. Remotely piloted aircraft in spatial multispectral modeling of stressors in coffee plantations. 2021. 106 f. Tese (Doutorado em Engenharia Agrícola) – Universidade Federal de Lavras, Lavras, 2021.

MARIN, D. B., et al. Remotely piloted aircraft and random forest in the evaluation of the spatial variability of foliar nitrogen in coffee crop. **Remote Sensing**, v.13, n.8, 1471, 2021.

MASSRUHÁ, S. M. F. S. et al. (org.). *Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas*. Brasília, DF: Embrapa, 2020.

MENESES, K. C. D. Estimating Potential Evapotranspiration in Maranhão State Using Artificial Neural Networks. [S.l.]: **Revista Brasileira de Meteorologia**, 2020.

MOLIN, J. P., DO AMARAL, L. R., COLAÇO, A. Agricultura de precisão. Oficina de textos, 2015.

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

MULLA, D. J. Twenty-five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. **Biosystems engineering**, v.114, n.4, 358-371, 2013.

MURUGAN, Deepak; GARG, Akanksha; SINGH, Dharmendra. Development of an adaptive approach for precision agriculture monitoring with drone and satellite data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, n. 12, p. 5322-5328, 2017.

NIE, P., et al. Hybrid combination of GIS,GPS,WSN and GPRS technology in modern digital agriculture application. **Adv. Mater. Res.**, v.108, 1158–1163, 2010.

NOGUEIRA MARTINS, R., et al. A novel vegetation index for coffee ripeness monitoring using aerial imagery. **Remote Sensing**, v.13, n.2, 263, 2021.

OLIVEIRA, H. C.; GUIZILINI, V. C.; NUNES, I. P.; SOUZA, J. R. Failure Detection in Row Crops From UAV Images Using Morphological Operators. **IEEE Geoscience and Remote Sensing Letters**, 2018.

PACCIORETTI, P.; CÓRDOBA, Mariano; BALZARINI, Mónica. FastMapping: Software to create field maps and identify management zones in precision agriculture. **Computers and Electronics in Agriculture**, v. 175, p. 105556, 2020.

PEREIRA, L. S., SILVA, D. O., PAMBOUKIAN, S. V. D. Sensoriamento remoto aplicado à agricultura de precisão no cultivo de bambu. **Revista Mackenzie de Engenharia e Computação**, v. 16, n. 1, p. 8-33, 2016.

QUEIROZ, Daniel Marçal et al. (Ed.). **Agricultura digital**. Oficina de Textos, 2022.

RENA, A. B. et al. Fisiologia do cafeeiro em plantios adensados. In: SIMPÓSIO INTERNACIONAL SOBRECAFÉ ADENSADO, 2., 1994, Londrina. Anais... Londrina:Instituto Agronômico do Paraná, 1996. p. 73-85.

RHINEY, K., KNUDSON, C., GUIDO, Z. Cultivating crisis: Coffee, smallholder vulnerability, and the uneven sociomaterial consequences of the leaf rust epidemic in Jamaica. **Ann. Am. Assoc. Geogr.**, v.11, 1–19, 2020.

ROSALEN, D. L. Challenges of digital agriculture in pest management. In: Nascimento, J. et al. Topics in Agricultural Entomology - XIII. Ponta Grossa: Atena Editora, 2022. p. 123 133.

SALCEDO-SANZ, Sancho et al. Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. **Information Fusion**, v. 63, p. 256-272, 2020.

SANCHEZ, R. B., et al. Variabilidade espacial de propriedades de Latossolo e da produção de café em diferentes superfícies geomórficas. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.9, 489-495, 2005.

SANTANA, L. S. et al. Vegetative vigor of maize crop obtained through vegetation indexes in orbital and aerial sensors images. **Revista Brasileira de Engenharia de Biosistemas**, v. 13,

n. 3, p. 195–206, 2019.

SANTANA, L. S., et al. Monitoring errors of semi-mechanized coffee planting by remotely piloted aircraft. **Agronomy**, v. 11, n. 6, p. 1224, 2021.

SANTANA, L. S., et al. Residual Ash Mapping and Coffee Plant Development Based on Multispectral RPA Images. **Remote Sensing**, v.16, n.11, 1917, 2024.

SANTOS, Luana Mendes dos et al. Use of remotely piloted aircraft in precision agriculture: A review. **Dyna**, v. 86, n. 210, p. 284-291, 2019a.

SANTOS, L. M. D. et al. Analysis of flight parameters and georeferencing of images with different control points obtained by RPA. **Agronomy Research**, v. 17, n. 5, p. 2054–2063, 2019b.

SANTOS, L. M.; FERRAZ, G. A. S.; BARBOSA, B. D. S.; DIOTTO, A. V.; MACIAL, D. T.; XAVIER, L. A. G. Biophysical parameters of coffee crop estimated by UAV RGB images. **Precision Agriculture**, p. 1-15, 2020.

SANTOS, S. A. et al. Supervised classification and NDVI calculation from remote piloted aircraft images for coffee plantations applications. **Coffee Science-ISSN 1984-3909**, v. 16, p. e161978-e161978, 2021a.

SANTOS, V. B. D.; SANTOS, A. M. F. D.; ROLIM, G. D. S. Estimation and forecasting of soybean yield using artificial neural networks. **Agronomy Journal**, 113, n. 4, Julho 2021b. 3193-3209 p.

SANTOS, S. A. D., et al. Evaluation of the water conditions in coffee plantations using RPA. **AgriEngineering**, v.5, n.1, 65-84, 2022a.

SANTOS, L. M. D., et al. Use of RPA Images in the Mapping of the Chlorophyll Index of Coffee Plants. **Sustainability**, 14(20), 13118, 2022b.

SANTOS, L. M. D., et al. Vegetation indices applied to suborbital multispectral images of healthy coffee and coffee infested with coffee leaf miner. **AgriEngineering**, 4(1), 311-319, 2022c.

SANTOS, L. M., et al. Use of Images Obtained by Remotely Piloted Aircraft and Random Forest for the Detection of Leaf Miner (*Leucoptera coffeella*) in Newly Planted Coffee Trees. **Remote Sensing**, 16(4), 728, 2024.

SANTHOSH, C. S., UMESH, K. K. (2022, December). A Compendium Probabilistic Prospective for Predicting Coffee Crop Yield Based on Agronomical Factors. In 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 1-8). IEEE.

SANTHOSH, C. S., UMESH, K. K. An ensemble approach for coffee crop yield prediction based on agronomic factors. **ASEAN Engineering Journal**, v. 13, n. 3, p. 29-38, 2023.

SHARMA, R., et al. A systematic literature review on machine learning applications for

sustainable agriculture supply chain performance. **Computers & Operations Research**, v.119, 104926, 2020.

YIN, Zhichao; SHI, Jianping. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. 2018. p. 1983-1992.

SILVA, F. M. D., et al. Variabilidade espacial de atributos químicos e de produtividade na cultura do café. **Ciência Rural**, v. 37, 401-407, 2007.

SILVA, F. M. D., et al. Variabilidade espacial de atributos químicos e produtividade da cultura do café em duas safras agrícolas. **Ciência e Agrotecnologia**, v.32, 231-241. 2008.

SILVA, S. D. A., et al. Spatial variability of chemical attributes of an Oxisol under coffee cultivation. **Revista Brasileira de Ciência do Solo**, v.34, 15-22., 2010.

SILVA, S. D. A., LIMA, J. S. D. S., SOUZA, G. S. D. Estudo da fertilidade de um Latossolo Vermelho-Amarelo húmico sob cultivo de café arábica por meio de geoestatística. **Revista Ceres**, v.57, 560-567, 2010.

SILVA, S. D. A., LIMA, J. S. D. S. Avaliação da variabilidade do estado nutricional e produtividade de café por meio da análise de componentes principais e geoestatística. **Revista Ceres**, v.59, 271-277, 2012.

SILVA, A. G. D. et al. Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. **Scientific Reports**, 12, n. 1, 2022.

SOUZA, V. C. O. de; VIEIRA, T. G. C.; ALVES, H. M. R. Uso do Sistema de Informação Geográfica para a implementação de um banco de dados da cafeicultura mineira e sua divulgação via web. In: EMPRESA DE PESQUISA AGROPECUÁRIA DE MINAS GERAIS. CENTRO TECNOLÓGICO SUL DE MINAS – LABORATÓRIO DE GEOPROCESSAMENTO, Minas Gerais, 2004.

SOUZA, Z. M. et al. Variabilidade espacial do pH, Ca, Mg e V% do solo em diferentes formas do relevo sob cultivo de cana-de-açúcar. **Ciência Rural**, v. 34, n. 6,p. 1763-1771, 2004.

SOUZA, K. X. S. de et al. Agricultura digital: definições e tecnologias. *Embrapa Informática Agropecuária*, 2020.

TOLEDO, S. V. de, BARROS, I. de. Influência da densidade de plantio e sistema de podas na produção de café. **Pesquisa Agropecuária Brasileira, Brasília**, v. 34, n. 8, p.1379-1384, 1999.

VALENTE, D. S., et al. Variabilidade espacial da condutividade elétrica aparente e das propriedades do solo em área de produção de café. **Engenharia Agrícola**, v.34, 1224-1233, 2014.

WAN, L., et al. Grain Yield Prediction of Rice Using Multi-Temporal UAV-Based RGB and Multispectral Images and Model Transfer—A Case Study of Small Farmlands in the South of China. **Agric. For. Meteorol.**, v.291, 108096, 2020.

YOST, M.A., et al. A long-term precision agriculture system sustains grain profitability. **Precis. Agric.**, v.20, 1177–1198, 2019.

ZANELLA, M. A., et al. Coffee yield prediction using high-resolution satellite imagery and crop nutritional status in Southeast Brazil. **Remote Sensing Applications: Society and Environment**, v.33, 101092, 2024.

ZHANG, Junzhe et al. UAV-derived imagery for vegetation structure estimation in rangelands: validation and application. **Ecosphere**, v. 12, n. 11, p. e03830, 2021.

## SEGUNDA PARTE - ARTIGOS

Artigo 1 – Elaborado conforme as normas do periódico	
Título do artigo	Spatial and temporal variability of productivity of coffee plants grown in an experimental field located in Três Pontas, Brazil.
Autores	S.A. Santos, G.A.S. Ferraz, V.C. Figueiredo, M.M.L. Volpato, C.S.M. Matos, A.B. Pereira, L. Conti, G. Bambi and D.B. Marin
Periódico	Agronomy Research
ISSN	1406-894X
DOI	<a href="https://doi.org/10.15159/ar.23.080">https://doi.org/10.15159/ar.23.080</a>

Fonte: elaboração própria.

### **Spatial and temporal variability of productivity of coffee plants grown in an experimental field located in Três Pontas, Brazil**

S.A. Santos<sup>1\*</sup>, G.A.S. Ferraz<sup>1</sup>, V.C. Figueiredo<sup>2</sup>, M.M.L. Volpato<sup>2</sup>, C.S.M. Matos<sup>2</sup>, A.B. Pereira<sup>2</sup>, L. Conti<sup>3</sup>, G. Bambi<sup>3</sup> and D.B. Marin<sup>3</sup>

<sup>1</sup>Federal University of Lavras, Department of Agricultural Engineering, Trevo Rotatório Professor Edmir Sá Santos, BR37200-900, Lavras, Brazil

<sup>2</sup>Agricultural Research Company of Minas Gerais, Av. José Cândido da Silveira 1647, Bairro União Belo Horizonte, BR31170-495 Belo Horizonte, Brazil

<sup>3</sup>University of Florence, Department of Agricultural, Food and Forestry Science and Technology, Piazza S. Marco 4, IT055 27571, Florence, Italy

\*Correspondence: [sthefany.santos1@estudante.ufla.br](mailto:sthefany.santos1@estudante.ufla.br)

Received: January 31st, 2023; Accepted: April 25th, 2023; Published: September 5th, 2023

**Abstract.** The coffee grower seeks to increase productivity, as well as reduce the operating costs of his crop. Precision Agriculture (PA) is composed of a cycle of tools and technologies that can bring a good return to coffee growers, seeking to optimize production processes, bringing better yields and minimizing costs. Therefore, the objective of this research is to evaluate the space-time behavior of productivity in a coffee plantation, aiming to apply AP techniques. The study was carried out in a coffee plantation of the species (*Coffea arabica*), cultivar Topázio MG1190, located in the municipality of Três Pontas, Brazil, with an area of 1.2 ha. With the aid of a GNSS RTK, 30 plants were georeferenced, from which their yields were later sampled in the years 2020, 2021 and 2022. The collected data were evaluated in two statistical processes in the RStudio software. The first stage consisted of a one-way analysis of variance with repeated measures, from the results it is concluded that there are differences between the production averages when buying the productivity of the years 2020, 2021 and 2022 and, in addition, the coefficient of variation for the three sets of samples was quite high ( $CV > 30\%$ ) indicating a heterogeneity between the data. The second stage consisted of a geostatistical analysis the data were fitted in a model and interpolated by ordinary kriging; the result was maps of spatial variability. Through these maps it was possible to evaluate the behavior of productivity spatially and temporally, as well as to quantify areas that had higher and lower levels of this attribute. It is concluded that productivity, even in the case of such a small productive area, can vary substantially in space and time, and the use of PA can help producers in decision making regarding management.

**Keywords:** biennial, geostatistics, precision agriculture in coffee trees

## INTRODUCTION

The coffee agro-industrial system is the focus of several studies due to its importance in the economic and social history of Brazil since the colonial period. This commodity played an important role in the accumulation of foreign exchange that started the country's industrialization process (Cardozo et al., 2019).

Coffee cultivation in Brazil occurs in environments with great diversity (Alves et al., 2011) and this cultivation is very sensitive to climatic conditions (Aparecido et al., 2015). Climate variability has a strong impact on agricultural activities (Sá Junior et al., 2012), being the main factor responsible for fluctuations and oscillations in coffee bean productivity (Camargo, 2010).

According to Santana et al. (2022) the Precision Agriculture applied to coffee production still needs to be developed and implemented. But the same authors indicate that there is still a propensity to research, defunded, and adopted due to the benefits and adopted due to the benefits it can bring as efficiency, environmental and economic sustainability.

Productivity maps are one of the precision agriculture tools and are an excellent indicator of problems that may occur during the productive cycle of a crop. Failure to use tools like these leads the producer to consider mean values for decision-making in relation to management, which can lead to decisive errors. Even if the crop is treated (fertilizing, spraying, controlling pests and diseases, etc) throughout this cycle, productivity is a factor that will not be uniform, as this is an attribute that is influenced by factors such as: biennial (Camargo & Camargo, 2001), soil fertility and leaf nutrition (Wadt & Dias, 2012; Scalco et al., 2014) occurrence of pests, diseases and weeds (Fialho et al., 2010; Carvalho et al., 2012, Lopes et al., 2012), physical attributes of the plant (Carvalho et al., 2013; Burak et al., 2016; Santos et al., 2023) among other factors.

Taking into account that coffee production is strongly influenced by several factors which may lead to a high rate of spatial variability of this attribute, in this context, Precision Agriculture (PA) arises with the objective of evaluating and quantifying this spatial variability. PA when applied to coffee growing is called Precision Coffee Growing (Ferraz et al., 2012a). Precision Agriculture (PA) approaches seek to understand the spatial and temporal variability present in agricultural variables within a production field (Martello et al., 2022).

The need to modernize agricultural production has encouraged more and more producers to adopt agricultural practices within the PA context (Dong et al., 2013). A more up-to-date definition of precision coffee farming was described by Santana et al. (2022) who define it as updated techniques and technologies use that aim to maximize crop profitability, increase operations efficiency, search for business sustainability, environmentally sustainable production, and unceasing search for maximizing productivity and improving final product quality.

Ferraz et al. (2017) state that geostatistics is an important methodology for data analysis, being used as a tool within PA to analyze the factors involved in production systems (Carvalho et al., 2013). Through geostatistical analysis, it is possible to identify whether there is spatial dependence for the analyzed factors, enabling the creation of thematic maps that help in decision making (Carvalho et al., 2013). The use of the GIS environment has great potential for accurate and rapid research, bringing clear results (Piri et al., 2019) especially when it comes to assessing climate variability and its consequences on productivity.

It is important to mention that the application of PA techniques and technologies in coffee growing is recent, mainly when it comes to the evaluation of heterogeneity of factors related to soil and plants, and bearing in mind that few studies evaluate small-scale productive crops. In this context, this work aims to evaluate the spatial variability of the productivity obtained for three consecutive years, 2020, 2021 and 2022 and, in this way, if the spatial dependence is proven, generate maps through interpolation by kriging, which will establish values of this attribute at unsampled locations within the study area.

## MATERIALS AND METHODS

### Workflow

The workflow of this research (Figure 1) was defined in the following stages: construction of the sampling grid, georeferencing of points, productivity sampling and geostatistical analysis.



Figure 1. Work flowchart

### Description of the Study Area

The study was conducted in a coffee plantation of the Experimental Field of the Minas Gerais Agricultural Research Company (EPAMIG, as abbreviated in Portuguese), located in the municipality of Três Pontas in the southern region of the state of Minas Gerais, Brazil, at an altitude of 905 m and a Universal Transverse Mercator (UTM) coordinate system position of 7640030.4 S and 449531.5E, Zone 23K. This municipality has a mean annual temperature of 20.3 °C and a mean annual rainfall of 1429 mm (Climate Data, 2023).

The soil in this area is classified as oxisol. The area of the experiment comprised 1.2 ha of a coffee plantation of the species *Coffea arabica* L. of the Topázio MG1190 cultivar (Figure 2). This crop was established in 1998 with spacings between rows of 3.70 m and between plants of 0.70 m (EPAMIG, 2023).

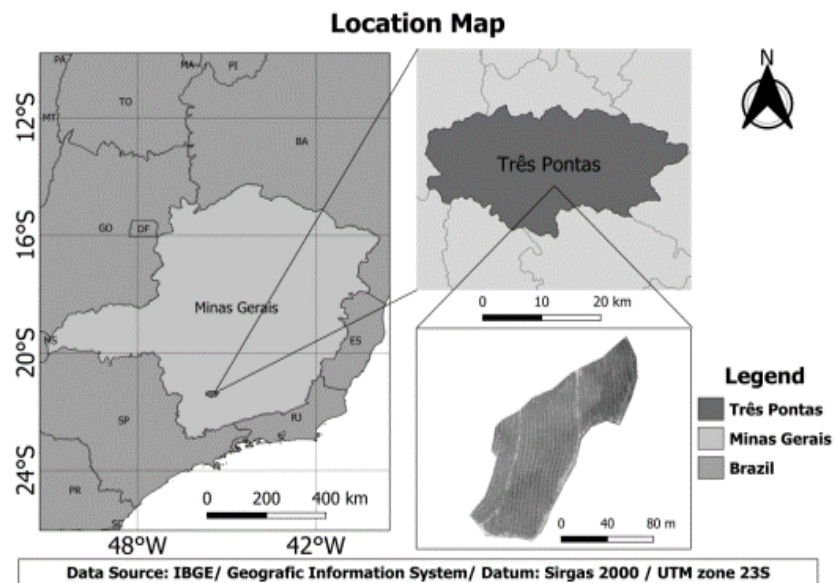
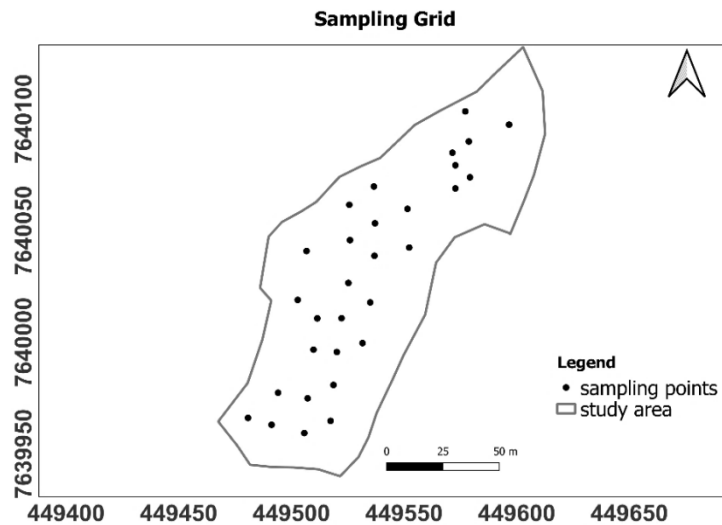


Figure 2. Location map

### Sampling grid

The sampling grid was prepared with QGIS software, version 3.4.8, and 30 sampling points representing a sampling density of 25 points per hectare were laid out (Figure 3).



**Figure 3.** Equidistant sampling grid used for sampling the 30 georeferenced points in this study

For the construction of the sampling mesh, the methodology of the equidistant sampling mesh proposed by Faria (2019) was used, this method consists of reducing the walking within the plot using walking routes. Within the sampling of 30 points, the smallest distance between them was 6m and the greatest distance was 175m.

For this study, the productivity data of the 30 georeferenced plants were sampled on the following dates: 1st collection (30 June 2020), 2nd collection (08 June 2021) and 3rd collection (13 June 2022).

### Productivity sampling

Productivity data were sampled at 30 georeferenced points within the study area. For each sampling point represented by 1 coffee plant, productivity was obtained through semi-mechanized harvesting using a derriçadora (Figure 4a). After harvesting each sampling point, the fruits that fell under the cloths were selected to separate them from leaves and branches (Figure 4b), and then they were deposited in a graduated container, which will indicate the productivity in liters per plant (L /plant) (Figure 4c).



**Figure 4.** Productivity sampling (a) semi-mechanized harvesting, (b) separation of leaves and branches and (c) productivity measurement in L/plant in graduated bucket.

### Statistical analysis

Taking into account that during the three years of sampling, the same 30 plants were sampled, in this way, as the dataset represents a group of samples where repeated measurements of productivity were obtained during a time interval (2020, 2021 and 2022), The one-way analysis of variance (ANOVA) statistical test with repeated measures was used in order to evaluate the mean productivity of each group as a whole. For this analysis, tests were first performed to verify outliers in the samples and then a Shapiro Wilk test to verify data normality.

Then, the analysis of variance model was built using the variables year (independent variable) and productivity (dependent variable), the result of this model is a table containing: degree of freedom of the evaluated variable, degree of freedom of error, sum of squares, F-value and p-value. For this F test to be valid, it is necessary to verify the sphericity of the data, and this will be done using the Mauchly test, this test considers the following hypotheses:

$H_{0\text{Mauchly}}$ : data are spherical when  $p > 0,05$ ;

$H_{1\text{Mauchly}}$ : data are not spherical when  $p \leq 0,05$ .

If sphericity is verified and the null hypothesis is accepted, the following hypotheses for the F test performed in the analysis of variance of a repeated measures way must be considered:

$H_{0\text{ANOVA}}$ : mean productivity 2020= mean productivity 2021= mean productivity in 2022;  $p > 0,05$

$H_{1\text{ANOVA}}$ : there is at least one difference between the means of productivity  $p \leq 0.05$

If the null hypothesis of the Mauchly test ( $H_{0\text{Mauchly}}$ ) is rejected and the alternative hypothesis ( $H_{1\text{Mauchly}}$ ) is accepted, it is necessary to correct the degrees of freedom of the variance analysis using the Greenhouse-Geisser method for the lack of sphericity, and from there, obtain a new p value, and only then apply the hypothesis analysis of the one-way ANOVA method with repeated measures. After performing the F test by one-way ANOVA with repeated measures and if the null hypothesis ( $H_{0\text{ANOVA}}$ ) of the test is rejected and the alternative hypothesis is accepted ( $H_{1\text{ANOVA}}$ ), that is, that there is at least one difference between the measurements of productivity it is necessary to verify what this difference is, through a pair comparison test using the Bonferroni adjustment, which will result in a comparison matrix between the mean productivity, where the following hypotheses must be considered:

$H_{0\text{Bonferroni}}$ : if  $p > 0.05$  there is no difference between the productivity means;

$H_{1\text{Bonferroni}}$ : if  $p \leq 0.05$  there is a difference between the productivity means.

And finally, to close the statistical analysis of the data, a descriptive statistics table was generated containing: productivity mean for the three years, standard deviation and coefficient of variation. Gomes & Garcia (2002) state that the variability of an attribute can be classified according to the magnitude of its coefficient of variation (CV), according to the authors, the CV is low when it is less than 10%, moderate when it is in the range from 10 to 20%, high when it is between 20 and 30% and very high when it is above 30%. All statistical analysis of the data was performed using the RStudio software using the dplyr, ez, reshape and rstatix libraries.

### Geostatistical analysis

Semivariograms were used to analyze the spatial dependence of the productivity. The semivariance is classically estimated by Equation (1) according to Vieira (2000):

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

Where  $N(h)$  is the number of experimental pairs of observations  $Z(x_i)$  and  $Z(x_i + h)$ , separated by a distance  $h$ . The semivariogram is represented by the graph  $\hat{\gamma}(h)$  versus  $h$ . From the fitting of a mathematical model to the calculated values of  $\hat{\gamma}(h)$ , the coefficients of the theoretical model are estimated for the semivariogram, called the nugget effect ( $C_0$ ), sill ( $C_0 + C_1$ ), and range ( $a$ ), as described by Bachmaier and Backes (2008).

The semivariograms were adjusted using the Ordinary Least Squares (OLS) method. The adjusted semivariogram model was the spherical one, since this is the most used in geostatistics studies related to soil and coffee culture (Grego & Vieira, 2005; Silva et al., 2007) mainly evaluating coffee

productivity (Ferraz et al., 2012a; Ferraz et al., 2012b; Carvalho et al., 2013, Carvalho et al., 2017 and Ferraz et al., 2017).

To verify that this model fits the cross-validation requirements, the mean error (ME) was calculated as described by Isaaks and Srivastava (1989). ME should be as close to zero as possible.

With the adjustment of the semivariograms, after the identification of the spatial variability, the data were interpolated by ordinary kriging. Thus, the variable was estimated in places where it was not sampled, which allowed us to visualize its distribution in space in the form of thematic maps.

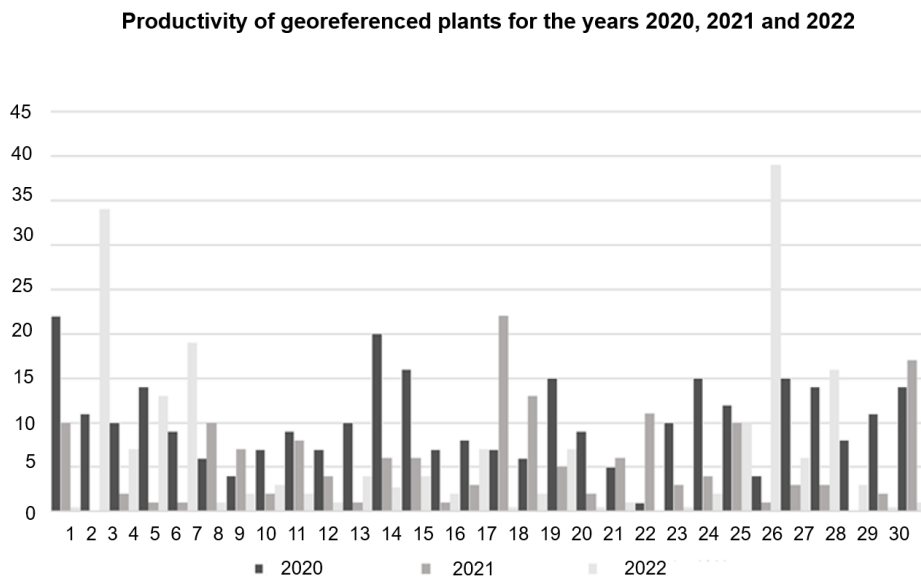
The calculation of the degree of spatial dependence (DSD) of the variables followed the classification proposed by Cambardella et al. (1994). In this classification, the authors point out that there is strong spatial dependence when the semivariogram shows a nugget effect equal to or less than 25% of the sill, moderate spatial dependence when this ratio is between 25% and 75%, and weak spatial dependence when it is greater than 75%.

The geoR package of the R software was used for geostatistical analysis and for the creation of the thematic maps (Ribeiro Junior and Diggle, 2001).

## RESULTS AND DISCUSSION

### Statistical analysis

Figure 5 represents a column chart containing the productivity data sampled in the 30 georeferenced plants in the years 2020, 2021 and 2022. It is observed that the highest values of productivity in coffee trees were found in the year 2022 for plants 25 (39 L/plant) and 2 (34 L/plant) respectively. The lowest productivity values were found for plants 2 and 8 in the year 2021 and for plant 21 in the year 2022, both with zero productivity.



**Figure 5.** Productivity in liters per plant (L/plant) obtained for the 30 plants georeferenced and sampled in the years 2020, 2021 and 2022.

Table 1 represents the results generated by the one-way analysis of variance with repeated measures, including the F test of hypotheses, as well as the Mauchly test to verify the sphericity of the data.

**Table 1.** One-way analysis of variance with repeated measures and Mauchly test to verify sphericity.

Variable	Analysis of variance					Mauchly's Test	
	DFn	DFd	SSn	F	P <sub>Anova</sub>	W	P <sub>Mauchly</sub>
Year	2	58	650.2889	7.6690	0.0011*	0.9271	0.3466

DFn: sample degrees of freedom, DFd: error degrees of freedom, SSn: sample sum of squares, SSd: error sum of squares, F: value of F, p ANOVA : probability of observation of the test F, W : statistics of the test, p<sub>Mauchly</sub> : observation probability of the Mauchly test , \*: significant at the 5% level of probability.

Through the analysis of variance modeling, it was observed that the  $p_{\text{Mauchly}}$  value was greater than 0.05 ( $p_{\text{Mauchly}} > 0.05$ ), that is, the Mauchly null hypothesis is accepted ( $H_{0\text{Mauchly}}$ ) that the data are spherical, so the hypotheses of the F test can be evaluated without a sphericity correction using the Greenhouse-Geisser method . The F test performed by analysis of variance indicated an  $p_{\text{ANOVA}} = 0.0011$  that is ( $p_{\text{ANOVA}} \leq 0.05$ ), thus rejecting the null hypothesis of the analysis of variance ( $H_{0\text{ANOVA}}$ ) and accepting the alternative hypothesis ( $H_{1\text{ANOVA}}$ ) stating that there is at least one difference between the average productivity in the coffee tree between the years of 2020, 2021 and 2022. To quantify this difference, a comparison between pairs was performed using the Bonferroni test, which resulted in a matrix (Table 2) for comparing means by pairs.

**Table 2.** Bonferroni test for comparing means between pairs.

Year	2020	2021
2021	0.2477	-
2022	0.0005*	0.1808

\*: significant at the 5% probability level

The results show that between the years 2021/2020 and 2022/2021 there were no differences between the productivity means ( $p > 0.05$ ), while between the years 2022/2020 there was a difference between the productivity means, that is, the average productivity for the year 2022 differs significantly from the average productivity for the year 2020 ( $p \leq 0.05$ ). After carrying out the analysis of variance of the set of data obtained in the field, the analysis of descriptive statistics was performed (Table 3).

**Table 3.** Data descriptive statistics

Year	Mean (L plant <sup>-1</sup> )	SD (L plant <sup>-1</sup> )	CV (%)
2020	8.47	4.64	54.78
2021	11.30	7.29	64.51
2022	15.00	7.65	51.00

SD: standard deviation, CV: coefficient of variation.

After the variance analysis of the coffee productivity data, sampled in the years 2020, 2021 and 2022, the following hypotheses are concluded:

- As much as the average productivity for the year 2022 is higher (15 L/plant), no it was necessarily the year that produced the most, and this was justified by the Bonferroni test, where the average productivity of the year 2022 differed significantly from the year 2020, the year in which most of the plants produced more compared to the years 2021 and 2022;
- The difference in productivity between the productive cycles of the coffee tree can be justified by the bienniality, a phenomenon found in coffee plantations, where one year there is a lot of production and the next one less production.
- The high productivity heterogeneity between plants during the same year can be explained by two factors: the first is the age of the coffee tree (25 years) and the second is the management adopted over all these years (conventional fertilization based on average fertility sampled in the field) which can generate a lack or excess of nutrients in the plants, which directly affects the individual production of each plant in this coffee tree.

Silva et al. (2007) evaluating the 2004 harvest of a 4.2ha Mundo Novo coffee tree with 4m spacing between rows and 1m between plants, found in their study an average productivity of 4.81(L/plant). Ferraz et al. (2012a) evaluating soil chemical attributes and productivity for the 2007-2008 season in a

coffee plantation of 22ha under the cultivation of coffee trees of the Topázio cultivar planted at spacing of 3.8m between rows and 0.7m between plants, found a productivity average of 1.45 (L/plant). With the aim of evaluating the spatial and temporal variability of productivity Ferraz et al. (2012b) the authors sampled this attribute in three harvests (2008, 2009 and 2010) in diferente sampling grids within the same coffee tree of 22ha. As a result, the authors obtained average productivity values of 1.45 (L/plant) for the 2008 season, 2.72 (L/plant) for the 2009 season and 4.93 (L/plant) for the 2009 season. 2010.

All the authors mentioned above found average productivity values much lower than those found in this study, and this is justified by several factors such as: coffee variety, spacing, climatic conditions, sampling mesh used in each study, management adopted in the field, etc.

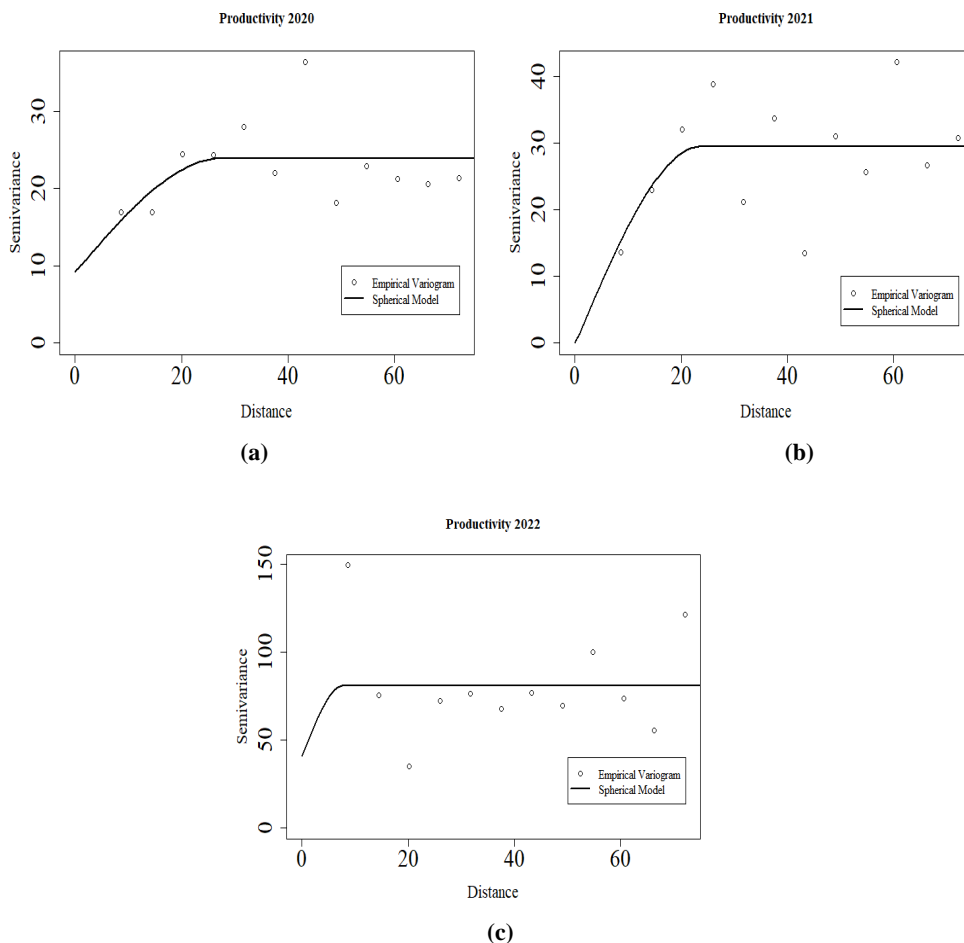
### Geostatistical Analysis

Table 4 represents the adjustment parameters of the semivariograms by the spherical model and by the ordinary least squares method. Figures 6 and 7 represent the semivariograms adjusted for the productivity variable and the thematic maps generated by kriging interpolation for the three years of collection, respectively.

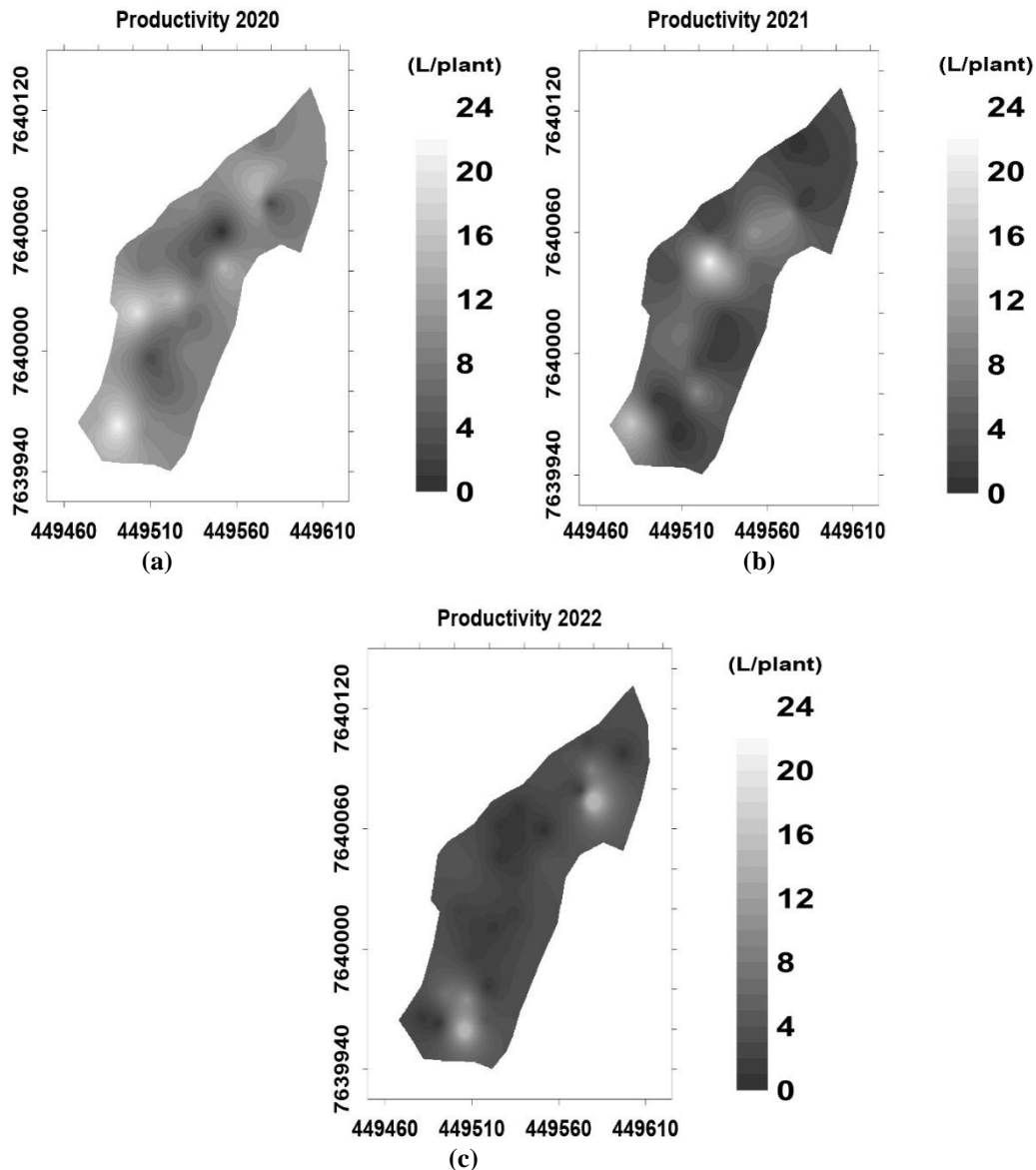
**Table 4.** Parameters for fitting the spherical and exponential semivariogram models of the variables evaluated by the OLS method

	Year	$C_0$	$C_1$	$C_0+C_1$	A	SDD	ME
Productivity	2020	0.10	24.00	24.01	25.00	0.41	strong
	2021	0.50	20.00	20.50	30.00	2.43	strong
	2022	0.01	80.00	80.01	12.00	0.01	strong

$C_0$  - Nugget effect;  $C_1$  - Contribution;  $C_0+C_1$  - Sill; A - Range (m); SDD - Spatial dependence degree; ME - Mean error.



**Figure 6.** Semivariograms adjusted for the spherical model (a) productivity 2020, (b) productivity 2021 and (c) productivity 2022.



**Figure 7.** Kriging maps (a) productivity 2020, (b) productivity 2021 and (c) productivity 2022

Through the geostatistical analysis of the productivity data obtained for the 2020, 2021 and 2022 harvests, it was possible to identify and quantify the spatial dependence of this variable. Camargo et al. (2004) states when an increase is observed between the absolute value of the difference between two samples and the increment of the separation distance between them until reaching a value in which there is no more spatial influence, thus causing the stabilization of the semivariogram the spatial dependence of the attribute is confirmed.

Regarding the adjustment parameters of the semivariograms, it is observed that the nugget effect was different from zero for all years of collection of the productivity attribute, it is an important parameter of the semivariograms as it indicates unexplained variability or even measurement errors, considering the sampling distance used (Carrasco, 2010). The nugget effect can be expressed as a percentage of the plateau, making it easier to compare the degree of spatial dependence.

All seasons (2020, 2021 and 2022) showed a strong degree of spatial dependence according to the criteria of Cambardella et al. (1994). Carvalho et al. (2017) state that the amplitude values relative to the semivariograms are of considerable importance in determining the limit of spatial dependence and can also be an indicator of the interval between the mapping units, important for optimizing future samplings. The highest range was for the 2021 productivity (25m) and the lowest was for the 2022 productivity samples (12m). The spherical model used in this study was efficient, as the mean error value calculated for all seasons was very close to zero, thus meeting the cross-validation criteria.

Geostatistics has been widely used in productivity mapping, as observed in the works Silva et al. (2007, 2008, 2010), Molin et al. (2010), by Ferraz et al. (2012a, b), Carvalho et al. (2013), Carvalho et al. (2017), Ferraz et al. (2017) and Ferraz et al. (2019), all these authors confirmed the spatial dependence of productivity, but no research evaluates the spatial dependence in such a small crop. In relation to the semivariogram adjustment parameters, all these authors found different nugget and range effect values. These differences can be justified, among other factors, by: crop management, crop age, chosen cultivar, soil management and type, climate.

As represented by figures 7a, 7b and 7c, it can be seen that there was a wide variation in productivity for all the years evaluated. The highest values of this attribute are represented by the lighter color (light gray/white), while the lower concentrations of productivity are represented by dark colors (dark gray/black). Visually, the largest contractions in productivity were represented by the year 2020 and the lowest recorded in the year 2022.

Ferraz et al. (2012a) state that productivity maps can be used in harvest management, be it manual, semi-mechanized or mechanized, that is, these maps are important for harvest logistics. In manual harvesting, in addition to estimating productivity, maps may be necessary to establish the number of workers to be hired. As for mechanized harvesting, in addition to the benefits mentioned above, the productivity maps helped in decision-making regarding the rent and/or purchase of machinery and equipment. And finally, in mechanized harvesting, they will be important for the construction of routes, especially when it comes to emptying the tractor support cart, avoiding unnecessary stops and maneuvers.

Figure 7a represents the spatial variability of the productivity evaluated in the year 2020, it is observed that a good part of the productivity was in the range of 7 to 13 L/plant and two points of high productivity were observed in the southwest of the area and two points of low productivity in the center of the area.

Figure 7b represents the map of spatial dependence of the productivity obtained in the year 2021, it was observed that in this year a good part of the productivity was concentrated in the low productivity interval (1 to 7 L/plant) and only a small point of high productivity was found in the center of the area.

The spatial variability of productivity for the year 2022 (Figure 7c) was also concentrated in the range of low productivity (1 to 7 L/plant) and unlike what happened in 2021, where there was a high concentration of productivity (center of the area for the 2021 harvest) in the year 2022, this point had low productivity.

The coffee tree, as a perennial plant with a biennial production cycle, also has different needs from one year to the next. In years of high load, the nutrient demand for fruit production, added to the demand for continuous plant growth, results in a greater need for fertilization (Mesquita et al., 2016). The biennial is a phenomenon considered to be a constant in the production of coffee trees. According to Mendonça et al. (2011) there is a high correlation between management and the biennial, that is, the inefficiency of the cultural management and the climatic adversities accentuate the biennial in the coffee tree, however the nature and magnitude of this influence still lack scientific clarification.

A visual comparison demonstrates the occurrence of biennial productivity, as the regions that in 2020 had the highest productivity had the lowest productivity for the years 2021 and 2022 in the following and consecutively. Plants that produced a lot in 2020 (regions with lighter coloring Figure 7a) used their reserves for fruiting, negatively influencing branch growth and, consequently, reducing productivity in 2021 and 2022 (regions with dark coloring). This can be confirmed by the difference between the evaluated harvests (Figure 7a, 7b and 7c), as these maps show that the areas with the greatest difference, positive or negative, coincide with areas with greater or lesser productivity respectively. Similar results are found in the work by Carvalho et al. (2017) where the authors evaluated two harvests (2012 and 2013) in a 22 ha field under the cultivation of *Coffea arabica* L, cultivar Topázio.

We can call this phenomenon internal variability, by definition the biennial is described as one year producing more and one year producing less, but taking into account that this crop is over 20 years old and the management system adopted is conventional, the needs The specific characteristics of each plant within that crop have not been met over the years, and thus, even with the biennial within coffee crops, for this specific crop the behavior was individual for each evaluated plant, that is, the highest productivities are observed in the 2020 crop, then the 2021 crop is smaller than the 2020 crop and finally the 2022 crop is smaller than the previous two.

The results obtained by a map of spatial variability of productivity and together with maps of

chemical and physical attributes of the soil can be useful to find the reasons for the occurrence of variability in productivity, mainly in the case of low productivity, which will enable the correction failures, allowing these problems to be minimized in the next harvest.

## CONCLUSION

By using one-way analysis of variance with repeated measures, it was possible to verify and quantify the difference between productivity means. The descriptive statistics analysis proved the existence of high variability among the data, by calculating the coefficient of variation.

Through the geostatistical analysis of the productivity data collected in different seasons, it was possible to verify that this attribute has spatial dependence. By adjusting the semivariogram and kriging interpolation, it was possible to prove the magnitude of this spatial dependence. The final product generated by this study were thematic maps, where through them it was possible to identify areas that have the highest and lowest concentrations of productivity.

It is important to emphasize that due to the biennial phenomenon that occurs in coffee growing, productivity is an attribute that will always present a high rate of spatial variability within the same crop, this effect can be minimized with practices and techniques of precision agriculture, that is, the use of specific and localized management can be a great ally to minimize the impacts caused by this variation in productivity. Geostatistics has shown good results in estimating results in unsampled locations, which directly benefits the producer, avoiding intensive and expensive sampling and bringing a quick and reliable return to the producer.

## REFERENCES

- Alves, M. C., Da Silva, F. M., Moraes, J. C., Pozza, E. A., de Oliveira, M. S., Souza, J. C., & Alves, L. S. (2011). Geostatistical analysis of the spatial variation of the berry borer and leaf miner in a coffee agroecosystem. *Precision Agriculture*, *12*, 18-31. Doi: [10.1007/s11119-009-9151-z](https://doi.org/10.1007/s11119-009-9151-z)
- Aparecido, L.E.O., de Souza Rolim, G., & de Souza, P. S. (2015). Sensitivity of newly transplanted coffee plants to climatic conditions at altitudes of Minas Gerais, Brazil. *Australian Journal of Crop Science*, *9*(2), 160-167.
- Bachmaier, M., & Backes, M. (2008). Variogram or semivariogram? Understanding the variances in a variogram. *Precision Agriculture*, *9*, 173-175.
- Burak, D.L., Santos, D.A., & Passos, R.R. Spatial variability of physical attributes: relief, organic matter and productivity in conilon coffee. (2016). *Coffee Science*, *11*(4), 455-466. (in portuguese).
- Camargo, Â.P.D., & Camargo, M.B.P.D. (2001). Definition and layout of the phenological stages of Arabica coffee under tropical conditions in Brazil. *Bragantia*, *60*, 65-68. (in portuguese). DOI: [10.1590/S0006-87052001000100008](https://doi.org/10.1590/S0006-87052001000100008)
- Camargo, E.C.G., Fucks, S.D., & Câmara, G. (2004). Spatial analysis of surfaces. Spatial analysis of geographic data. *Planaltina: Embrapa Cerrados*, 79-122. (in portuguese).
- Camargo, M.B.P.D. (2010). The impact of climatic variability and climate change on arabic coffee crop in Brazil. *Bragantia*, *69*, 239-247. Doi: [10.1590/S0006-87052010000100030](https://doi.org/10.1590/S0006-87052010000100030)
- Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F., & Konopka, A.E. (1994). Field-scale variability of soil properties in central Iowa soils. *Soil science society of America journal*, *58*(5), 1501-1511.
- Carrasco, P.C., Nugget effect, artificial or natural? *The Journal of the Southern African Institute of Mining and Metallurgy*, *110*, 299-306, 2010.
- Cardozo, D.P., Las Schaab, L., & Parré, J.L. (2019). Spatial analysis of coffee productivity in the Southeast region of Brazil: 1990-2015. *Revista Economia Ensaio* *34*(1), 176-198. (in portuguese). Doi: [10.14393/REE-v34n1a2019-40853](https://doi.org/10.14393/REE-v34n1a2019-40853)
- Carvalho, A.M.D., Mendes, A.N.G., Botelho, C.E., Oliveira, A.C.B.D., Rezende, J.C.D., & Rezende, R.M. (2012). Agronomic performance of rust resistant coffee cultivars in the State of Minas Gerais, Brazil. *Bragantia*, *71*, 481-487. DOI: [10.1590/S0006-87052013005000007](https://doi.org/10.1590/S0006-87052013005000007)
- Carvalho, L.C.C., Da Silva, F.M., Ferraz, G.A., da Silva, F.C., & Stracieri, J. (2013). Spatial variability of soil physical attributes and agronomic characteristics of coffee crop. *Coffee Science*, *8*(3), 265-275.
- Carvalho, L.C., Silva, F.M.D., Ferraz, G.A.S., Stracieri, J., Ferraz, P.F., & Ambrosano, L. (2017). Geostatistical analysis of Arabica coffee productivity in two crop seasons. *Revista Brasileira de Engenharia Agrícola e*

- Ambiental*, **21**, 410-414. DOI: [10.1590/1807-1929/agriambi.v21n6p410-414](https://doi.org/10.1590/1807-1929/agriambi.v21n6p410-414)
- Climate Date 2023. [Dados climáticos para cidades mundiais - Climate-Data.org](https://www.climate-data.org). Accessed 20.11.2022.
- Dong, X., Vuran, M.C., & Irmak, S. (2013). Autonomous precision agriculture through integration of wireless underground sensor networks with center pivot irrigation systems. *Ad Hoc Networks*, **11**(7), 1975-1987. Doi:[10.1016/j.adhoc.2012.06.012](https://doi.org/10.1016/j.adhoc.2012.06.012)
- EPAMIG (Minas Gerais Agricultural Research Company). EPAMIG – Empresa de Pesquisa Agropecuária de Minas Gerais. Accessed 20.11.2022.
- Ferraz, G.A.S., Silva, F.M.D., Carvalho, L.C., Alves, M.D.C., & Franco, B.C. (2012a). Spatial and temporal variability of phosphorus, potassium and productivity in a coffee crop. *Engenharia Agrícola*, **32**, 140-150. (in portuguese). DOI: [10.1590/S0100-69162012000100015](https://doi.org/10.1590/S0100-69162012000100015)
- Ferraz, G.A.S., Da Silva, F.M., de Carvalho Alves, M., de Lima Bueno, R., & da Costa, P.A.N. (2012b). Geostatistical analysis of fruit productivity and detachment force in coffee. *Precision Agriculture*, **13**, 76-89. DOI: [10.1007/s11119-011-9223-8](https://doi.org/10.1007/s11119-011-9223-8)
- Ferraz, G.A. S., Silva, F.M.D., Oliveira, M.S.D., Custódio, A.A.P., & Ferraz, P.F.P. (2017). Spatial variability of plant attributes of a coffee crop. *Revista Ciência Agronômica*, **48**, 81-91. Doi: [10.5935/1806-6690.20170009](https://doi.org/10.5935/1806-6690.20170009)
- Ferraz, G.A.S., Ferraz, P.F.P., Martins, F.B., Silva, F.M., Damasceno, F.A., & Barbari, M. (2019). Principal components in the study of soil and plant properties in precision coffee farming. *Agronomy Research* **17**(2), 418-429. Doi: [10.15159/ar.19.114](https://doi.org/10.15159/ar.19.114)
- Fialho, C.M.T., Silva, G.R., Freitas, M.A.M., França, A.C., Melo, C.A.D., & Silva, A.A. (2010). Weed competition with the coffee crop in two infestation times. *Planta Daninha*, **28**, 969-978. (in portuguese). DOI: [10.1590/S0100-83582010000500005](https://doi.org/10.1590/S0100-83582010000500005)
- Gomes, F.P., & Garcia, C.H. (2002). Statistics applied to agronomic and forestry experiments: presentation with examples and guidelines for using applications. Piracicaba: FEALQ, v. 11, p. 21.
- Grego, C.R., & Vieira, S.R. (2005). Spatial variability of soil physical properties in an experimental plot *Revista Brasileira de Ciência do Solo*, **29**, 169-177. (in portuguese).
- Isaaks, E.H.; Srivastava, R.M. (1989). An introduction to applied geostatistics. New York: Oxford Universit, 561p.
- Lopes, P.R., Araújo, K.C.S., Ferraz, J.M.G., Lopes, I.M., & Fernandes, L.G. (2012). Agroecological coffee production in southern Minas Gerais: alternative systems to intensive agrochemical production. *Revista brasileira de agroecologia*, **7**(1), 25-38. (in portuguese).
- Martello, M., Molin, J.P., Bazame, H.C., Tavares, T.R., & Maldaner, L.F. (2022). Use of Active Sensors in Coffee Cultivation for Monitoring Crop Yield. *Agronomy*, **12**(9), 2118. DOI: [10.3390/agronomy12092118](https://doi.org/10.3390/agronomy12092118)
- Mendonça, R., Rodrigues, W., Martins, L., & Tomaz, M.A. (2011). Approach to biennial production in coffee plants. *Enciclopédia Biosfera*, **7**(13). (in portuguese).
- Mesquita, C.M., Rezende, J.E., Carvalho, J.S., Fabri Júnior, M.A., Moraes, N.C., Dias, P.T., Carvalho, R.M., & Araújo, W.G. (2016). Coffee manual: management of coffee plantations in production. Belo Horizonte: EMATER-MG, 2016. 72 p. (in portuguese).
- Molin, J.P., Motomiya, A.V.D.A., Frasson, F.R., Faulin, G.D.C., & Tosta, W. (2010). Test procedure for variable rate fertilizer on coffee. *Acta Scientiarum. Agronomy*, **32**, 569-575. DOI: [10.4025/actasciagron.v32i4.5282](https://doi.org/10.4025/actasciagron.v32i4.5282)
- Piri, I., Moosavi, M., Taheri, A.Z., Alipur, H., Shojaei, S., & Mousavi, S.A. (2019). The spatial assessment of suitable areas for medicinal species of Astragalus (Astragalus hypsogeton Bunge) using the Analytic Hierarchy Process (AHP) and Geographic Information System (GIS). *Remote sensing and space sciences*, **22**(2), 193-201. DOI: [10.1016/j.ejrs.2018.02.003](https://doi.org/10.1016/j.ejrs.2018.02.003).
- Ribeiro Júnior, P.J., & Diggle, P. (2001). GeoR: a package for geostatistical analysis. *R-News*, **1**(2), 15-18.
- Sá Júnior, A., de Carvalho, L.G., Da Silva, F.F., & de Carvalho Alves, M. (2012). Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. *Theoretical and Applied Climatology*, **108**, 1-7. DOI:[10.1007/s00704-011-0507-8](https://doi.org/10.1007/s00704-011-0507-8)
- Santana, L.S., Ferraz, G.A., Santos, S.A.D., & Dias, J.E.L. (2022). Precision coffee growing: a review. *Coffee Science* **17**. DOI:[10.25186/v17i.2007](https://doi.org/10.25186/v17i.2007)
- Santos, S.A., Ferraz, G.A.S., Figueiredo, V.C., Volpato, M.M.L., Machado, M.L., & Silva, V.A. (2023). Evaluation of the water conditions in coffee plantations using RPA. *AgriEngineering*, **5**, 65-84. DOI: [10.3390/agriengineering5010005](https://doi.org/10.3390/agriengineering5010005)
- Scalco, M.S., Alvarenga, L.A., Guimarães, R.J., Dominghetti, A.W., Colombo, A., Assis, G.A., & Abreu, G.F. (2014). Leaf contents of phosphorus and zinc, productivity, and growth of irrigated coffee. *Pesquisa Agropecuária Brasileira*, **49**, 95-101. (in portuguese). DOI: [10.1590/S0100-204X2014000200003](https://doi.org/10.1590/S0100-204X2014000200003)
- Silva, F.M.D., Souza, Z.M.D., Figueiredo, C.A.P.D., Marques Júnior, J., & Machado, R.V. (2007). Spatial variability of chemical attributes and productivity in coffee crops. *Ciência Rural*, **37**, 401-407. (in portuguese).
- Silva, F.M.D., Souza, Z.M.D., Figueiredo, C.A.P.D., Vieira, L.H.D.S., & Oliveira, E.D. (2008). Spatial variability of chemical attributes and productivity of the coffee crop in two agricultural seasons. *Ciência e Agrotecnologia*, **32**, 231-241. (in portuguese). Doi: [10.1590/S1413-70542008000100034](https://doi.org/10.1590/S1413-70542008000100034)

- Silva, F.M.D., Alves, M.D.C., Souza, J.C.S., & Oliveira, M.S.D. (2010). Effects of manual harvesting on coffee biennial in Ijaci, Minas Gerais. *Ciência e Agrotecnologia*, *34*, 625-632. (in portuguese). DOI: [10.1590/S1413-70542010000300014](https://doi.org/10.1590/S1413-70542010000300014)
- Wadt, P.G.S., & Dias, J.R.M. (2012). Regional and interregional DRIS norms in the nutritional assessment of Conilon coffee. *Pesquisa Agropecuária Brasileira*, *47*, 822-830. (in portuguese). DOI: [10.1590/S0100-204X2012000600013](https://doi.org/10.1590/S0100-204X2012000600013)

Artigo 2 – Elaborado conforme as normas do periódico	
Título do artigo	Spatial variability of chlorophyll and NDVI obtained by different sensors in an experimental coffee field
Autores	S.A.S. Silva, G.A.S. Ferraz, V.C. Figueiredo, M.M.L. Volpato, M.L. Machado, V.A. Silva, C.S.M. Matos, L. Conti and G. Bambi
Periódico	Agronomy Research
ISSN	1406-894X
DOI	<a href="https://doi.org/10.15159/AR.24.037">https://doi.org/10.15159/AR.24.037</a>

Fonte: elaboração própria.

### **SPATIAL VARIABILITY OF CHLOROPHYLL AND NDVI OBTAINED BY DIFFERENT SENSORS IN AN EXPERIMENTAL COFFEE FIELD**

S.A.S. Silva<sup>1</sup>, G.A.S. Ferraz<sup>1\*</sup>, V.C. Figueiredo<sup>2</sup>, M.M.L. Volpato<sup>2</sup>, M.L. Machado<sup>2</sup>, V.A. Silva<sup>2</sup>, C.S.M. Matos<sup>2</sup>, L. Conti<sup>3</sup> and G. Bambi<sup>3</sup>

<sup>1</sup>Federal University of Lavras, School of Engineering, Department of Agricultural Engineering, Rotary Clover Professor Edmir Sá Santos, BR37200-900 Lavras, Brazil

<sup>2</sup>Agricultural Research Company of Minas Gerais, Av. José Cândido da Silveira 1647, Bairro União Belo Horizonte, BR31170-495 Belo Horizonte, Brazil

<sup>3</sup>University of Florence – UniFI, Department of Agriculture, Food, Environment and Forestry (DAGRI), Via San Bonaventura, 13, IT50145 Florence, Italy

\*Correspondence: [gabriel.ferraz@ufla.br](mailto:gabriel.ferraz@ufla.br)

Received: January 31st, 2024; Accepted: May 1st, 2024; Published: May 8th, 2024

**Abstract.** The objective of this research was to study the spatial variability of NDVI and chlorophyll sampled by different sensors, as well as to evaluate the correlation between them in a coffee field. The study was carried out on a coffee farm located in the municipality of Três Pontas, Minas Gerais. A sampling grid containing 30 points was created for the study area. Each sampling point was represented by one plant, which was georeferenced by a GNSS RTK. For each sample plant, NDVI and chlorophyll were obtained by the optical and active sensors GreenSeeker and ClorofiLOG, respectively. In addition, it was carried out a flight with an RPA equipped with a passive and multispectral sensor. Using the data obtained by active sensors, a geostatistical analysis was carried out to evaluate the spatial variability of NDVI and chlorophyll. The geostatistical analysis verified the existence of spatial dependence for the two attributes, and thus it was possible to generate spatialization maps through kriging. The images obtained by the passive sensor resulted in five multispectral orthomosaics, making it possible to calculate the NDVI, thus generating a spatialization map of this index. It was possible to observe in the generated maps, points that presented a certain similarity and for this purpose a correlation analysis was carried out for the values of each attribute, sampled directly in the maps, and in different sampling grids (30, 60, 90 and 120 points). By analyzing the Pearson coefficient (R) it was possible to quantify the level of correlation between the data obtained by the different sensors and through the t test it was possible to find significant correlations between them.

**Keywords :** Geostatistics, Spatial distribution, Spectral indices, active sensor, passive sensor

## INTRODUCTION

The term Precision Coffee Farming originated from the application of Precision Agriculture (PA) techniques and technologies to coffee cultivation (Alves et al., 2006). It is characterized as a set of techniques, technologies, and tools that efficiently characterize the spatial variability of coffee tree parameters (Ferraz et al., 2012; Santana et al., 2022). Through this set of characteristics, it is possible to obtain an integrated base of information and to understand the relationships between the production system, the attributes involving soil and plants, and their behavior in variations in space, time, and climate.

In PA, a very important technological tool is remote sensing (RS), widely disseminated due to its various applications. Precision agriculture combined with computational tools has been studied and widely disseminated in terms of coffee crops (Santos et al., 2023). According to Amaral et al. (2020), sensors and applications in remote sensing (RS) at all levels (orbital, aerial and terrestrial) have significantly evolved.

Spectral remote sensing enables early, efficient, objective and non-destructive assessment of plant responses to environmental stress factors (Li et al., 2010). According to Shiratsuchi et al. (2014) the sensors used by RS can be divided into two categories: passive or active. Passive sensors record electromagnetic energy reflected or emitted by the target, such as reflected solar radiation or emitted thermal radiation. Active sensors provide their own source of electromagnetic energy, such as radars, sonars, active canopy sensors (such as GreenSeeker and ClorofiLOG, for example).

The combination of spectral data from two or more bands creates vegetation spectral indices. These improve the relationship between spectral data and the biophysical parameters of vegetation (Zanzarini et al., 2013). Vegetation indices allow determining the health status of crops, based on different characteristics (Main et al., 2011; Yu et al., 2014). The normalized difference vegetation index (NDVI) is one of the technologies widely used in the field of remote sensing and has a strong relationship with morphophysiological variables, such as leaf health, leaf area index (LAI), biomass, plant productivity and chlorophyll concentration (Kim et al., 2022).

In recent years, RS methodologies have been widely used in monitoring agricultural crops and in decision-making for better management practices (Marin et al., 2019). Jesus et al. (2014) states that RS techniques have been widely used to evaluate vegetation indices and chlorophyll levels, with the aim of identifying, in real time, possible changes due to variations that may occur in cultivation. Barbosa et al. (2019) and Santos et al. (2019a) stated that the use of Remotely Piloted Aircraft (RPA) in PA has increasing potential for agricultural monitoring through obtaining data with RS techniques. RPA's can be used in smaller areas or in specific locations to obtain data in less time, being able to monitor crop growth.

By identifying the spatial variability of the vegetative development of plants, it is possible to observe differences in productive potential in coffee plantations, where in most rural properties, they are treated in a uniform manner in terms of management (Rodrigues et al., 2019). Campos et al. (2022) state that adequate management of coffee plantations can be carried out in order to make the plants well-nourished and productive through geospatial and temporal monitoring of coffee trees.

The chlorophyll content of leaves is an indicator that represents the growth status of crops, and its monitoring in crops is crucial for agricultural practices (Pereira et al., 2019). Chlorophyll can be subdivided into two classes: chlorophyll a and chlorophyll b, together with carotenoids, are part of the primary photosynthetic pigments of plants (Moreira, 2012). These photosynthetic pigments have an important role in plant physiology, as they are correlated with attributes such as nitrogen and magnesium (Marengo & Lopes, 2007) that are related to plant nutrition. Furthermore, chlorophyll absorbs energy at different wavelengths (Moreira, 2011), which makes it possible to use remote sensing (RS) techniques to observe the behavior of this attribute in different crops (Santos et al., 2019b).

Therefore, this work aims to study the spatial variability of the normalized difference vegetation index and chlorophyll obtained by active and passive sensors, as well as to evaluate the correlation between the data obtained by them in different sampling grids in coffee fields.

## MATERIALS AND METHODS

The study was carried out in a coffee plantation located in an Experimental Field belonging to the Agricultural Research Company of Minas Gerais (EPAMIG, acronym in portuguese), located in the municipality of Três Pontas, southern region of the state of Minas Gerais, Brazil, at mean altitude of 905 m above sea level in the coordinates of the Universal Transverse Mercator (UTM) system S 7.640.030,4 and E 449.531,5, Zone 23K. This municipality has at mean annual temperature of 20.3°C and a mean annual precipitation of 1429 mm. The soil in this area is classified as Oxisol.

The experimental area comprised 1.2 ha of coffee trees of the *Coffea arabica* L. species, cultivar Topázio MG 1190. This crop was established in 1998 with spacing between rows of 3.70m and 0.70 m between plants (Figure 1).

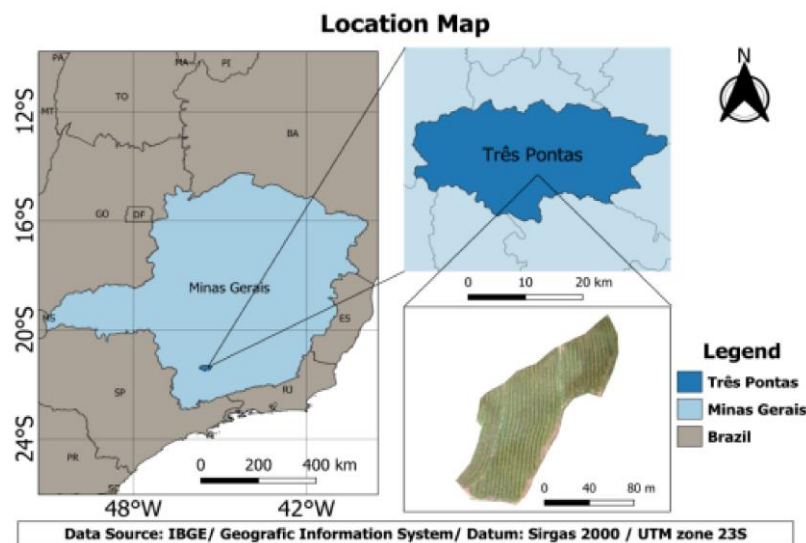


Figure 1. Location map

In this area, a sampling grid was developed containing 30 georeferenced points (Figure 2), where sample data for the variables  $NDVI_{ps}$  (NDVI obtained from RPA images),  $NDVI_{as}$ ,  $NDVI_{as.east}$  and  $NDVI_{as.west}$  (NDVI obtained by GreenSeeker readings) and Chlorophyll. Each sampling point was represented by a plant. Both the study area and sampling points were georeferenced by a global navigation satellite system (GNSS) real-time kinematic (RTK).

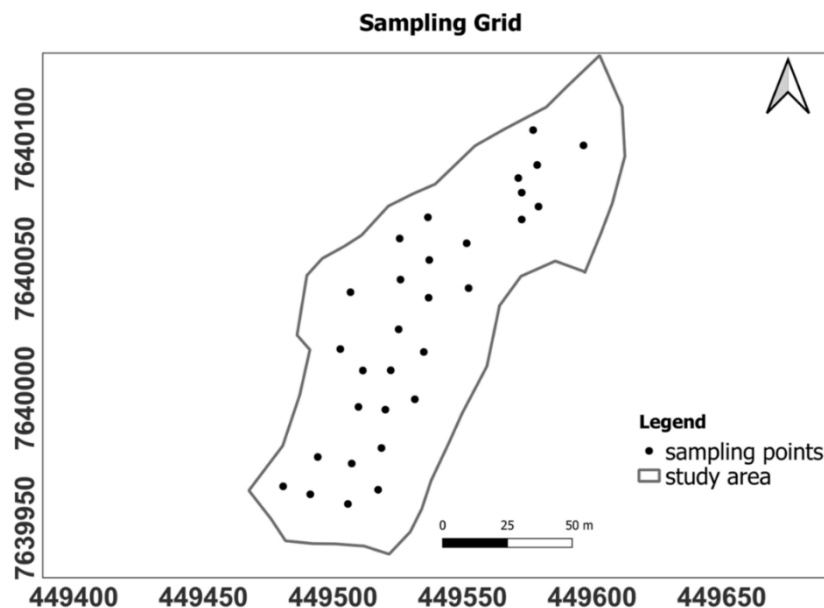


Figure 2. Sample grid

The data collection for this study was conducted in August 2022. All measurements with the different sensors were carried out simultaneously in the interval from 11:00 am to 12:00 pm. The methodology involved several steps, including field data collection, data processing, generation of maps, and comparison of values.

### Field data collection

#### Chlorophyll

Chlorophyll data were acquired employing an active sensor, designated as chlorofiLOG, produced by Falker. The measurement by this sensor is conducted optically, utilizing the optimal light frequencies absorbed by chlorophyll during photosynthesis. This instrument assesses three frequency bands and can discriminate between the two chlorophyll types: Chlorophyll a and Chlorophyll b. Each sampled plant was stratified into three sections (upper, middle, and lower), and the chlorophyll measurements was specifically taken from the middle section. Ten leaves were selected from either the third or fourth node from the apex of the plagiotropic branch, ensuring they were healthy and devoid of signs of pests or diseases (Santos et al., 2022; Bento et al., 2022; Barata et al., 2023). Sampling occurred around 11:00 a.m., coinciding with the timing of the Remote Piloted Aircraft (RPA) flight for image collection. In total, 300 chlorophyll readings were recorded, resulting in a dataset comprising 30 attribute values, derived from the mean of 10 readings per sampled plant.

#### NDVI active sensor (NDVI<sub>as</sub>)

For the Normalized Difference Index (NDVI) readings obtained by an active sensor (NDVI<sub>as</sub>), the commercial GreenSeeker® 505 HandHeld Sensor was used, which is a non-image-forming optical sensor capable of measuring this index. The measurements of this index using this device were carried out using a method similar to scanning.

For NDVI<sub>as</sub> data sampling, the plant was divided into three sections, referred to as thirds (upper, middle, and lower), and two sides of solar exposure (east and west) were considered. On each side of sun exposure, 9 readings were taken, with 3 in each third represented by Figure 3. So, 540 NDVI<sub>as</sub> values were obtained, divided into 18 readings for each sample plant (considering the 9 readings obtained for the east side and the 9 readings obtained for the west side). The readings were divided as follows: NDVI<sub>as</sub> (mean of 18 readings obtained in the three thirds of the plants for the east and west sides), NDVI<sub>as.east</sub> (mean of 3 readings obtained from the upper third of the sample plants facing east) and NDVI<sub>as.west</sub> (mean of 3 readings obtained in the upper third of the sample plants facing west).

This methodology for NDVI readings obtained by the GreenSeeker sensor was adopted considering that due to the high height of the coffee trees (above 2m) it would be impossible to read the crown of the plants.



Figure 3. NDVI data sampling using the active sensor

### NDVI passive sensor (NDVI<sub>ps</sub>)

To generate NDVI values by the passive sensor (NDVI<sub>ps</sub>), a flight was carried out with an RPA, model EBee SQ from the manufacturer senseFly. This aircraft has the following characteristics: fixed wing, a wingspan of 110 cm, rated radio range of 3 km, cruising speed of 40-110 km/h, wind resistance of up to 45 km/h (12 m/s), electric motor, maximum payload of 1.1 kg (including camera and batteries) and flight autonomy of up to 55 min.

The aircraft was equipped with a Parrot camera (Sequoia model) with a high-resolution red-green-blue (RGB) sensor with a focal length of 4.88 mm. This camera also has four monochromatic sensors for the spectral bands: green (550 ± 40 nm), red (660 ± 40 nm), near infrared (NIR) (790 nm ± 40 nm) and red edge (735 ± 40 nm). The resolution is 1280 × 960, with a pixel size of 3.75 μm and a focal length of 3.98 mm; the ground sample distance (GSD) is 6.8 cm at a flight height of 50 m (above ground level – AGL), which was adopted for the described study.

In addition to the RGB and monochromatic sensors, Sequoia has a luminosity sensor to correct the influence of the sun by obtaining data with radiometric corrections. This sensor records not only the current lighting, but also the location of the center of the photo and inertial data.

Flight planning and execution was carried out through the base station, which was developed by the same aircraft manufacturer (senseFly) with the following set: the eMotion software, responsible for flight programming and execution of the aircraft's path, and a transmitting antenna that allows real-time monitoring of the overflight, as well as sending commands for landing, direction changes and image acquisition. During the flight, images are stored on the memory card contained in the multispectral sensor. After the flight, these images will be downloaded and sent to processing software. The flight plan followed the parameters represented in Table 1.

Table 1. flight planning parameters

<b>Camera</b>	Parrot Sequoia
<b>Resolution of the RGB camera</b>	16 megapixels
<b>Resolution of the multispectral camera</b>	1.2 megapixels
<b>Focal length</b>	3.98mm
<b>Vertical cover</b>	70%
<b>Horizontal cover</b>	70%
<b>Spatial resolution</b>	6.8 cm
<b>Flight altitude</b>	50 m
<b>Speed</b>	12 m/s

### Processing of data obtained by active Chlorophyll and NDVI<sub>as</sub> sensors

Chlorophyll measurements using the ClorofiLOG sensor were downloaded using specific software provided by the equipment manufacturer. In this way, it was possible to download the measurements stored in the device and determine chlorophyll a (Chla) and chlorophyll b ( Chlb). The sum of these two types of chlorophyll (Chla + Chlb) results in Total Chlorophyll (TC), which was used in this work.

Mean values of NDVI<sub>as</sub>, NDVI<sub>as.east</sub>, NDVI<sub>as.west</sub>, and chlorophyll were utilized to analyse the spatial dependence of these attributes using semivariograms. Semivariance is classically estimated by Equation (1), according to Vieira (2000).

(1)

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 N(\mathbf{h})} \sum_{i=1}^{N_i(\mathbf{h})} [Z(x_i) - Z(x_i + \mathbf{h})]^2$$

Where N(h) is the number of experimental pairs of observations Z(xi) and Z(xi + h), separated by a distance h. The semivariogram is represented by the graph  $\hat{\gamma}(\mathbf{h})$  versus h. From the adjustment of a mathematical model to the calculated values of  $\hat{\gamma}(\mathbf{h})$ , the coefficients of the theoretical model are estimated for the semivariogram, called nugget effect (C<sub>0</sub>), contribution (C<sub>0</sub> + C<sub>1</sub>) and range (a), as described by Bachmaier & Backers (2008).

For this study, the ordinary least squares (OLS) method and spherical model were used to fit the semivariogram. To check whether the model adjustments met the cross-validation requirements, the mean error (ME) was calculated according to *Isaaks & Srivastava (1989)*. The ME must have a value as close to zero as possible.

After adjusting the semivariograms and verifying spatial variability, the data were interpolated using ordinary kriging. Thus, the variables were estimated in locations where they were not sampled, which made it possible to visualize their distribution in space through thematic maps.

The calculation of the degree of the spatial dependence (DSD) of the variables followed the classification proposed by Cambardella et al. (1994). In this classification, the authors point out that there is strong spatial dependence when the semivariogram presents a nugget effect equal to or less than 25% of the sill variance, moderate spatial dependence when this relationship is between 25% and 75%, and weak spatial dependence when it is greater than 75%.

The geostatistical analysis carried out by adjusting the semivariograms and ordinary kriging was carried out in the RStudio software using the geoR package (Ribeiro Jr & Diggle, 2001). Kriging generated interpolated and georeferenced points, which were exported to the QGIS version 3.22.9 software to create isocolour maps for the attributes  $NDVI_{as}$ ,  $NDVI_{as.east}$ ,  $NDVI_{as.west}$  and chlorophyll.

### Processing of NDVI data obtained by the passive sensor

The normalized difference index (NDVI) consists of calculating the difference between emission and reflection of two wavelengths of the electromagnetic spectrum: near infrared (0.725-1.1  $\mu m$ ) and red (0.58-0.68  $\mu m$ ) (Rouse et al., 1973).

After completing the flight and capturing images using the multispectral camera, the images were processed using the Pix4D software, resulting in four orthomosaics for the spectral bands and one orthomosaic for the RGB composition. The orthomosaics were imported to the geoprocessing software QGIS, where NDVI was calculated using the raster calculator tool and using Equation 2, enabling to obtain the NDVI values for the study area.

(2)

$$NDVI = \frac{NIR-RED}{NIR+RED}$$

Where:

NIR: near infrared spectral band

RED: red spectral band

The NDVI value varies from -1 to 1 and shows us the vigor of the crop. Values close to 1 mean that the more intense the green, the more vigor there is in the vegetation and vegetation cover. We must take into account whether we are working with extensive or intensive cultivation, or if there is no bare soil, as all of this will be taken into account by the index. And this will also measure the vigor of the weeds. Values close to 0 correspond to areas with little vegetation, initial stages of cultivation, bare soil or non-productive areas. Negative values are generally associated with areas of water, snow, or clouds.

### Map generation and correlation analysis

For this study, five maps were generated, one map through the calculation of NDVI using multispectral images ( $NDVI_{ps}$ ) and four maps obtained by geostatistical analysis through ordinary kriging ( $NDVI_{as}$ ,  $NDVI_{as.east}$ ,  $NDVI_{as.west}$ , and chlorophyll). All maps were created using QGIS software. To compare map values of the studied variables, 4 sample grids of 30, 60, 90 and 120 points were created within the area. To construct the grids, the following requirements were used:

- Points have a minimum distance of 2m from each other;
- The points are 4m away from the area's borders, so that there would be no interference in the results.

Polygons of 0.8 m were created using the zonal statistical tool for each sampling point within the grids. These polygons were then sampled using the mean pixel values within the 0.8m buffer for each sample point across the different grids.

The sampling of NDVI data, for both sensors, and chlorophyll values resulted in 4 tables, each of

which is represented by a sampling grid. With the data from the tables, it was possible to perform correlation analysis using the RStudio software through the ggplot2 library. Correlation analysis summarizes the degree of relationship of two or more variables, its calculation results in the Pearson Correlation Coefficient (R). For its validation, the test followed the criteria:

$\rho \leq 0.05$ : significant at the 5% probability level

$\rho > 0.05$ : not significant

## RESULTS AND DISCUSSION

### Descriptive statistics

Table 2 presents the data from the descriptive statistical analysis of the variables under study that were obtained by the sensors directly and indirectly in the field. In the table it is possible to observe the minimum, maximum, median, mean, variance, standard deviation and coefficient of variation values for the attributes: NDVI<sub>as</sub>, NDVI<sub>as.east</sub>, NDVI<sub>as.west</sub> and chlorophyll.

Table 2. Descriptive statistics of data

Attribute	Minimum	Maximum	Median	Mean	Variance	SD	CV (%)
NDVI <sub>ps</sub>	0.27	0.77	0.51	0.51	0.01	0.13	25.87
NDVI <sub>as</sub>	0.80	0.89	0.86	0.86	3.73	0.01	2.28
NDVI <sub>as.east</sub>	0.40	0.79	0.65	0.64	0.01	0.09	15.56
NDVI <sub>as.west</sub>	0.43	0.75	0.64	0.62	0.01	0.10	16.14
Chlorophyll	40.10	64.20	52.44	52.44	29.71	5.45	10.39

SD: standard deviation; CV: coefficient of variation

From descriptive statistics it is possible to observe that the mean for NDVI values was higher when considering all thirds and the east and west faces for GreenSeeker readings (mean NDVI<sub>as</sub> = 0.86), while the means for NDVI<sub>ps</sub>, NDVI<sub>as.east</sub> and NDVI<sub>as.west</sub> were very close (0.51, 0.64, 0.62) respectively. Regarding chlorophyll, the mean for this attribute was 52.44.

Gomes & Garcia (2002) state that the variability of an attribute can be classified according to the magnitude of its coefficient of variation (CV), which according to the authors can be: low, when this is less than 10%; moderate when it is in the range of 10 to 20%, high when it is between 20 and 30%; and very high when it is above 30%. Frogbrook et al. (2002) state that the first indicators of data heterogeneity are high CV values. According to the CV values represented in Table 2, we can state that the attribute NDVI<sub>ps</sub> represent high variability, while the attributes NDVI<sub>as.east</sub>, NDVI<sub>as.west</sub> and chlorophyll represent moderate variability in their data sets, the NDVI<sub>as</sub> presented low variability (2,28%).

### Geostatistical analysis

Table 3 shows the semivariogram adjustment parameters, as well as the mean error values and degree of spatial dependence.

Table 3. Semivariogram adjustment parameters for the variables under study

variable	C <sub>0</sub>	C <sub>1</sub>	C <sub>0</sub> +C <sub>1</sub>	a	DSD (%)	ME	
NDVI <sub>as</sub>	0.01	3.50	3.51	40	0.28	Strong	0.0000
NDVI <sub>as.east</sub>	0.00	0.01	0.01	45	0.00	Strong	0.0001
NDVI <sub>as.west</sub>	0.00	0.01	0.01	25	0.00	Strong	0.0009
Chlorophyll	0.01	20.00	20.01	65	0.05	Strong	0.0031

C<sub>0</sub>: nugget effect; C<sub>1</sub>: sill; C<sub>0</sub>+C<sub>1</sub>: contribution; a: range; DSD: degree of the spatial dependence and ME: mean error

Through the results of the geostatistical analysis, it was observed that, for all variables, the ME presented very low values and close to zero, demonstrating that the adjustments made by the spherical model were well made and met the cross-validation requirements. All variables represented ME values very close to zero.

This study identified that all variables studied showed a strong degree of spatial dependence (DSD).

## Thematic maps

Figure 4 represents the maps of NDVI (Passive sensor), NDVI (Active sensor), NDVI (Active sensor - East), NDVI (Active sensor - West), and Chlorophyll respectively.

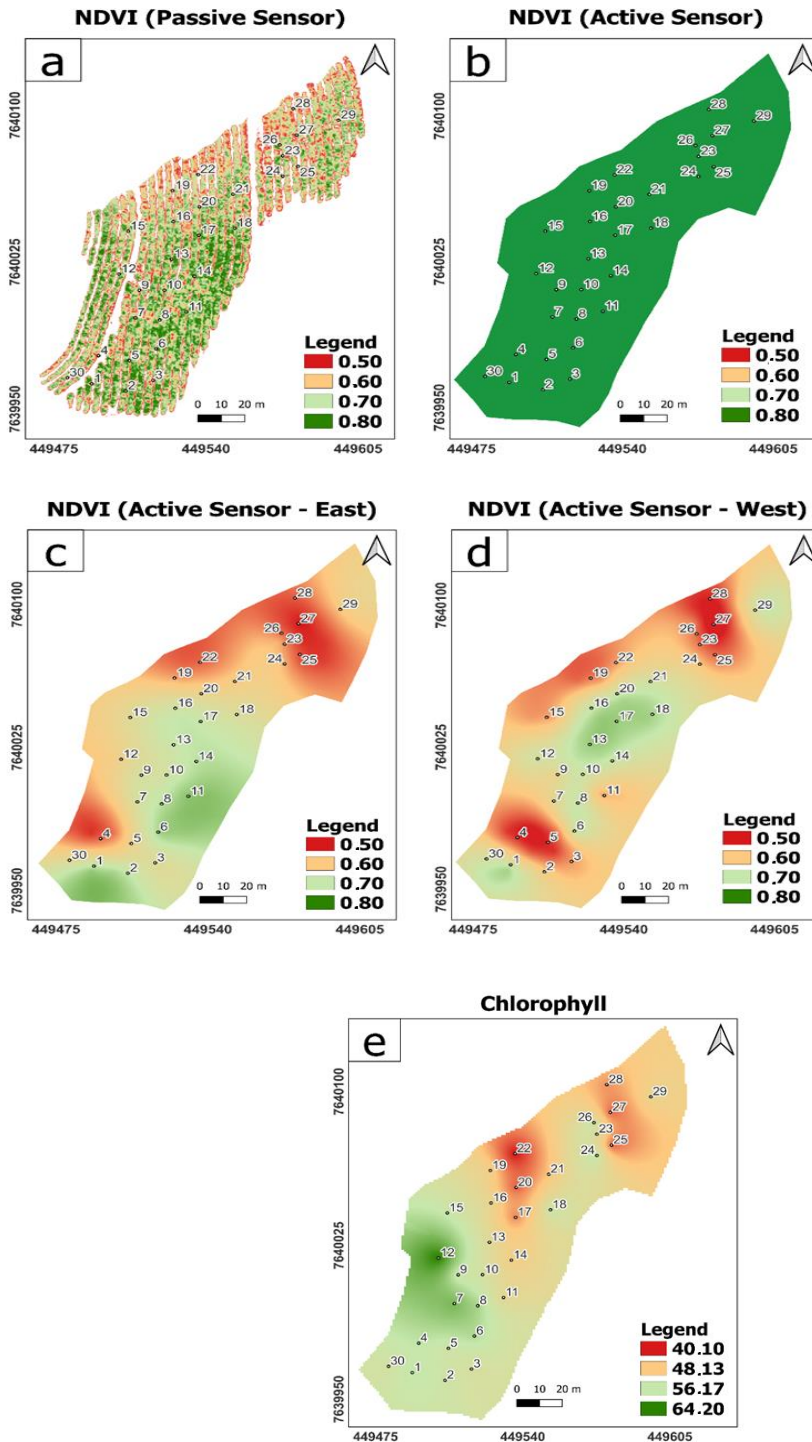


Figure 4. Thematic maps for the variables  $NDVI_{ps}$ ,  $NDVI_{as}$ ,  $NDVI_{as.east}$ ,  $NDVI_{as.west}$  and chlorophyll.

The maps represented by Figure 4 were created in the same color palette and also in the same range of values for the NDVI attribute (Figures 4a, 4b, 4c and 4d).

When analyzing Table 2, it is possible to observe that the NDVI index obtained by the passive

sensor varied from the lowest value of 0.27 to the highest value of 0.77, however, during the construction of the thematic map, it was observed that NDVI values below 0.50 were related to exposed soil, low vegetation and drought between the rows of coffee trees, so to facilitate the visualization of only the NDVI values in the coffee plants, NDVI values below 0.50 were omitted in the image, facilitating the comparison between the values of NDVI obtained actively by the interpolated maps.

The visual analyzes that will be presented below will disregard Figure 4b, due to the little variation in its values (0.80 to 0.89 Table 2) the thematic map was represented by only one tone (dark green) which makes discussion difficult. of this image with the results obtained by the other methods in images 4a, 4c, 4d and 4e.

When looking at Figure 4a, it is possible to state that most of the coffee trees had NDVI values in the range of 0.60 to 0.70. It is also possible to state that when drawing an imaginary line dividing the area into two hemispheres, the north of the area (above point 16) presents lower NDVI values, unlike the south side of the area (below point 16) where it is possible to find some concentrations of higher values (NDVI around 0.80). Throughout the area it is possible to observe some points with low NDVI values (around 0.50) mainly in the coffee trees close to points 19 and 22 and also close to points 23, 26, 27 and 28, which is also observed in the figures 4c and 4d.

In relation to Figures 4c and 4d, it is possible to observe that the NDVI values coincide in a large part of the area. A small difference is observed between the NDVI values only in coffee trees close to sampling points 3,5, 11, 15, 29 and in the zone below points 1,2 and 30. Furthermore, it is possible to observe that to the north of the area in Figures 4c and 4d the NDVI values are lower when compared to the values in the southern part of the area, this same situation is found in figure 4a. This can be justified due to the northern part of the area, mainly around points 19, 22, 23, 25, 26, 27 and 28, which are next to a very busy rural road that accesses a coffee processing plant, therefore, Coffee trees closer to both the road and the plant are affected by dust as well as higher temperatures, generating a greater set of stress on the coffee plants.

When analyzing the map represented by Figure 4e, it is observed that lower chlorophyll values (around 40.10) are concentrated around points 19 and 22 and also close to points 23, 25, 27 and 28, coinciding with the lower NDVI values of figures 4a, 4c and 4d. Regarding the highest chlorophyll values, these are found in coffee trees close to point 12 (southwest of the area), values that coincide with the NDVI values found in Figure 4a.

Visual analysis of the maps is the first step to identify the existence of a possible correlation between the variables and the methods used, however a mathematical analysis is necessary to quantify the relationship between them, as well as verify their significance. Therefore, a correlation analysis was carried out by calculating the Pearson coefficient (R) as well as performing the t test to determine significance.

### **Correlation analysis**

In the literature, there are studies that evaluate the correlation between  $NDVI_{ps}$  x Total Chlorophyll (Bento et al., 2022; Santos et al., 2022 and Barata et al., 2023) as well as researches that evaluate the correlation of NDVI obtained by actives and passive sensors (Gomes et al., 2021 and Campos et al., 2022) in coffee cultivation. This research will be important to provide a basis for discussing the results presented in this paper.

In Figure 5 it is possible to observe the correlation graphs for the variables  $NDVI_{ps}$ ,  $NDVI_{as}$ ,  $NDVI_{as.east}$ ,  $NDVI_{as.west}$ , and chlorophyll. Table 4 shows the  $\rho$  values for each variable and sampling grid used in this study in order to verify their significance using the t test.

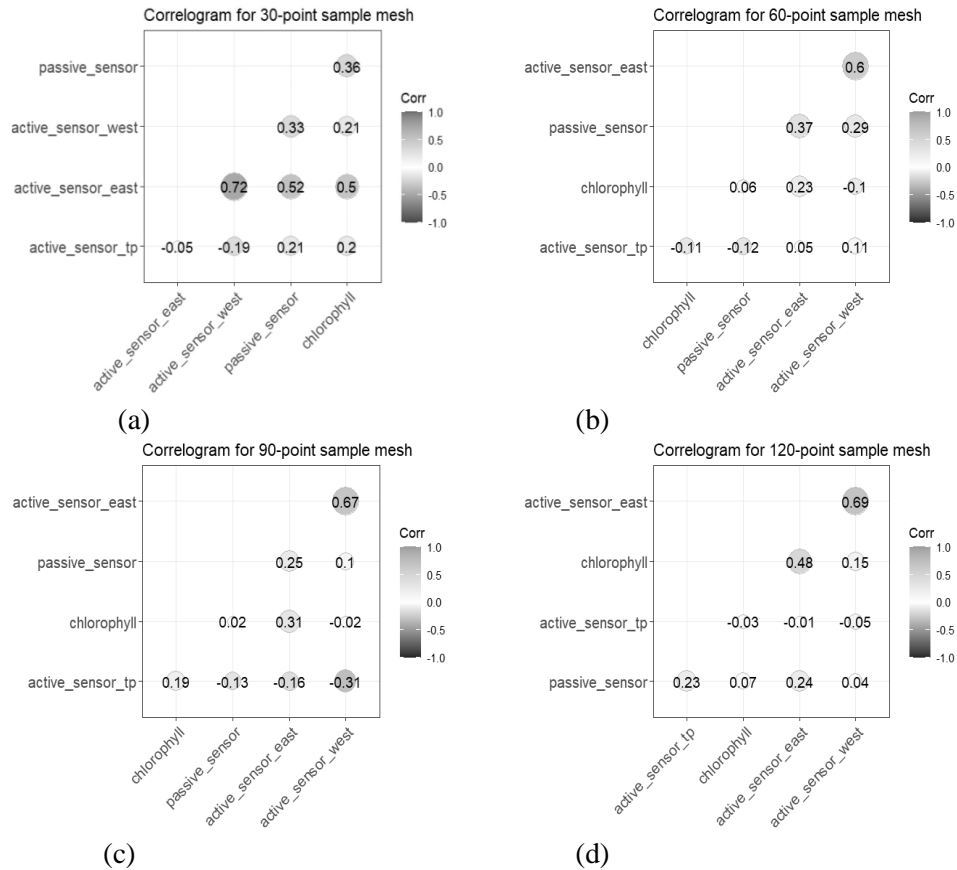


Figure 5. Correlograms for sample grides with 30 points (a), 60 points (b), 90 points (c) and 120 points (d)

Table 4.  $\rho$  value for correlations between variables in different sampling grids

\*\* : significant at the 5% level; ns : not significant

Grid	Variable	NDVI <sub>ps</sub>	NDVI <sub>as</sub>	NDVI <sub>as.east</sub>	NDVI <sub>as.west</sub>	Chl
30	NDVI <sub>ps</sub>	0.00	0.27ns	0.00 **	0.07ns	0.05ns
	NDVI <sub>as</sub>	0.26ns	0.00 ns	0.78ns	0.32ns	0.30ns
	NDVI <sub>as.east</sub>	0.00ns	0.78ns	0.00	0.00 **	0.00 **
	NDVI <sub>as.west</sub>	0.07ns	0.32ns	0.00 **	0.00	0.27ns
	Chl	0.05ns	0.30ns	0.00 **	0.27ns	0.00 ns
60	NDVI <sub>ps</sub>	0.00	0.37ns	0.00 **	0.02 **	0.66ns
	NDVI <sub>as</sub>	0.37ns	0.00	0.07ns	0.38ns	0.42ns
	NDVI <sub>as.east</sub>	0.00 **	0.73ns	0.00	0.00 **	0.08ns
	NDVI <sub>as.west</sub>	0.02 **	0.38ns	0.00 **	0.00	0.44ns
	Chl	0.66ns	0.42ns	0.08ns	0.44ns	0.00
90	NDVI <sub>ps</sub>	0.00	0.23ns	0.01 **	0.35ns	0.81ns
	NDVI <sub>as</sub>	0.23ns	0.00	0.13ns	0.00 **	0.08ns
	NDVI <sub>as.east</sub>	0.02 **	0.13ns	0.00	0.00 **	0.00 **
	NDVI <sub>as.west</sub>	0.35ns	0.00 **	0.00 **	0.00	0.87ns
	Chl	0.81ns	0.08ns	0.00 **	0.86ns	0.00
120	NDVI <sub>ps</sub>	0.00	0.01 **	0.00 **	0.62ns	0.46ns
	NDVI <sub>as</sub>	0.01 **	0.00	0.90ns	0.55ns	0.71ns
	NDVI <sub>as.east</sub>	0.00 **	0.90ns	0.00	0.00 **	0.00 **
	NDVI <sub>as.west</sub>	0.62ns	0.55ns	0.00 **	0.00	0.09ns
	Chl	0.47ns	0.71ns	0.00 **	0.09ns	0.00

Considering the data presented by the correlogram in Figure 5a, as well as the  $\rho$  values present in Table 4, it can be stated that the results of the 30-point sampling grid indicate that the NDVI<sub>as.east</sub> x NDVI<sub>ps</sub> presented a correlation of 52%, NDVI<sub>as.east</sub> x NDVI<sub>as.west</sub> of 72% and NDVI<sub>as.east</sub> x Chlorophyll of

50%.

In relation to the 60-point sampling grid, it can be observed that there was a significant correlation between  $NDVI_{ps} \times NDVI_{as.east}$  and  $NDVI_{ps} \times NDVI_{as.west}$ , with correlations of 37% and 29% respectively. Also, for this grid, a significant correlation of 60% can be observed when comparing the NDVI values measured by GreenSeeker on the east and west faces ( $NDVI_{as.east}$  and  $NDVI_{as.west}$ ).

When considering the sampling grid of 90 points, a significant correlation can be observed between  $NDVI_{ps} \times NDVI_{as.east}$  (25%) and a significant and inverse correlation of -31% when comparing  $NDVI_{as}$  and  $NDVI_{as.west}$ . The  $NDVI_{as.east}$  variable, in addition to its correlation with  $NDVI_{ps}$ , this variable also showed correlations of 67% and 31% with the  $NDVI_{as.west}$  and Chlorophyll variables.

For the 120-point sampling grid, significant correlations can be observed between  $NDVI_{ps} \times NDVI_{as}$  (23%) and  $NDVI_{ps} \times NDVI_{as.east}$  (24%), as well as between  $NDVI_{as.east} \times NDVI_{as.west}$  (69%) and  $NDVI_{as.east} \times$  Chlorophyll (48%).

A general comparison of the variables that were correlated, it was observed that the variable that presented the highest number of correlations was the  $NDVI_{as.east}$  (11 correlations) and it was also possible to observe that for the grids of 60, 90 and 120 points this variable presented a correlation significant with  $NDVI_{ps}$ . The variable that presented the lowest number of correlations was  $NDVI_{as}$  (only 2 significant correlations), this was already reflected in the discussion topic of the maps, where it was possible to observe the little variation in values of this variable. Regarding Chlorophyll, it presented only 3 significant correlations, and all cases were significantly correlated with the  $NDVI_{as.east}$  for the 30, 90 and 120 point grids.

Another noteworthy observation is the comparison between the  $NDVI_{as.east}$  and  $NDVI_{as.west}$  variables across all sample grids. A significant correlation was observed between these variables; however, despite their statistical significance, they exhibited moderate correlations (72% for the 30-point grid, 60% for the 60-point grid, 67% for the 90-point grid and 69% for the 120-point grid). This indicates that despite utilizing the same data collection method with the GreenSeeker and considering only the readings from the upper third of the plant, these variables did not display high correlations. This observation can be attributed to the environmental conditions during data collection. Specifically, the shaded west side resulted in lower NDVI values, while the east-facing side of the plants, illuminated by sunlight, generated higher NDVI values. Additionally, there were more correlations observed with the  $NDVI_{ps}$  variable, as the RPA image collection coincided with peak solar irradiance and absence of cloud cover.

Santos et al. (2022) investigated the effectiveness of various vegetation indices derived from multispectral imagery captured by a remotely piloted aircraft (RPA)-mounted sensor in predicting chlorophyll content in both coffee tree leaves ( $Chl_{leaf}$ ) and canopy ( $Chl_{dossel}$ ). They also examined the correlation between these indices and chlorophyll levels across different seasons (rainy and dry). The study reported a significant correlation of 61% between  $NDVI_{ps}$  and Total Chlorophyll during the dry season (the same season and chlorophyll content used in our study). However, our research did not observe a significant correlation between these same attributes. By collecting data such as height, diameter, and chlorophyll, and together with high-resolution multispectral images obtained by RPA, Bento et al. (2022) evaluated the relationship between vegetation indices with total chlorophyll (TC) content and leaf area index to characterize different types of coffee cultivars. To evaluate the correlation between these attributes and vegetation indices, the authors used the Spearman correlation index. Through correlation between the  $NDVI_{ps}$  and TC, the authors found inverse and non-significant correlations of -0.05, -0.15 and -0.18 for the cultivars Catucaí 2SL, Catucaí IAC62 and Bourbon IACJ10 respectively, corroborating the results of this study. work, despite the difference between age and cultivars of coffee trees.

Barata et al. (2023), evaluating coffee trees transplanted to areas with different liming methods (surface and deep) through field measurements and vegetation indices obtained by RPA images, carried out a correlation analysis between VI and parameters such as height, crown diameter, chlorophyll, leaf area index, chlorophyll a, chlorophyll b and total chlorophyll. The results obtained by the authors for the correlation between  $NDVI_{ps}$  and Total Chlorophyll do not show significant correlations and once again corroborate the results found in this research.

Gomes et al. (2021) compared the NDVI obtained by active (GreenSeeker) and passive sensors (Mica Sense and MAPIR). These authors found high and significant correlations (around 80 to 90%) when comparing the  $NDVI_{ps}$  (MicaSense multispectral sensor) in relation to the NDVI values obtained

by GreenSeeker. These results differ from those found in this research, which despite significant correlations between these attributes, none presented such high values.

The discrepancy between the findings of this study and those presented by Gomes et al. (2021) lies in the methodology used by the authors to collect the index. For each sampling point, the authors carried out 3 readings using GreenSeeker, but these readings were carried out in the coffee tree canopy (at a distance of 30cm from the plant canopy), therefore, the authors obtained values very close to those obtained by NDVI<sub>ps</sub>. In this study, it was not possible to obtain NDVI measurements using the GreenSeeker from the crown of the sample plants, since, as it was a very old coffee tree (around 25 years old), the plants were very tall (above 2m in height, on mean), making it possible to obtain only data from the upper third.

Campos et al. (2022), evaluating the modeling of NDVI in coffee trees through the use of a passive RGB sensor coupled to RPA, evaluated the correlation of the NDVI obtained by GreenSeeker with red, blue, green bands and the normalized relationship between the RGB sensor bands. The authors find inverse and significant correlations (mean of 70% correlation between the bands evaluated). Like Gomes et al. (2021), the authors carried out NDVI readings with GreenSeeker in the coffee tree canopy, at a distance of 1m, following a methodology recommended by Ali & Ibrahim (2020).

Enciso et al. (2019), evaluating the correlation of the GreenSeeker sensor and the NDVI obtained by a multispectral sensor embedded in an RPA, observed a non-significant correlation ( $\rho < 0.05$ ), the justification is that the GreenSeeker measurements consider the plant canopy, while the NDVI calculated from multispectral images consider the entire vegetative area, resulting in an R of less than 0.45.

## CONCLUSION

The geostatistical analysis was efficient to evaluate the spatial variability of chlorophyll and NDVI data obtained by the GreenSeeker and ClorofiLOG sensors. Through geostatistical analysis it was possible to model the data and generate semivariograms and perform ordinary kriging. The data resulting from ordinary kriging generated spatial distribution maps of these two attributes, and through visual analysis it was possible to observe the behavior of these two variables throughout the study area, indicating points of highest and lowest concentration of NDVI and chlorophyll attributes.

In addition to the maps generated by kriging, the flight and processing of images obtained by RPA resulted in the calculation and generation of the NDVI<sub>ps</sub> map. Considering the very high resolution of the images, it was possible to verify with detail and precision the spatial distribution of this vegetation index in the studied area, facilitating the identification of points with higher and lower concentrations of vegetative vigor.

By calculating the Pearson correlation coefficient (R), it was possible to find significant correlations between the attributes, even when evaluated in different sampling grids. The results generated indicate the effectiveness of using sensors in coffee crops, benefiting producers in making decisions regarding the management of their crops quickly and efficiently.

**ACKNOWLEDGEMENTS.** The authors would like to thank the financial support of the funding agencies CNPQ (project 305953/2020-6), FAPEMIG (projects PPE-00118-22 and BPD-00040-22), EMBRAPA Café - Consórcio Pesquisa Café (projects 10.18.20.023.00.00 and 10.18.20.041.00.00), CAPES, and the UFLA Postgraduate Program in Agricultural Engineering.

## REFERENCES

- Ali, A.M., & Ibrahim, S.M. 2020. Wheat grain yield and nitrogen uptake prediction using at Leaf and GreenSeeker portable optical sensors at jointing growth stage. *Information Processing in Agriculture*, 7 (3), 375–383.
- Alves, E.A., Queiroz, D.M. & Pinto, F.A.C. 2006. Precision coffee farming. In: Zambolim, L. ( ed): Boas agricultural practices in coffee production, e. g. 190 (in Portuguese).
- Amaral, L.R.D., Zerbato, C., Freitas, R.G.D., Barbosa Júnior, M.R. & Simões, I.O.P.D.S. 2020. UAV applications in Agriculture 4.0. *Revista Ciência Agrônômica*, 51, e20207748.

- Bachmaier, M. & Backers, M. Variogram or semivariogram? Understanding the variances in a variogram. 2008. *Precision Agriculture*, **9**, 173–175.
- Barata, R.A.P, Ferraz, G.A.S., Bento, N.L., Soares, D.V., Santana, L.S., Marin, D.B., Mattos, D.G., Schwerz, F., Rossi, G., Conti, L. & Bambi, G. 2023. Evaluation of Coffee Plants Transplanted to an Area with Surface and Deep Liming Based on Multispectral Indices Acquired Using Unmanned Aerial Vehicles. *Agronomy-Basel*, **13**, 2623.
- Barbosa, B. D. S., Ferraz, G. A. S., Gonçalves, L. M., Marin, D. B., Maciel, D. T., Ferraz, P. F. P., & Rossi, G. 2019. RGB vegetation indices applied to grass monitoring: A qualitative analysis. *Agronomy Research*, **17**(2), 349–357.
- Bento, N.L., Ferraz, G.A.S., Barata, RAP, Soares, DV, Santos, LM Dos., Santana, LS, Ferraz, PFP, Conti, L. & Palchetti, E. 2022. Characterization of Recently Planted Coffee Cultivars from Vegetation Indices Obtained by a Remotely Piloted Aircraft System. *Sustainability*, **14**, p.1446.
- Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B., Karlen, D.L., Turco, R.F. & Konopka, A.E. 1994. Field-scale variability of soil properties in central Iowa soils. *Soil Sci. Soc. Am. J.*, **58**, 1501–1511.
- Campos, G.A. De Oliveira., Alves, M. De Carvalho., Miranda, J. Da Rocha., Resende, M.L.V. & Carvalho, G.R. 2022. Modeling coffee NDVI using passive RGB sensor embedded in UAS. *Theoretical and Applied Engineering*, **6** (3), 1–11 (in Portuguese).
- Enciso, J., Avila, C.A., Jung, J., Elsayed-Farag, S., Chang, A., Yeom, J., ... & Chavez, J.C. 2019. Validation of agronomic UAV and field measurements for tomato varieties. *Computers and Electronics in Agriculture*, **158**, 278–283.
- Ferraz, G.A.S, Silva, F.M.D, Carvalho, L.C., Alves, M.D.C. & Franco, B.C. 2012. Spatial and temporal variability of phosphorus, potassium and productivity of a coffee crop. *Engineering Agrícola*, **32**(1), 140–150 (in Portuguese).
- Frogbrook, Z.L., Oliver, M.A., Salahi, M., & Ellis, R.H. 2002. Exploring the spatial relations between cereal yield and soil chemical properties and the implications for sampling. *Soil Use and Management*, **18** (1), 1–9.
- Gomes, A., Queiroz, D.M.D., Valente, D.S., Pinto, F.D.A.D.C., & Rosas, J.T. 2021. Comparing a single-sensor camera with a multisensor camera for monitoring coffee using crop unmanned aerial vehicles. *Agricultural Engineering*, **41**, 87–97.
- Gomes, F.P. & Garcia, C.H. 2002. *Statistics applied to agronomic and forestry experiments*. FEALQ, Piracicaba-Brazil, 305 pp (in Portuguese).
- Isaaks, E.H. & Srivastava, R.M. 1989. *An introduction to applied geostatistics*. New York: Oxford University Press, Oxford, 560pp.
- Jesus, M.H., Bredemeier, C., Vian, A.L., Almeida, D. & Silva, J.A. 2014. Variation in the vegetation index by normalized difference in corn as a function of productive potential and plant density. In: *Brazilian Congress of Precision Agriculture (ConBAP)*. São Pedro, São Paulo, BR (in Portuguese).
- Kim, K.H., Shawon, M.R.A., An, J.H., Lee, H.J., Kwon, D.J., Hwang, I.C., ... & Choi, K.Y. 2022. Effect of shade screen on sap flow, chlorophyll fluorescence, NDVI, plant growth and fruit characteristics of cultivated paprika in greenhouse. *Agriculture*, **12** (9), 1405.
- Li, G., Wan, S., Zhou, J., Yang, Z. & Qin, P. 2010. Leaf chlorophyll fluorescence, hyperspectral reflectance, pigments content, malondialdehyde and proline accumulation responses of castor bean (*Ricinus communis* L.) seeds to salt stress levels. *Industrial crops and products*, **31**, 13–19.
- Main, R., Cho, M.A., Mathieu, R., O'Kennedy, M.M., Ramoelo, A. & Koch, S. 2011. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS*, **66** (6), 751–761.
- Marengo, R.A. & Lopes, N.F. 2007. *Plant physiology: photosynthesis, respiration, water relations and mineral nutrition*. Editora UFV, Viçosa, Minas Gerais, Brazil, 469pp (in Portuguese).
- Marin, D.B, Alves, M.D.C, Pozza, E.A., Gandia, R, M., Cortez, M.L.J., & Mattioli, M.C. 2019. Multispectral remote sensing in the identification and mapping of biotic and abiotic variables in coffee. *Revista Ceres*, **66**, 142–153 (in Portuguese).
- Moreira, M.A. 2011. *Fundamentals of Remote Sensing and Application Methodologies*. Editora UFV, Viçosa, Minas Gerais, Brazil, 422pp (in Portuguese).
- Pereira, A., Ribeiro, G., Oliveira, A., Oliveira, A., Ribeiro, A., Oliveira, A., Oliveira, Z. & Yang, X. 2019. Modeling aboveground biomass of maize based in machine learning approaches using

- remote sensing data from UAVs. *Plant Methods*, **15**, 1–19.
- Ribeiro Junior, P.J. & Diggle, P.J. 2001. GeoR a package for geostatistical analysis. *R-News*, *New York*, **1(2)**, 14–18.
- Rodrigues, G.C., Grego, C.R., Luchiari, A., & Speranza, E.A. 2019. Spatial characterization of vegetation indices relative chlorophyll index in specialty coffee production areas in southern Minas Gerais. In: X Brazilian Coffee Research Symposium, Vitória, Espírito Santo, Brazil, 6pp (in Portuguese).
- Rouse, J.W., Haas, R.H.; Schell, J.A., Deering, D.W. & Harlan, J.C. 1973. Monitoring the vernal advancement of retrogradation of natural vegetation. *Greenbelt : National Aerospace Spatial administration*, 371pp.
- Santana, L.S., Ferraz, G.A e S., Santos, S.A. Dos. & Dias, J.E.L. 2022. Precision coffee growing: A review. *Coffee Science*, **17**, p. e172007.
- Santos, L. M., Ferraz, G. A. S., Andrade, M. T., Santana, L. S., Barbosa, B. D. S., Maciel, D. T., & Rossi, G. 2019a. Analysis of flight parameters and georeferencing of images with different control points obtained by RPA. *Agronomy Research*, **17(5)**, 2054–2063.
- Santos, P.L.F., Oliveira, R.M.M. & Gazola, R.P.D. 2019b. Photosynthetic pigments and their correlation with foliar nitrogen and magnesium in Bermuda grass grown in substrates. *Acta Iguçu*, **8**, 92–101 (in Portuguese).
- Santos, L. M., Ferraz, G. A. S., Carvalho, M. A. F., Vilela, M. S., & Estima, P. H. O. 2023. Preliminary study on the potential use of RPA images to quantify the influence of the defoliation after coffee harvesting to its yield. *Agronomy Research*, **21(S3)**, 1555–1566.
- Santos, L.M., Ferraz, G.A.S., Carvalho, M.A. De F., Teodoro, S.A., Campos, A.A.V. & Menicucci Neto, P. 2022. Use of RPA Images in the Mapping of the Chlorophyll Index of Coffee Plants. *Sustainability*, **14**, p. 13118.
- Shiratsuchi, L.S., Brandão, Z.N., Vicente, L.E., Victoria, D.D.C., Ducati, J.R., Oliveira, R.D. & Vilela, M.D.F. 2014. Remote Sensing: basic concepts and applications in Precision Agriculture. In: Bernardi, A.C. de C., Naime, J.M., Resende, A.V., Bassoi, L.H. & Inamasu, R. Y. ( eds), *Precision Agriculture (Results from a New Look)*. Embrapa Instrumentação (ALICE), São Carlos- São Paulo, Brazil, 58–73 (in Portuguese).
- Vieira, S.R. 2000. Geostatistics in studies of spatial soil variability. In: Novais, RF; Alvarez V., V.H; Schaefer, C.E.G.R. ( eds), *Special topics in soil sciences*. Brazilian Society of Soil Science, Viçosa, 1–54 (in Portuguese).
- Yu, K., Lenz-Wiedemann, V., Chen, X. & Bareth, G. 2014. Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. *ISPRS*, **97**, 58–77.
- Zanzarini, F.V., Pissarra, T.C., Brandão, F.J. & Teixeira, D.D. 2013. Spatial correlation of the vegetation index (NDVI) from Landsat /ETM+ images with soil attributes. *R. Bras. Agricultural Eng. Environmental*, **17(6)7**, 608–614.

Artigo 3 – Elaborado conforme a norma NBR 6022 (2018)	
Título do artigo	Desempenho de algoritmos de machine learning para a predição de produtividade de uma lavoura cafeeira
Autores	S.A.S. Silva, G.A.S. Ferraz, V.C. Figueiredo, M.M.L. Volpato, D.D. Ferreira, F.E.M. Borges

Fonte: elaboração própria.

## **DESEMPENHO DE ALGORITMOS DE MACHINE LEARNING PARA A PREDIÇÃO DE PRODUTIVIDADE DE UMA LAVOURA CAFEIEIRA**

S.A.S. Silva<sup>1</sup>, G.A.S. Ferraz<sup>1</sup>, V.C. Figueiredo<sup>2</sup>, M.M.L. Volpato<sup>2</sup>, D.D. Ferreira<sup>3</sup>, F.E.M. Borges<sup>3</sup>

<sup>1</sup> Universidade Federal de Lavras, Departamento de Engenharia Agrícola, Trevo Rotatório Professor Edmir Sá Santos, BR37200-900, Lavras, Brasil

<sup>2</sup> Empresa de Pesquisa Agropecuária de Minas Gerais, Av. José Cândido da Silveira 1647, Bairro União Belo Horizonte, BR31170-495 Belo Horizonte, Brasil

<sup>3</sup> Universidade Federal de Lavras, Departamento de Automática, Trevo Rotatório Professor Edmir Sá Santos, BR37200-900, Lavras, Brasil

### **RESUMO**

A aplicação de técnicas de machine learning na agricultura é uma ferramenta poderosa para aprimorar a previsão de produtividade, permitindo tomadas de decisões mais precisas e eficazes. Este estudo teve como objetivo avaliar o desempenho de diferentes algoritmos de aprendizado de máquina na predição de produtividade do café, considerando variáveis de manejo e atributos fisiológicos das plantas. A pesquisa foi conduzida em uma lavoura de *Coffea arabica* de 1,2 ha, situada no Campo Experimental da EPAMIG em Três Pontas, Minas Gerais, Brasil. Os dados foram coletados ao longo das safras 2020/2021 e 2021/2022, incluindo umidade do solo, potencial hídrico foliar, fertilidade do solo, nutrição das folhas, índices de vegetação, diâmetro da copa e índice de área foliar. Para a modelagem preditiva, os dados foram divididos em três cenários: Cenário 1 (banco de dados original, contendo todas as variáveis), Cenário 2 (banco de dados contendo somente as variáveis selecionadas pela correlação significativa com a produtividade) e Cenário 3 (banco de dados com as variáveis selecionadas pela correlação e aplicação de técnica de oversampling (SMOTE) para aumentar o banco de dados). Para cada cenário foram gerados modelos de regressão para as safras 2020/2021, 2021/2022 e safras combinadas (2020/2021 + 2021/2022). Para a modelagem, foram testados os algoritmos Random Forest (RF), Gradient Boosting (GB), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN) e Decision Tree (DT) utilizando a linguagem Python e a técnica de validação cruzada Leave-One-Out (LOO-CV). Os resultados da análise de correlação destacaram que variáveis como umidade gravimétrica, atributos químicos das folhas e os índices NDRE e NDVI apresentaram maior correlação com a produtividade do café. Em relação ao desempenho dos modelos, o KNN teve o melhor desempenho considerando as safras individuais dentro do cenário 1, no cenário 2 e 3, GB foi o algoritmo mais estável na geração dos modelos. A aplicação de algoritmos de machine learning na previsão da produtividade cafeeira pode ser uma ferramenta valiosa para otimizar a gestão agrícola, desde que sejam selecionadas variáveis relevantes e abordagens estatísticas adequadas para cada conjunto de dados.

**Palavras-chave:** algoritmos de regressão, agricultura digital, previsão de safra.

## 1. INTRODUÇÃO

A produção de café é uma cadeia de valor agrícola de alto perfil que fornece um meio de subsistência para milhões de produtores que estão concentrados na sua grande maioria nos países em desenvolvimento (Borrela et al., 2015), configurando como uma das maiores agroindústrias do mundo. No Brasil, o café é um dos principais produtos agrícolas na pauta de exportação, impactando um volume de recursos na balança comercial nacional e com grande influência em aspectos sociais e econômicos (Rocha et al., 2016). A renda de subsistência do agricultor varia conforme os rendimentos que ele obtém de suas colheitas.

No cenário nacional, a cultura cafeeira desempenha um papel significativo na economia do país, gerando renda e empregos substanciais (Sesso et al., 2021). Como maior produtor e exportador mundial de café, o Brasil enfrenta desafios para agregar valor à sua produção e manter a participação de mercado (Takano et al., 2020). A cadeia produtiva do café é sensível às flutuações cambiais, com variáveis de produção inversamente relacionadas às variações cambiais (Soares et al., 2021). As previsões de rendimento do café também podem ajudar governos e formuladores de políticas a proteger a agricultura e o comércio domésticos, oferecendo alertas antecipados sobre possíveis riscos de desastres e impactos de perda de safra, devido a eventos climáticos adversos (Kouadio et al., 2021).

Na busca de reforçar e apoiar o processo de tomada de decisão é essencial reconhecer a inter-relação real existente entre o rendimento da cultura e os vários fatores que o influenciam (Holzman et al., 2018, Whetton et al., 2017). As quatro variáveis mais importantes que afetam o rendimento das culturas são a fertilidade do solo, acessibilidade à água, clima, pragas e doenças. Esses fatores podem representar um risco eminente para os agricultores, quando não são verificados e supervisionados com precisão (Elavarasan et al., 2018).

No contexto de monitoramento e tomada de decisão, a previsão de safras utilizando aprendizado de máquinas pode ser uma ferramenta poderosa para os produtores. Ao analisar grandes volumes de dados históricos e atuais, como condições climáticas, qualidade do solo, práticas agrícolas e indicadores econômicos, os algoritmos de aprendizado de máquinas podem gerar previsões mais precisas sobre a produtividade das safras (Aono et al., 2022).

Na literatura alguns autores avaliaram o potencial dos algoritmos de machine learning na previsão de safra de café utilizando séries temporais de dados de satélites (Abreu Júnior et al., 2022; Martello et al., 2022), dados climáticos (Kittichotsatsawat, Tippayawong e Tippayawong, 2022; Carielo, Prestes e Marinho, 2021; Lorençone et al., 2021; Rodríguez et al., 2021, Freitas et al., 2024), atributos do solo (Faria et al., 2024; Kouadio et al., 2018), manejo

das culturas (Rodríguez et al., 2021; Freitas et al., 2024), índice e área foliar, diâmetro da copa e valores individuais das bandas RGB obtidas por Aeronave Remotamente Pilotada (Barbosa et al., 2021). Os resultados obtidos com o uso de machine learning na previsão de safra evidenciam sua relevância como ferramenta de apoio à tomada de decisão, especialmente em um setor altamente sensível a variações de oferta e demanda.

Com uma previsão mais confiável, os produtores podem se preparar melhor para as flutuações de mercado, ajustando suas estratégias de venda e negociação com base em uma estimativa mais realista da oferta futura de café. Isso ajuda a minimizar riscos e otimizar o retorno financeiro, proporcionando maior estabilidade e eficiência na gestão da produção.

Considerando que a estimativa da produtividade é uma informação valiosa para os cafeicultores, o objetivo deste artigo é avaliar a eficácia da previsão da produtividade de uma lavoura cafeeira utilizando técnicas de aprendizado de máquina, comparando diferentes modelos e abordagens para identificar aqueles que proporcionam maior precisão e confiabilidade nas estimativas de produtividade. Uma análise de variáveis mais relevantes para a estimativa da produtividade é também apresentada.

O artigo está organizado da seguinte forma: a Seção 2 descreve os materiais e métodos adotados neste estudo, começando pela caracterização da área experimental e das variáveis utilizadas para prever a produtividade dos cafeeiros. Em seguida, são apresentados os procedimentos de coleta e organização dos dados, além do detalhamento dos algoritmos de aprendizado de máquina aplicados. Por fim, são definidas as métricas utilizadas para avaliar o desempenho dos modelos, como  $R^2$  e MAPE.

## **2. MATERIAL E MÉTODOS**

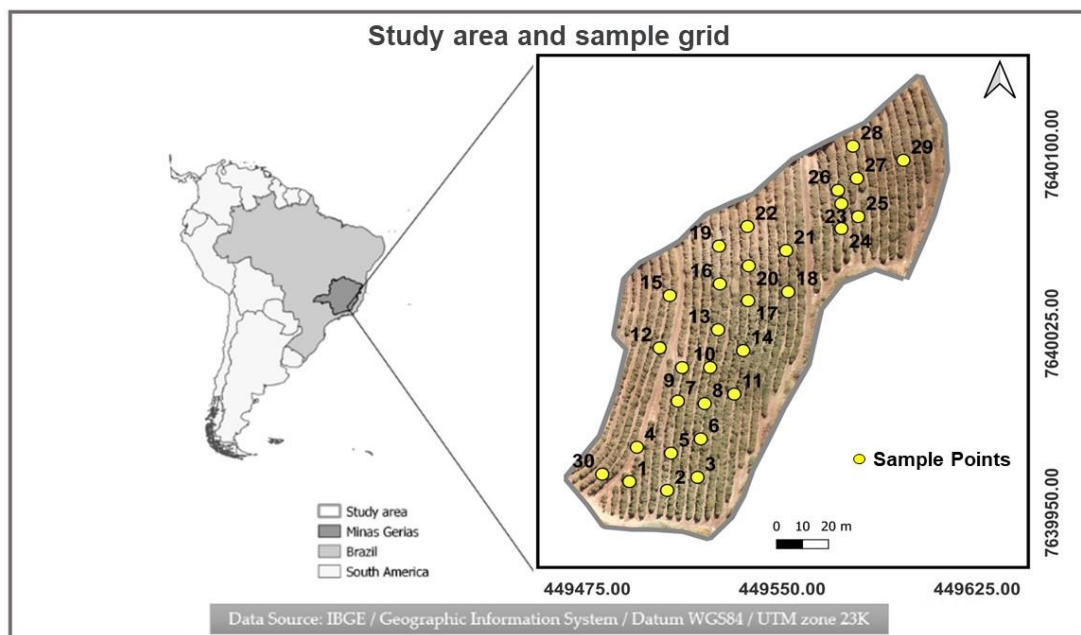
### **2.1 Caracterização da área de estudo, grade e amostragem**

O estudo foi desenvolvido em uma lavoura cafeeira situada no campo experimental da Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG), no município de Três Pontas, região sul do estado de Minas Gerais, Brasil, a uma altitude de 905 m e uma posição do sistema de coordenadas Universal Transversa de Mercator (UTM) de S 7640030.4 e E 449531.5, Zona 23K. Este município tem uma temperatura média anual de 20,3 °C e uma precipitação média anual de 1429 mm. O solo nesta área é classificado como latossolo vermelho (LV).

A lavoura possui 1,2ha de cafeeiros da espécie *Coffea arabica* L., pertencentes a cultivar Topázio MG1190. Esta lavoura foi implantada no ano de 1998, no espaçamento entre linhas de

3,70 m e 0,70 m entre plantas. Para este estudo, foi confeccionada no software QGis uma malha amostral de 30 pontos (vide Figura 1). Esta malha amostral representa uma densidade de 25 pontos por hectare. O contorno da área e o georreferenciamento dos pontos foram realizados por um (GNSS) Trimble RTK, modelo R8. Cada ponto amostral foi representado por uma planta, das quais foram devidamente identificadas.

Figura 1 - Localização da área de estudos



Fonte: Do Autor (2025)

## 2.2 Coleta de dados

Após o georreferenciamento dos pontos amostrais, os seguintes atributos foram amostrados na área: produtividade, umidade do solo, potencial hídrico foliar, fertilidade do solo, nutrição das folhas, além de voos realizados por uma Aeronave Remotamente Pilotada (RPA) equipada com sensor multiespectral.

## 2.3 Produtividade

Os dados de produtividade foram amostrados em junho de cada ano/safra (2020/2021 e 2021/2022) nos 30 pontos georreferenciados na área de estudo. Para cada ponto amostral (representado por uma planta de café) a produtividade foi obtida através da colheita semimecanizada com o auxílio de uma derriçadora. Após o uso do equipamento, os frutos

caíram sob panos, foi realizada a limpeza para separação de folhas e galhos, e em seguida os grãos foram depositados em um recipiente graduado para medir a produtividade em litros por planta (Figura 2).

Figura 2 - Colheita semimecanizada dos pontos georreferenciados



Fonte: Do Autor (2025)

### 2.3.1 Umidade do solo

A umidade do solo foi obtida pela retirada de amostras indeformadas de solo na profundidade de 0-20cm, com o auxílio de um trado. A amostragem foi realizada nas seguintes ocasiões:

- Safra 2020/2021: amostras de umidade referentes a agosto de 2020 (época de seca) e janeiro de 2021 (época chuvosa)
- Safra 2021/2022: amostras de umidade referentes a agosto de 2021 (época de seca) e janeiro de 2022 (época chuvosa)

Cada amostra indeformada retirada nos pontos georreferenciados, foi armazenada em sacos plásticos adequados para determinação da umidade, os quais foram identificados de acordo com a numeração de cada ponto demarcado na área, e em seguida enviadas para o laboratório.

Para obter a umidade gravimétrica, as amostras foram submetidas ao método da estufa (padronizado pela norma NBR 6457/2016). Cada amostra de solo foi pesada a fim de constatar a massa úmida, e posteriormente, foram armazenadas em uma estufa a 105° C por 24 horas. Após esse período, as amostras foram pesadas novamente para obter a massa seca. Com esses valores foi possível calcular a umidade gravimétrica através da Equação (1).

(1)

$$UG(\%) = \left( \frac{\text{massa úmida} - \text{massa seca}}{\text{massa seca}} \right) \times 100$$

### 2.3.2 Potencial hídrico

Para a coleta do potencial hídrico das folhas ( $\Psi_w$ ), utilizou-se o método da bomba de Scholander (Scholander et al., 1965), que é uma câmara de pressão utilizada para mensurar o potencial hídrico das folhas (tensão da seiva no xilema). As amostragens do potencial hídrico ocorreram na mesma época que a amostragem da umidade, safra 2020/2021 ( $\Psi_w$  referente a agosto de 2020 e janeiro de 2021) e safra 2021/2022 ( $\Psi_w$  referente a agosto de 2021 e janeiro de 2022).

Neste estudo, foi mensurado o  $\Psi_w$  de antemanhã (amostragens realizadas entre 04:00 a 06:00 horas da manhã). Esta medida pode indicar a disponibilidade hídrica no solo, uma vez que há uma tendência de equilíbrio entre as condições hídricas da planta e do solo, em situações em que o déficit hídrico não é severo (Silva et al., 2003).

Para cada ponto georreferenciado coletaram-se três folhas do terço médio, estando contidas entre o 3º ou 4º par de folhas. Tais folhas continham seus pecíolos intactos. Após a coleta, as mesmas foram identificadas e armazenadas para não ocorrer perda de umidade.

Em seguida a coleta, as amostras foram submetidas à bomba de Scholander (Figura 3). Este método consiste em colocar a folha dentro de um cilindro, fazendo com que somente o pecíolo fique para fora, atravessando um pequeno orifício revestido de borracha e a prova de gás. Uma pressão é gradativamente fornecida à câmara do cilindro até que se verifique o primeiro sinal de expulsão da seiva pelo pecíolo, e assim a injeção de gás (inerte-nitrogênio) deve ser interrompida e a leitura do manômetro deve ser registrada. O valor expresso pelo manômetro representa o potencial de pressão do xilema em Mpa (Mega Pascal).

Figura 3 - Medição do  $\Psi_w$  pela bomba de Scholander



Fonte: Do Autor (2025)

### **2.3.3 Fertilidade do solo**

As coletas das amostras para estabelecer a fertilidade do solo foram realizadas em abril de 2021 e abril de 2022. Para cada ponto amostral foram retiradas amostras na profundidade de 0-20cm. Posteriormente após as amostragens, as mesmas foram enviadas para um laboratório de análise de solo. Os atributos químicos analisados foram fósforo (P), fósforo remanescente (Prem), potássio (K), Cálcio (Ca), Magnésio (Mg), potencial hidrogeniônico (pH), matéria orgânica (MO) e acidez total (H + Al).

### **2.3.4 Nutrição das folhas**

A amostragem foliar em cafeeiros é realizada quando os grãos de café estão na fase “chumbinho”, fase esta, que ocorre entre dezembro e meados de janeiro.

Para cada ponto amostral foram coletadas 50 folhas (folhas pertencentes a planta amostral + plantas em seu entorno). As folhas coletadas estavam entre o terceiro e quarto par de folhas no terço médio de cada planta e foram selecionadas folhas que estavam livres de lesões causadas por doença ou deficiência, fenômenos climáticos e práticas culturais. Após a amostragem das folhas, estas foram enviadas para um laboratório de análises onde foram determinados os seguintes atributos: N, P, K, Ca, Mg, S, Mn, Zn, B, Cu e Fe.

Para este estudo a amostragem das folhas para estimar a concentração de seus atributos químicos foi realizada somente para a safra 2021/2022.

### **2.3.5 Imagens multiespectrais**

#### **2.2.6.1 Execução do voo**

Simultaneamente com as datas de coleta de umidade do solo e potencial hídrico foliar (agosto 2020, janeiro de 2021, agosto 2021 e janeiro de 2022), realizaram-se voos com uma RPA equipada com sensor multiespectral. Para os voos, utilizou-se uma aeronave, modelo eBee SQ, da fabricante senseFly (Figura 4) que possui as seguintes características: asa fixa, envergadura de 110 cm, alcance de rádio de 3 km nominais, velocidade de cruzeiro de 40-110 km/h, resistência ao vento de até 45 km/h (12 m/s), motor elétrico, carga útil máxima de 1,1 kg (incluindo câmera e baterias) e uma autonomia de voo de até 55 minutos.

Figura 4 - Aeronave remotamente pilotada eBee SQ



Fonte: Do Autor (2024)

O planejamento e execução do voo foram executados por meio da estação base, desenvolvida pela mesma fabricante da aeronave, com o seguinte conjunto: o software eMotion, responsável pela programação do voo e execução do trajeto da aeronave, e uma antena transmissora, que permite o acompanhamento em tempo real do sobrevoo, bem como o envio de comandos de pouso, mudanças de direção ou tomada de imagens.

A interface do programa mostra informações importantes sobre o nível de bateria, temperatura ambiente, altitude, posição, duração e velocidade do voo, velocidade do vento, resolução e sobreposição longitudinal e latitudinal da área a ser sobrevoada, altitude e link do rádio. O plano de voo seguiu os seguintes parâmetros:

- Distância focal: 3,98 mm
- Recobrimento vertical: 70%
- Recobrimento horizontal: 70%
- Altitude de voo: 50 m
- Tamanho do pixel: 4,71 cm px<sup>-1</sup>
- Velocidade: 12 m/s

#### 2.2.6.2 Processamento das imagens

O conjunto de imagens obtidas pelo sensor acoplado a RPA foi processado pelo software Pix4DMapper. O processamento digital envolveu várias etapas como: a fototriangulação dos blocos de imagens, que permitiu obter os parâmetros de orientação externa e interna das imagens; a geração da nuvem de pontos; a produção do Modelo Digital de Superfície (MDS); e a criação do ortomosaico.

Para a fototriangulação, foram utilizados cinco pontos de controle, estrategicamente distribuídos de maneira uniforme pela propriedade. Esses pontos foram levantados com o

auxílio de um GNSS modelo R8 da Trimble, operando em modo RTK (Real Time Kinematic) com dupla frequência e precisão inferior a 1 mm em uma banda de 1 Hz.

Posteriormente, foi realizada a correção radiométrica dos ortomosaicos, convertendo os valores de ND (Número Digital) em refletância de superfície. Esse processamento também foi efetuado no software Pix4DMapper, utilizando recursos do sensor Parrot Sequoia.

Antes do voo, realizou-se a calibração radiométrica da câmera com o uso de uma placa de referência. Com os valores obtidos dessa placa, o software foi capaz de calibrar e corrigir a refletância da imagem, levando em conta as condições de iluminação presentes no momento do voo.

O resultado final do processamento das imagens são ortomosaicos para as seguintes refletâncias: Red, RedEdge, Nir, Green e também um ortomosaico RGB e também foram gerados os modelos digitais de terreno e superfície (MDT e MDS), respectivamente.

### 2.2.6.3 Índices de Vegetação

Com o intuito de utilizar as informações das imagens multiespectrais, índices de vegetação foram calculados, a fim de amostrar seus valores compreendidos dentro de cada ponto amostral presentes na malha de amostragem. Através das bandas Nir, Red e RedEdge, os seguintes índices espectrais de vegetação foram calculados (Tabela 1).

Tabela 1 - Índices de vegetação

<b>Índice</b>	<b>Sigla</b>	<b>Função</b>	<b>Equações</b>	<b>Referências</b>
Normalized Difference Vegetation Index	NDVI	Medida da vegetação verde saudável.	$\frac{NIR + RED}{NIR - RED}$	Rouse et al. (1973)
Normalized Difference RedEdge	NDRE	Avaliar teor de clorofila nas plantas, assim como sua captação de nitrogênio e a demanda de fertilizantes.	$\frac{NIR - RED\ Edge}{NIR + RED\ Edge}$	Barnes et al., (2000)
Enhanced Vegetation Index 2	EVI2	Melhoramento do índice EVI, sem a faixa do azul, minimizando as distorções causadas pela atmosfera.	$2,5 * \frac{NIR - RED}{(NIR + 2,4 * RED + 1)}$	Jiang et al., (2008)

O cálculo dos índices de vegetação foi realizado no software QGis (v 3.22.9) pela ferramenta calculadora raster. Posterior ao cálculo de cada índice criou-se um shapefile contendo 30 polígonos de 30 cm cada, compreendendo exatamente os 30 pontos amostrais na área. Através desses polígonos, foram amostrados valores dos índices para cada ponto pela

ferramenta estatística zonal, também pelo software QGis.

A ferramenta estatística zonal calcula o valor médio dos pixels compreendidos dentro de uma limitação (cada polígono de 30cm). Para cada índice de vegetação calculado e submetido a ferramenta de estatística zonal, foram gerados 30 valores médios para cada um deles.

### 2.3.6 Altura de planta, diâmetro de copa e índice de área foliar dos cafeeiros

As estimativas das alturas das plantas, a partir das imagens UAV, foi realizada seguindo o fluxo de trabalho descrito por Panagiotidis et al. (2017), onde é calculada a diferença entre MDS e MDT (Equação (2)) e através da ferramenta amostragem de pontos é possível diferenciar os pixels dos modelos digitais.

(2)

$$H = MDS - MDT$$

Os diâmetros das copas foram estimados utilizando a metodologia empregada no trabalho de Barbosa et al. (2021). Após exportar o ortomosaico RGB para o software QGis, os valores do diâmetro médio para cada planta amostral foram selecionados manualmente através da caixa delimitadora.

O índice de área foliar (IAF) foi calculado com base na metodologia descrita por Favarin et al. (2002), por ser uma metodologia rápida e não destrutível, utilizando as informações de diâmetro da copa (D) e altura de planta (H) descritos na Equação (3).

(3)

$$IAF = 0,0134 + 0,7276 \times D^2 \times H$$

## 2.4 Análise do conjunto de dados

Após a coleta e processamento dos dados (produtividade, umidade do solo, potencial hídrico, fertilidade do solo, nutrição das folhas, índices de vegetação, altura das plantas, diâmetro da copa e índice de área foliar), foram elaborados três arranjos para gerar a modelagem da produtividade.

Arranjo 1: uma variável dependente (produtividade) e vinte e duas variáveis independentes, total de 23 colunas e 30 linhas;

Arranjo 2: uma variável dependente (produtividade) e trinta e três variáveis

independentes, total de 34 colunas (inclusão dos dados de nutrição das folhas) e 30 linhas;

Arranjo 3: junção dos dados do arranjo 1 e 2 (excluindo os dados de nutrição das folhas) totalizando uma tabela de 23 colunas e 60 linhas.

Os dados de produtividade foram submetidos a uma análise de estatística descritiva onde foram calculados os valores de: média, mínimo, máximo, desvio padrão, assimetria e coeficiente de variação (CV).

## 2.5 Seleção de atributos e normalização dos dados

A análise de correlação foi usada como uma seleção de atributos para avaliar o grau de correlação entre as variáveis preditoras (variáveis de entrada) e a variável a ser predita (variável alvo → produtividade). Dessa maneira é possível trabalhar com os modelos de previsão somente com as variáveis que têm maior impacto na produtividade. Esta seleção de variáveis é importante para garantir modelos menos complexos (densos) e evitar o mal da dimensionalidade, já que o número total de variáveis é maior do que o número de amostras disponíveis (Haykin, 2001).

A correlação de Pearson é uma medida estatística que quantifica a relação linear entre duas variáveis numéricas. Seu coeficiente, chamado Coeficiente de Correlação de Pearson (R), varia entre -1 e 1, e é calculado pela fórmula da Equação (4).

(4)

$$R = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \cdot \sum(Y_i - \bar{Y})^2}}$$

Onde:

R = coeficiente de correlação

$X_i$  = valores da variável X em uma amostra

$\bar{X}$  = média dos valores da variável X

$Y_i$  = valores da variável Y em uma amostra

$\bar{Y}$  = média dos valores da variável Y

Se:

$R=1$  → Correlação positiva perfeita (as variáveis aumentam juntas);

$R > 0$  → Correlação positiva (Valores altos de X tendem a estar associados a valores altos de Y);

$R = 0$  → sem correlação linear (as variáveis não têm relação linear clara).

$R < 0$  → Correlação inversa (quando uma variável aumenta, a outra tende a diminuir);

$R = -1$  → Correlação inversa perfeita (relação linear inversa perfeita).

Após a análise de correlação, os dados foram submetidos à normalização do tipo Z-score (ou padronização), um processo essencial quando se trabalha com Machine Learning, especialmente para algoritmos sensíveis à escala dos dados. Esse método ajusta os valores de forma que a média se torne 0 e o desvio padrão 1, garantindo que todas as variáveis possuam a mesma influência no modelo.

## 2.6 Modelos de previsão de produtividade

Para este estudo, foram realizadas análises de regressão múltipla com o auxílio de algoritmos de aprendizado de máquinas. O intuito é avaliar a capacidade desses algoritmos em gerar modelos de previsão de produtividade para cafeeiros, utilizando um conjunto de dados com número limitados de amostras.

A análise de regressão múltipla é uma técnica que permite explorar e inferir a relação de uma variável dependente (variável resposta) com variáveis independentes (variáveis explicatórias). Neste caso, as variáveis explicatórias são os dados de umidade do solo, potencial hídrico, fertilidade do solo, nutrição foliar, altura da planta, diâmetro de copa, índice de área foliar (LAI) e valores extraídos de índices de vegetação. Já a variável resposta é a produtividade.

Os dados obtidos para este estudo serão divididos de três formas (Tabela 2), para avaliar qual delas terá o melhor desempenho na previsão de produtividade.

Tabela 2 - Banco de dados para o aprendizado de máquina

Banco de dados	Variáveis explicatórias	Variável resposta
Safra 2020/2021 (1)	- GH e WP (agosto 2020 e janeiro 2021) - Fertilidade do solo: pH, K, P, Ca, Mg, H+Al, SB, MO e Prem (abril 2021) - Índices de vegetação: NDVI NDRE EVI2 (agosto 2020 e janeiro 2021) - Altura, Diâmetro e IAF (janeiro 2021)	Produtividade 2021
Safra 2021/2022 (2)	- GH e WP (agosto 2021 e janeiro 2022) - Fertilidade do solo: pH, K, P, Ca, Mg, H+Al, SB, MO e Prem (abril 2022) - Índices de vegetação: NDVI, NDRE e EVI2 (agosto 2021 e janeiro 2022) - Altura, Diâmetro e IAF (janeiro 2022) - Nutrição foliar: N, P, K, Ca, Mg, Mn, Zn, B, Cu e Fe (dezembro 2021)	Produtividade 2022
Safras Combinadas (1 + 2)	- GH e WP (agosto 2020 e 2021; janeiro 2021 e 2022) - Fertilidade do solo: pH, K, P, Ca, Mg, H+Al, SB, MO e Prem (abril 2021 e 2022) - Índices de vegetação: NDVI, NDRE e EVI2 (agosto 2020 e 2021; janeiro 2021 e 2022) - Altura, Diâmetro e IAF (janeiro de 2021 e 2022)	Produtividades 2021 e 2022

GH: Gravimetric Humidity, WP: Water Potential, pH: potencial hidrogeniônico, N: nitrogênio, K: potássio, P: fósforo, Ca: cálcio, Mg: magnésio, Mn: manganês, Zn: zinco, B: boro, Cu: cobre, Fe: Ferro, H+AL: acidez total, SB: soma de bases, MO: matéria orgânica, Prem: fósforo remanescente, NDVI: normalized difference vegetation index, NDRE: normalized difference red edge vegetation index, EVI2: Enhanced Vegetation Index, IAF: índice de área foliar

Para o conjunto de dados obtidos para este estudo, serão realizadas análises de regressão múltipla, testando e validando algoritmos de aprendizado de máquinas na linguagem Python. Os modelos serão desenvolvidos para validar a aplicação de algoritmos de machine learning na previsão da produtividade do café. O objetivo é identificar quais variáveis exercem maior impacto sobre a produção e construir modelos preditivos com base nesses fatores.

Os modelos serão elaborados anualmente com base em um conjunto de dados obtidos no ano anterior à safra, considerando as variáveis relacionadas ao solo, às características das plantas e aos índices espectrais. Essa abordagem permite gerar modelos mais ajustados à realidade da lavoura, fornecendo ao produtor informações atualizadas e relevantes que podem auxiliar no preparo e no planejamento para a safra seguinte.

Neste estudo, serão consideradas apenas variáveis manejáveis, incluindo umidade do solo, potencial hídrico foliar, fertilidade do solo, nutrição das folhas, índices de vegetação, diâmetro da copa e índice de área foliar. No entanto, é importante destacar que fatores ambientais, como clima, eventos climáticos extremos, precipitação e características naturais do solo, também exercem um papel fundamental na produtividade do cafeeiro. Embora não possam ser controlados diretamente, esses fatores influenciam o crescimento e desenvolvimento das plantas, interagindo com as variáveis de manejo e podendo afetar a eficácia das estratégias

adotadas na lavoura.

Para a geração dos modelos de regressão os dados foram organizados em três cenários e para cada cenário foram gerados modelos para as safras 2020/2021, 2021/2022 e safras combinadas.

- **Cenário 1:** Geração de modelos de regressão para todos os atributos disponíveis, utilizando o banco de dados original (considerando todas as variáveis);
- **Cenário 2:** Geração de modelos de regressão apenas para os atributos selecionados com base em suas correlações com a produtividade;
- **Cenário 3:** Geração de modelos de regressão apenas para os atributos selecionados com base em suas correlações com a produtividade e aplicando a Synthetic Minority Over-sampling Technique (SMOTE) desenvolvida por Chawla et al. (2002). A técnica foi utilizada para triplicar o número de amostras dos conjuntos de dados, gerando dados sintéticos.

Os modelos de regressão foram gerados a partir dos algoritmos: Random Forest (RF), Gradient Boosting (GB), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN) e Decision Tree (DT).

O algoritmo Random Forest (RF) é uma extensão da Decision Tree (DT), na qual várias árvores são construídas a partir de amostras do conjunto de dados, e suas previsões são combinadas. Esse algoritmo foi desenvolvido por Breiman (2001). Em relação à DT, o RF reduz o risco de sobreajuste e melhora a precisão, porém demanda uma interpretação mais detalhada e cuidadosa.

O Gradient Boosting (GB) é um algoritmo de aprendizado supervisionado baseado na técnica de boosting. Foi desenvolvido por Friedman (2001) e combina vários modelos fracos (geralmente árvores de decisão) para criar um modelo mais forte e preciso.

O Multilayer Perceptron (MLP) é um tipo de rede neural artificial feedforward, composta por múltiplas camadas de neurônios. Foi desenvolvido por Rumelhart, Hinton e Williams (1986). É amplamente utilizado em problemas de regressão e classificação e é um dos modelos mais básicos e populares dentro do aprendizado profundo.

O algoritmo K-Nearest Neighbors (KNN) é baseado em instâncias, fazendo previsões com base na média das k instâncias mais próximas no espaço de atributos (Fix; Hodges, 1951). Ele é de fácil compreensão e funciona bem para problemas não lineares. Para grandes conjuntos de dados, esse algoritmo pode ser computacionalmente intensivo e sensível ao valor de k. No

entanto, o modelo dispensa treinamento e, geralmente, apresenta bom desempenho para problemas com poucos dados.

A Decision Tree (DT) é um algoritmo de machine learning desenvolvido por Quinlan (1986). Trata-se de um modelo que divide o espaço dos dados em regiões baseadas em decisões do tipo "sim" ou "não", formando uma estrutura em forma de árvore. Cada nó da árvore representa uma decisão baseada em uma das variáveis independentes. Diferentemente do algoritmo de Regressão Linear Múltipla (RLM), a DT não requer suposições de linearidade e é capaz de capturar relações não lineares. Uma limitação desse algoritmo é que, caso não seja devidamente estruturado, o modelo pode sofrer de sobreajuste (*overfitting*).

## **2.7 Validação dos algoritmos de regressão múltipla**

Quando as análises são realizadas por regressão múltipla utilizando algoritmos de aprendizado de máquina, a validação cruzada (VC) desempenha um papel crucial para garantir que o modelo seja robusto e generalize bem para novos dados e, além disso, ajuda a garantir que o modelo não esteja ajustando-se demais aos dados de treino, e sim capturando padrões reais. A VC permite obter estimativas mais precisas do desempenho do modelo em novos dados, já que divide os dados em várias partes para testes e treinos.

Para a validação cruzada dos modelos de regressão, os dados foram divididos em 80% (treinamento) e 20% (teste). A técnica de validação utilizada foi a Leave-One-Out (LOO-CV). Ela é recomendada para conjunto de dados pequenos, pois maximiza o uso dos dados para treinamento.

O grande problema da regressão é que em boa parte dos casos ela não é representada de forma linear. Levando em consideração que as regressões são baseadas na soma de erros, para avaliar os algoritmos utilizados são adotadas as seguintes métricas: raiz quadrada do erro médio ao quadrado (RMSE) e erro médio absoluto percentual (MAPE).

## **3. RESULTADOS E DISCUSSÃO**

### **3.1 Estatística Descritiva**

O conjunto de dados relacionados à produtividade foi analisado quanto à média, mínimo-máximo, desvio padrão, assimetria e coeficiente de variação (CV), apresentados na Tabela 3.

Tabela 3 - Estatística descritiva das safras 2020/2021 e 2021/2022

<b>Estatística</b>	<b>Safra 2020/2021</b>	<b>Safra 2021/2022</b>
<b>Média</b>	10,20	5,47
<b>Mínimo - Máximo</b>	1,00 – 22,00	0,00 – 22,00
<b>Desvio Padrão</b>	4,79	5,25
<b>Assimetria</b>	0,43	1,12
<b>CV (%)</b>	46,96	96,15

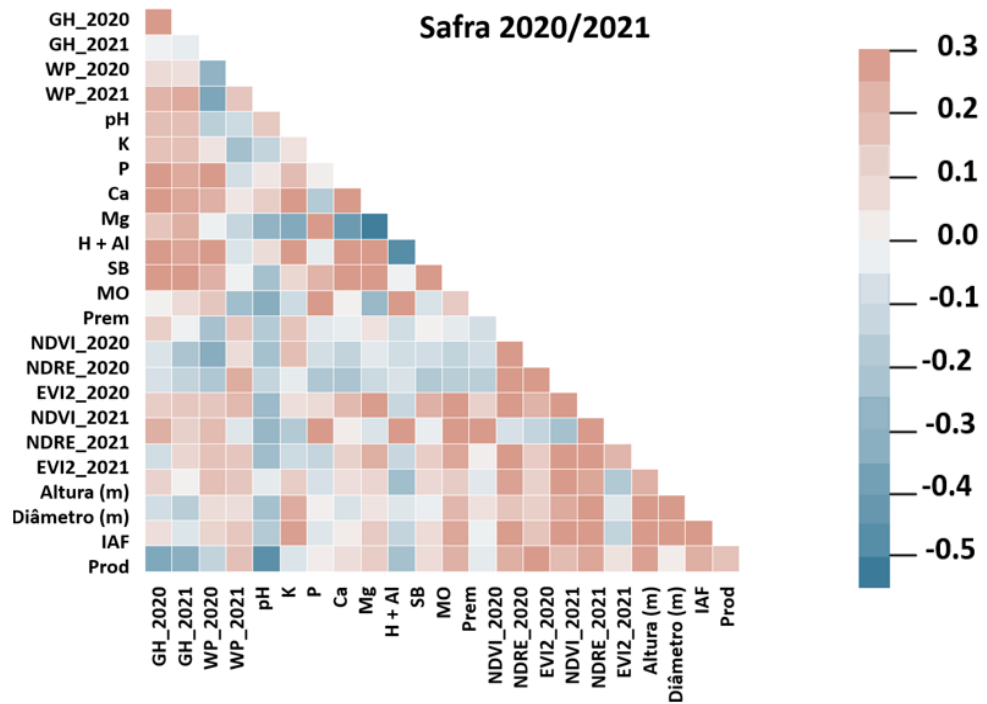
A estatística descritiva para as safras 2020/2021 e 2021/2022 demonstra uma queda significativa na média da produção, de 10,20 litros/planta para 5,47 litros/planta. A amplitude de variação é semelhante, mas o desvio padrão aumentou ligeiramente, indicando maior dispersão dos dados na safra 2021/2022. A assimetria passou de 0,43 para 1,12, sugerindo que a distribuição dos valores da safra 2021/2022 está mais distorcida à direita, possivelmente devido a muitos valores baixos e alguns extremos elevados. Além disso, o coeficiente de variação (CV%) quase dobrou (46,96% para 96,15%), reforçando a maior variabilidade na safra 2021/2022.

De acordo com Carvalho et al. (2004), a variabilidade espacial e temporal da produtividade das plantas é um problema recorrente nas regiões produtoras de café. Além das diferenças entre plantas vizinhas, a bienalidade exerce um impacto significativo na produção. Durante os anos de alta produtividade, o elevado consumo das reservas energéticas da planta compromete o desenvolvimento de novos ramos produtivos, resultando em uma redução na produção no ciclo seguinte (DaMatta et al., 2007; Silva et al., 2008). A limitação no crescimento dos ramos plagiotrópicos, responsáveis pela frutificação, acentua a alternância entre safras abundantes e escassas. Além disso, a bienalidade interfere diretamente na previsão de safra, tornando a estimativa de produtividade um desafio para os produtores (Miranda, Reinato e Silva, 2014).

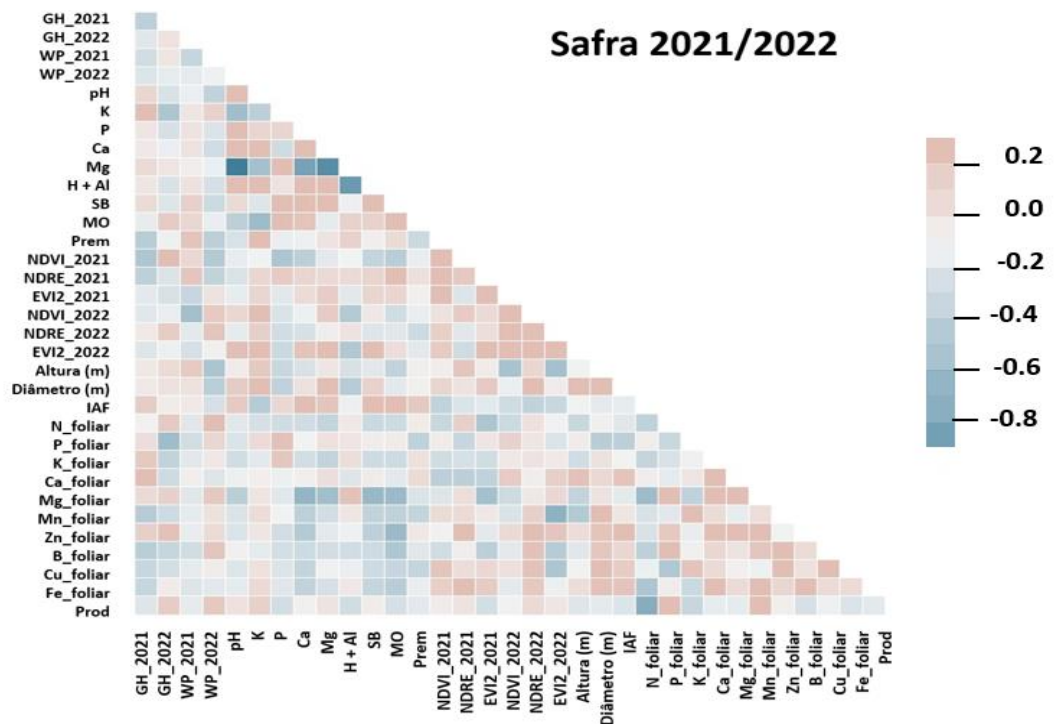
### 3.2 Análise de Correlação

As Figuras 5a, 5b e 6 representam os gráficos de correlação para os cenários 1, 2 e 3, respectivamente.

Figura 5- Gráficos de correlação para as safras 2020/2021 (a) e 2021/2022 (b)



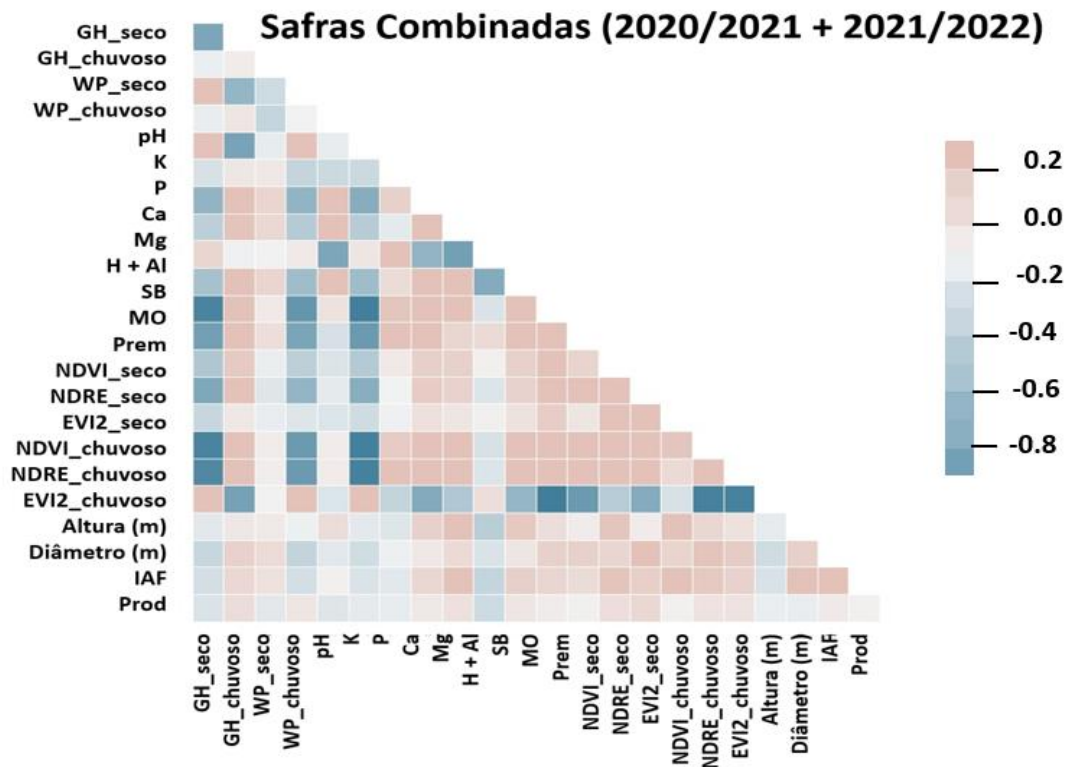
(a)



(b)

Fonte: Do Autor (2025)

Figura 6- Gráficos de correlação para safras combinadas (2020/2021 + 2021/2022)



Fonte: Do Autor (2025)

Para a safra 2020/2021 foram selecionados os atributos GH\_2020, GH\_2021, pH (correlação inversa variando entre 0,3 e 0,5) e NDRE (correlação positiva de aproximadamente 0,3). Ou seja, os atributos que tiveram maior impacto de correlação para o arranjo de dados da safra 2020/2021 foi a umidade gravimétrica amostrada em agosto de 2020 e janeiro de 2021, o pH e o índice NDRE referente a agosto de 2020.

Para a safra 2021/2022 foram selecionados os atributos WP\_2022, P\_foliar e S\_foliar (correlação positiva entre 0,2 e 0,4, respectivamente) e os atributos N\_foliar e K\_foliar (correlação inversa de aproximadamente 0,5 para N e 0,2 para K). Para a safra 2021/2022 as variáveis de nutrição foliar tiveram um maior impacto em relação a correlação juntamente com os valores do potencial hídrico foliar obtido em janeiro de 2022.

Para as safras combinadas os resultados da correlação dos atributos com a produtividade não foram tão significativos. Os atributos NDRE\_2020, GH\_2021, NDVI\_2020, H+Al e GH\_2020 (correlação positiva entre 0,1 e 0,2 aproximadamente).

Os resultados deste estudo, de certa forma, corroboram pesquisas anteriores que investigaram a relação entre variáveis espectrais, atributos do solo e nutrição foliar na

produtividade do café.

Assim como observado por Martello et al. (2022), que encontraram correlações positivas significativas entre NDVI e GNDI com a produtividade para imagens adquiridas um ano antes da colheita, este estudo também identificou correlações positivas entre NDRE\_2020 e a produtividade, indicando a influência do vigor vegetativo da planta em estágios fenológicos anteriores.

A relação entre bandas espectrais e produtividade também foi destacada por Abreu Júnior et al. (2022), que identificaram correlações positivas de até 0,72 para bandas visíveis e correlações inversas para os índices TCARI e GNDVI, o que sugere que diferentes índices espectrais podem capturar distintas condições fisiológicas da planta, influenciando sua produtividade.

No que se refere aos atributos nutricionais, os resultados obtidos para a safra 2021/2022 mostraram correlações significativas para os nutrientes foliares N, K, P e S, corroborando os achados de Faria et al. (2024), que também identificaram relações inversas para N e K e positivas para P e S, reforçando a importância do equilíbrio nutricional na produtividade do cafeeiro.

Além disso, os valores de correlação observados neste estudo para alguns atributos do solo, como H+Al e matéria orgânica, estão próximos dos valores relatados por Kouadio et al. (2018), que encontraram correlações entre 0,24 e 0,26 para atributos químicos do solo e a produtividade do café. Entretanto, ao analisar as safras combinadas, verificou-se que a correlação entre os atributos e a produtividade foi menos expressiva, o que pode indicar que os efeitos dessas variáveis apresentam variação temporal e sazonal, exigindo uma abordagem mais detalhada na modelagem preditiva para diferentes ciclos produtivos do café.

### **3.3 Modelos de previsão**

As Tabelas 4, 5 e 6 representam os valores de RMSE e MAPE (teste e treinamento) para cada algoritmo, considerando os cenários 1 (Tabela 4), 2 (Tabela 5) e 3 (Tabela 6).

Tabela 4 - Métricas RMSE e MAPE para as Safras 2020/2021, 2021/2022 e safras combinadas (2020/2021 e 2021/2022) considerando o banco de dados original

Modelos	Banco de dados original											
	Safra 2020/2021				Safra 2021/2022				Safras Combinadas 2020/2021 + 2021/2022			
	Treinamento		Teste		Treinamento		Teste		Treinamento		Teste	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
<b>RF</b>	4,47 ± 3,62	-	6,54	3,38	8,02 ± 8,12	4,50 ± 6,11	8,59	-	5,54 ± 5,60	-	11,11	0,98
<b>GB</b>	4,72 ± 3,72	-	9,86	4,02	8,81 ± 9,27	4,74 ± 9,61	13,66	-	5,35 ± 5,75	-	11,23	1,15
<b>MLP</b>	5,71 ± 4,06	-	5,54	3,11	9,35 ± 10,62	3,97 ± 4,95	10,46	-	6,42 ± 5,70	-	12,41	1,53
<b>KNN</b>	4,45 ± 3,87	-	4,21	2,51	6,71 ± 8,60	2,27 ± 2,59	12,19	-	6,38 ± 5,94	-	11,80	1,13
<b>DT</b>	4,96 ± 3,87	-	13,06	6,16	10,30 ± 10,77	6,41 ± 15,57	15,41	-	6,58 ± 7,38	-	13,35	1,71

Uma análise comparativa dos modelos RF, GB, MLP, KNN e DT, representados na Tabela 4, revelou diferenças significativas no desempenho preditivo, com base nas métricas RMSE e MAPE para as safras avaliadas.

Devido à presença de pontos georreferenciados com produtividade zero (0 litros/planta) e à utilização da validação Leave-One-Out, na qual cada iteração considera apenas uma amostra, alguns valores de MAPE foram desconsiderados. Isso ocorre porque a equação do MAPE tem como denominador os valores reais de produtividade ( $y_i$ ), e quando  $y_i = 0$ , a fração se torna indefinida, inviabilizando seu cálculo.

Na safra 2020/2021, os resultados de teste indicam que o modelo KNN apresentou um menor valor de RMSE (4,21 L plantas<sup>-1</sup>), demonstrando maior precisão nas previsões absolutas. Além disso, o KNN obteve o menor MAPE (2,51%), indicando o menor erro percentual relativo às previsões. Por outro lado, o DT apresentou o maior valor de RMSE (13,06 L plantas<sup>-1</sup>) e o maior MAPE (6,16%) evidenciando um desempenho significativamente inferior em comparação aos demais modelos.

Em relação à safra 2021/2022 o RF demonstrou o melhor desempenho em termos de RMSE nos valores de teste (8,59 L plantas<sup>-1</sup>), indicando previsões mais precisas. Já o KNN apresentou o menor MAPE nos resultados de treinamento (2,27%), evidenciando que o modelo conseguiu ajustar bem os dados dessa etapa. Novamente o DT para essa safra apresentou o pior desempenho com RMSE de (15,41 L plantas<sup>-1</sup>) no teste e um MAPE de 6,41% no treinamento reforçando sua baixa capacidade de generalização. O GB também demonstrou um RMSE elevado de (13,66 L plantas<sup>-1</sup>) no teste, sugerindo que este modelo pode não ter sido suficientemente robusto para lidar com os dados desta safra.

Ao considerar os dados combinados das duas safras, o RF apresentou melhores resultados gerais, obtendo um RMSE de 11,11 L plantas<sup>-1</sup> e o menor MAPE (0,98%) no teste. Esses resultados indicam que o RF foi o modelo mais eficiente ao considerar todas as

informações disponíveis, conseguindo generalizar melhor suas previsões. O DT manteve seu padrão de desempenho inferior, apresentando o maior RMSE (13,35 L plantas<sup>-1</sup>) e o maior MAPE (1,71%) no conjunto de testes, reforçando sua limitação para a modelagem da produtividade considerando este banco de dados.

Em todas as safras o DT apresentou os piores resultados de desempenho considerando as métricas de validação.

Tabela 5 - Métricas RMSE e MAPE para as Safras 2020/2021, 2021/2022 e safras combinadas (2020/2021 e 2021/2022) considerando somente as variáveis selecionadas em relação às suas correlações com a produtividade para cada safra.

Banco de dados com variáveis selecionadas (variáveis com as maiores correlações com a produtividade)												
Modelos	Safras 2020/2021 (GH_2020, GH_2021, pH e NDRE)				Safras 2021/2022 (N_foliar, S_foliar, P_foliar, WP_2022 e K_foliar)				Safras Combinadas 2020/2021 + 2021/2022 (H+AI, NDRE_2020, GH_2020, GH_2021, NDVI_2020)			
	Treinamento		Teste		Treinamento		Teste		Treinamento		Teste	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
<b>RF</b>	5,01 ± 3,73	-	2,61	1,43	6,92 ± 7,89	-	9,68	-	6,97 ± 7,65	-	5,26	3,27
<b>GB</b>	5,73 ± 3,99	-	2,66	0,89	7,52 ± 9,57	-	9,14	-	7,97 ± 8,86	-	5,53	2,48
<b>MLP</b>	4,72±3,52	-	1,78	0,74	11,92 ± 13,45	-	15,58	-	8,06 ± 6,91	-	5,34	1,80
<b>KNN</b>	4,51 ± 3,00	-	2,76	1,55	6,77 ± 7,95	-	11,18	-	6,30 ± 6,70	-	5,26	2,26
<b>DT</b>	6,20 ± 4,92	-	3,02	0,66	8,34 ± 8,12	-	15,92	-	8,41 ± 9,41	-	10,88	3,79

Em relação aos resultados representados pela Tabela 5 foi possível observar que os cinco algoritmos de regressão apresentaram variações consideráveis em desempenho em relação às safras analisadas.

Considerando a safra 2020/2021 o menor erro de treinamento foi representado pelo KNN (4,51 L plantas<sup>-1</sup>), seguido pelo MLP (4,72 L plantas<sup>-1</sup>), indicando que esses modelos se ajustaram melhor aos dados. Nos resultados de teste o MLP apresentou um desempenho considerável com o menor valor de RMSE (1,78 L plantas<sup>-1</sup>) e o segundo menor valor de MAPE (0,74%). O DT apresentou o pior desempenho, com valores de RMSE de 6,20 L plantas<sup>-1</sup> no teste e 3,02 L plantas<sup>-1</sup> no treinamento, sugerindo que este modelo não foi adequado para este conjunto de dados.

Referentes aos dados apresentados para a safra 2021/2022 pode-se afirmar que o menor erro de treinamento foi novamente para o modelo KNN (RMSE= 6,77 L plantas<sup>-1</sup>), seguido pelo RF (6,92 L plantas<sup>-1</sup>). No teste, o GB apresentou o menor erro (RMSE= 9,14 L plantas<sup>-1</sup>). O DT novamente apresentou o pior desempenho para esta safra, com RMSE= 15,92 L plantas<sup>-1</sup>, indicando baixa precisão.

Em relação aos resultados apresentados pelas safras combinadas (2020/2021 + 2021/2022) o KNN obteve o menor RMSE no treinamento (6,30 L plantas<sup>-1</sup>), seguido pelo RF

(6,97 L plantas<sup>-1</sup>). No teste, os algoritmos RF e KNN apresentaram o menor valor de RMSE (5,26 L plantas<sup>-1</sup>) seguidos pelo algoritmo MLP com RMSE de (5,34 L plantas<sup>-1</sup>) o qual apresentou o menor valor de MAPE (1,80%). Novamente, o DT apresentou o pior desempenho com RMSE de  $8,41 \pm 9,41$  no treinamento e RMSE de 10,88 L plantas<sup>-1</sup> e MAPE= 3,79% no teste, sugerindo que ele não conseguiu modelar bem os dados combinados.

Tabela 6 - Métricas RMSE e MAPE para as Safras 2020/2021, 2021/2022 e safras combinadas (2020/2021 e 2021/2022) considerando somente as variáveis selecionadas em relação as suas correlações com a produtividade e com o uso da técnica SMOTE para triplicar o banco de dados.

Banco de dados com variáveis selecionadas utilizando a técnica SMOTE (triplicar o volume de dados)												
Modelos	Safr 2020/2021 (GH_2020, GH_2021, pH e NDRE)				Safr 2021/2022 (N_foliar, S_foliar, P_foliar, WP_2022 e K_foliar)				Safras Combinadas 2020/2021 + 2021/2022 (H+AI, NDRE_2020, GH_2020, GH_2021, NDVI_2020)			
	Treinamento		Teste		Treinamento		Teste		Treinamento		Teste	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
<b>RF</b>	1,39 ± 2,31	-	2,09	0,83	1,75 ± 2,29	-	9,04	-	2,69 ± 4,63	-	6,89	4,41
<b>GB</b>	1,24 ± 2,49	-	2,28	0,88	0,91 ± 1,99	-	12,42	-	2,56 ± 4,64	-	7,84	4,24
<b>MLP</b>	1,72 ± 2,41	-	2,82	1,77	2,29 ± 2,11	-	13,11	-	3,34 ± 4,06	-	5,61	2,26
<b>KNN</b>	1,74 ± 2,23	-	5,14	2,90	3,15 ± 4,12	-	17,64	-	3,32 ± 4,97	-	5,49	2,88
<b>DT</b>	0,90 ± 2,67	-	3,14	0,67	1,06 ± 3,86	-	10,53	-	2,15 ± 5,98	-	11,31	6,37

Em relação à safra 2020/2021 RF e GB tiveram os menores RMSE no teste (2,09 e 2,28 L plantas<sup>-1</sup>, respectivamente) indicando um bom desempenho absoluto. O menor valor de MAPE no conjunto de dados de teste foi verificado no modelo DT (0,67%), sugerindo que suas previsões são percentualmente mais precisas. O KNN teve a maior diferença de RMSE entre treino e teste (+3,40 L plantas<sup>-1</sup>), sugerindo um problema de underfitting. O DT obteve um RMSE muito baixo no treino (0,90 L plantas<sup>-1</sup>) e um aumento significativo no teste (+2,24 L plantas<sup>-1</sup>).

De acordo com os resultados apresentados para a safra 2021/2022 pode-se observar que todos os modelos apresentaram aumento expressivo no RMSE de teste, indicando overfitting severo. O KNN teve o maior aumento no erro (+14,49 L plantas<sup>-1</sup>), sugerindo que se ajustou demais ao treino. RF obteve o menor RMSE no teste (9,04 L plantas<sup>-1</sup>), mas ainda assim é um erro alto, sugerindo que o modelo não conseguiu generalizar bem.

Já em relação aos resultados das safras combinadas o menor RMSE no teste foi de (5,61 L plantas<sup>-1</sup>) para o algoritmo MLP, indicando um melhor desempenho na previsão absoluta, quando comparado com os demais valores. O menor MAPE foi de 2,26% representando pelo modelo MLP pelo conjunto de dados de teste. O DT teve um aumento expressivo no RMSE de teste (+9,16 L plantas<sup>-1</sup>), indicando overfitting. RF e GB mantiveram estabilidade, mas apresentaram valores de erro absoluto mais altos.

A aplicação da técnica SMOTE resultou na redução significativa dos erros de treinamento em todos os cenários avaliados, o que já era esperado devido ao balanceamento e ampliação do conjunto de dados. No entanto, para os dados de teste, esse comportamento não se repetiu de forma uniforme. Na safra 2020/2021, o SMOTE contribuiu para a melhora do desempenho dos modelos, com redução expressiva nos valores de RMSE de teste em comparação aos cenários anteriores. Em contrapartida, na safra 2021/2022, os erros de teste aumentaram consideravelmente, evidenciando indícios de overfitting. Já nas safras combinadas, embora os valores de RMSE de teste também tenham aumentado, o crescimento foi menos acentuado. Portanto, a técnica SMOTE demonstrou benefícios limitados ao cenário da safra 2020/2021, não sendo tão eficaz para os demais contextos avaliados.

Ao comparar as três situações é possível afirmar que:

a) Safra 2020/2021:

- Banco de Dados original (Cenário 1): obteve os maiores valores de RMSEs no teste, com o DT atingindo um erro muito alto de 13,06 L plantas<sup>-1</sup>, indicando overfitting extremo;
- Variáveis selecionadas (Cenário 2): melhorou significativamente os resultados. O RMSE caiu para valores abaixo de 3,5 L plantas<sup>-1</sup> para todos os modelos, com o MLP tendo o menor RMSE (1,78 L plantas<sup>-1</sup>) e segundo menor valor de MAPE (0,74%);
- Utilizando a técnica SMOTE para as variáveis selecionadas (Cenário 3): todos os valores de RMSE de treinamento reduziram em comparação com o Cenário 2. Já em relação aos resultados de teste em comparação com o Cenário 2, os valores de RMSE e MAPE reduziram para os modelos RF e GB.

b) Safra 2021/2022:

- Banco de Dados original (Cenário 1): O RF apresentou o menor RMSE de teste nos cenários 1 e 2, sendo o segundo menor entre os modelos no cenário 2, mas os valores do erro ainda são muito altos (8,59 L plantas<sup>-1</sup>) considerando o cenário 1;
- Variáveis selecionadas (Cenário 2): a seleção de variáveis não trouxe grandes melhoras para o desempenho dos algoritmos;
- Utilizando a técnica SMOTE para as variáveis selecionadas (Cenário 3): apesar dos valores de RMSE de treinamento terem caído bastante em relação ao cenário 2, os valores de RMSE obtiveram um aumento muito significativo nos resultados

de teste, sugerindo overfitting.

c) Safras Combinadas (2020/2021 + 2021/2022):

- A junção de dados das duas safras não trouxe melhoras significativas nos modelos, com desempenho semelhante ou inferior ao observado nas safras individuais, especialmente nos valores de RMSE e MAPE no teste.

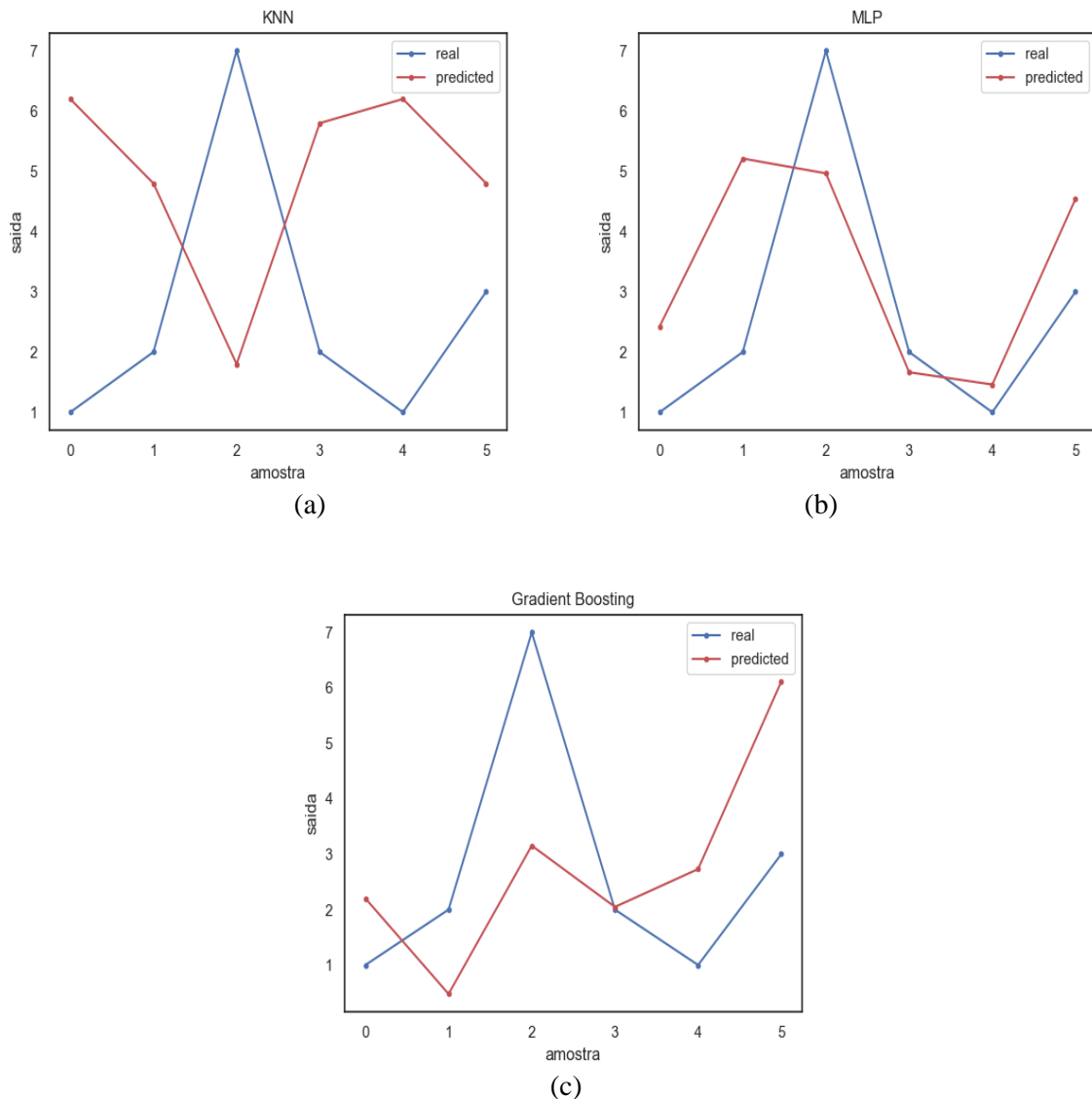
De forma geral, os resultados indicam que o uso de variáveis com alta correlação com a produtividade contribuiu para melhorar o desempenho dos modelos. No entanto, a aplicação do SMOTE, apesar de reduzir os erros de treinamento, deve ser utilizada com cautela, visto que sua eficácia variou entre as safras e apresentou tendência ao overfitting, especialmente para dados mais recentes (2021/2022).

Dentre todos os algoritmos avaliados para cada cenário, selecionou-se os três melhores modelos para cada cenário. As figuras 7a, 7b e 7c representam os gráficos de dispersão dos melhores modelos de previsão considerando dados de teste. Os gráficos de dispersão representam os valores reais de produtividade no eixo X e os valores preditos de produtividade no eixo Y. Se o modelo for perfeito, os valores estarão alinhados na diagonal.

Devido ao tamanho reduzido do conjunto de dados disponível, foi possível obter apenas seis amostras para o teste dos modelos (20% dos dados originais) de previsão de produtividade. Esse número limitado de amostras pode impactar a capacidade de avaliação do desempenho preditivo, tornando os resultados mais sensíveis a variações individuais dos dados. No entanto, mesmo com um conjunto restrito, os gráficos comparando os valores reais e preditos fornecem uma visão inicial sobre a precisão dos modelos, permitindo identificar padrões e possíveis ajustes para futuras melhorias.

Para o cenário 1, o modelo que representou o melhor desempenho foi do algoritmo KNN para a safra 2020/2021, com RMSE de treinamento de 4,45 L plantas<sup>-1</sup> e um RMSE de teste de 4,21 L plantas<sup>-1</sup>, além de um MAPE de 2,51%. Para o cenário 2, o MLP apresentou um melhor desempenho também para a safra 2020/2021 com RMSE de treinamento de 4,72 L plantas<sup>-1</sup> e um RMSE de 1,78 L plantas<sup>-1</sup> de teste, além de um MAPE de 0,74%. No cenário 3 o melhor desempenho foi para o GB, com RMSE de teste e treinamento de 1,24 e 2,28 L plantas<sup>-1</sup> respectivamente e um MAPE de teste de 0,88% resultantes da safra 2020/2021.

Figura 7- Gráficos de dispersão para os melhores modelos para os cenários 1 (a), 2 (b) e 3 (c)



Fonte: Do Autor (2025)

No geral, o banco de dados original apresentou os piores resultados. A seleção de variáveis melhorou significativamente os resultados para 2020/2021, mas não ajudou tanto para 2021/2022, confirmando que a inserção de dados de nutrição foliar para a safra 2021/2022 não trouxe grandes melhorias aos modelos, quando comparada com a safra 2020/2021 (safra que não possuía dados de nutrição foliar). O uso da técnica SMOTE trouxe ganhos para 2020/2021, mas gerou overfitting severo para 2021/2022. RF foi o modelo mais estável nas diferentes abordagens, enquanto DT teve os piores desempenhos gerais.

Até o momento, não foram encontrados artigos que utilizassem exatamente o mesmo arranjo de variáveis preditoras empregadas neste estudo. No entanto, algumas pesquisas que exploram variáveis similares, como imagens multiespectrais obtidas por satélites e aeronaves remotamente pilotadas, índices de vegetação, fertilidade do solo, diâmetro da copa e índice de

área foliar, podem fornecer referências relevantes. Dessa forma, a discussão dos resultados será embasada na comparação com esses estudos, permitindo uma análise mais abrangente sobre a eficácia das variáveis utilizadas na predição da produtividade.

Entre os estudos analisados, destaca-se o de Abreu Júnior et al. (2022), que utilizou um banco de dados com 208 pontos amostrais, todos localizados na mesma área de estudo. Os resultados indicaram que a banda azul e o índice GNDVI apresentaram as maiores correlações com a produtividade. O algoritmo de redes neurais (NN) obteve o melhor desempenho, com RMSE de 23% e MAPE de 20%.

Já nos trabalhos conduzidos por Faria et al. (2024), foi utilizado um banco de dados contendo 222 linhas de informações coletadas ao longo de duas safras, contemplando variáveis de fertilidade do solo e nutrição foliar. Os algoritmos com melhor desempenho foram o AdaBoost (RMSE de 8,02 sc. ha<sup>-1</sup>), seguido pelo Random Forest (RMSE de 8,77 sc. ha<sup>-1</sup>).

Em outra abordagem, Kouadio et al. (2018) utilizaram um banco de dados com 1.000 linhas contendo informações de fertilidade do solo para desenvolver modelos preditivos de produtividade do café. Foram testados os algoritmos Extreme Learning Machine (ELM), Regressão Linear Múltipla (MLR) e Random Forest (RF). As variáveis matéria orgânica do solo (MOS), potássio e enxofre apresentaram melhor desempenho preditivo com o ELM, que alcançou RMSE de 496,35 kg ha<sup>-1</sup>.

No estudo de Barbosa et al. (2021), o objetivo foi construir modelos de previsão de produtividade utilizando um conjunto de 114 linhas de dados, contendo informações sobre o índice de área foliar, diâmetro da copa e valores das bandas RGB captados por sensor acoplado a uma ARP. O algoritmo NEAT apresentou o melhor desempenho, com MAPE de 31,75%.

Os estudos analisados indicam que todos os autores utilizaram bancos de dados significativamente maiores do que o empregado nesta pesquisa, o que contribuiu para o desempenho dos modelos desenvolvidos. No entanto, não houve um padrão único de melhor algoritmo, pois cada estudo obteve melhores resultados com técnicas distintas. Os modelos foram construídos com diferentes arranjos de variáveis, algumas com maior relevância na previsão da produtividade, enquanto outras tiveram menor impacto.

A previsão da produtividade dos cafeeiros continua sendo um grande desafio. Os estudos mostram que ainda não existe um modelo fixo, com um conjunto de variáveis pré-definidas para essa finalidade, devido à complexidade dos fatores que influenciam a produção. Elementos como localização, altitude, tipo de solo, variações climáticas, manejo agrícola, estresses ambientais e características das cultivares afetam diretamente o rendimento das

lavouras. Essa variabilidade ressalta a necessidade de abordagens flexíveis e personalizadas, adaptadas às condições específicas de cada região e sistema produtivo.

Nos estudos de Abreu Júnior et al. (2022) os autores afirmam que o café é uma cultura perene que necessita de dois anos para completar seu ciclo fenológico, diferentemente da maioria das culturas, cujo ciclo reprodutivo se conclui em um ano. A alternância entre anos de alta e baixa produtividade como foi observado neste estudo, caracteriza o ciclo bianual do café, o que representa um desafio adicional para a previsão da produtividade utilizando algoritmos de machine learning. Essa oscilação pode introduzir padrões sazonais complexos nos dados, impactando a estabilidade e a capacidade dos modelos preditivos de generalizar corretamente entre diferentes safras.

Considerar o ciclo fenológico do café na construção dos modelos é fundamental para reduzir erros preditivos e melhorar a acurácia das estimativas. Estratégias como a inclusão de variáveis temporais, séries temporais multianuais e técnicas de aprendizado de padrões não lineares podem ser empregadas para capturar essa variabilidade, garantindo previsões mais robustas e condizentes com a realidade produtiva da cultura.

O fenômeno da bienalidade tem sido apontado como um fator relevante a ser considerado em modelos espectrais para estimar a produtividade do café (Chemura, Mutanga, Dube, 2017). O desenvolvimento de uma ferramenta eficiente para analisar esse padrão e representá-lo no espaço poderia aprimorar significativamente a modelagem da produtividade cafeeira.

A inteligência artificial (IA) ainda é uma ferramenta em fase de testes quando se trata da cafeicultura, e por isso, ajustes e aprimoramentos são necessários para que os modelos se tornem cada vez mais precisos e eficientes. Mesmo assim, a IA já se mostra uma ferramenta poderosa, com grande potencial para trazer muitos benefícios aos produtores. Com sua capacidade de analisar grandes volumes de dados e identificar padrões, ela pode ajudar na previsão de produtividade, no monitoramento da lavoura e na otimização do uso de insumos. Isso permite uma tomada de decisão mais segura, reduzindo desperdícios e aumentando a eficiência da produção. No futuro, com mais pesquisas e melhorias, a inteligência artificial pode se tornar uma aliada indispensável na cafeicultura, contribuindo para uma produção mais sustentável e rentável.

#### 4. CONCLUSÃO

Este estudo mostrou que algoritmos de aprendizado de máquina podem ser ferramentas poderosas para prever a produtividade do café, desde que as variáveis relevantes sejam corretamente escolhidas. As análises de correlação indicaram que fatores como a umidade do solo, os elementos químicos nas folhas (fósforo, enxofre, nitrogênio e potássio) e os índices NDRE e NDVI tiveram maior correlação com a produtividade. O desempenho dos modelos variou conforme o cenário: o KNN foi o mais eficaz nas safras individuais, enquanto o Gradient Boosting (GB), com variáveis selecionadas e aplicação da técnica SMOTE, apresentou maior estabilidade.

Conclui-se que o uso de algoritmos de aprendizado de máquina é uma abordagem promissora para otimizar a gestão agrícola, proporcionando previsões mais precisas e antecipadas. Isso favorece decisões mais informadas sobre práticas de cultivo e colheita, além de contribuir para a sustentabilidade e a eficiência da produção cafeeira. Ressalta-se que, com o aumento e o aprimoramento contínuo do banco de dados, espera-se uma melhoria progressiva no desempenho dos modelos, ampliando sua capacidade de generalização e sua aplicabilidade em diferentes contextos produtivos.

## REFERÊNCIAS

- ABREU JÚNIOR, C. A. M. et al. Estimating coffee plant yield based on multispectral images and machine learning models. **Agronomy**, v. 12, n. 12, p. 3195, 2022.
- AONO, A. H. et al. Machine learning for crop science: applications and perspectives in maize breeding. **Revista Brasileira de Milho e Sorgo**, v. 21, 2022.
- ASSAD, E. D. et al. Sistema de previsão da safra de soja para o Brasil. **Pesquisa Agropecuária Brasileira**, v. 42, p. 615-625, 2007.
- BARBOSA, B. D. S. et al. UAV-based coffee yield prediction utilizing feature selection and deep learning. **Smart Agricultural Technology**, v. 1, 100010, 2021.
- BARNES, E. M. et al. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In: **Proceedings of the 5th International Conference on Precision Agriculture**, 2000.
- BORRELLA, I.; MATAIX, C.; CARRASCO-GALLEGO, R. Smallholder farmers in the speciality coffee industry: opportunities, constraints and the businesses that are making it possible. **IDS Bulletin**, v. 46, n. 3, p. 29-44, 2015.
- BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.
- CARIELO, M. S.; PRESTES, J. A. L.; MARINHO, W. A. T. Modelagem preditiva com aprendizagem de máquina para a produção de café dos municípios de Minas Gerais. Apresentado no **L Congresso Brasileiro de Engenharia Agrícola - CONBEA 2021**, 08 a 10 de novembro de 2021.
- CARVALHO, L.G.; SEDIYAMA, G.C.; CECOM, P.R.; ALVES, H.M.R. A regression model to predict coffee productivity in Southern Minas Gerais, Brazil. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v.8, n.2/3, p.204-211, 2004.
- CHEMURA, A.; MUTANGA, O.; DUBE, T. Remote sensing leaf water stress in coffee (*Coffea arabica*) using secondary effects of water absorption and random forests. **Physics and Chemistry of the Earth, Parts A/B/C**, v. 100, p. 317-324, 2017.
- DAMATTA, F.M.; RONCHI, C.P.; MAESTRI, M.; BARROS, R.S. Ecophysiology of coffee growth and production. **Brazilian Journal of Plant Physiology**, v.19, p.485-510, 2007.
- ELAVARASAN, D. et al. Forecasting yield by integrating agrarian factors and machine learning models: A survey. **Computers and electronics in agriculture**, v. 155, p. 257-282, 2018.
- FARIA, R. et al. Models for predicting coffee yield from chemical characteristics of soil and leaves using machine learning. **Journal of the Science of Food and Agriculture**, v. 104, n. 9, p. 5197-5206, 2024.
- FAVARIN, J. L., et al. Equations for estimating the coffee leaf area index. **Pesquisa Agropecuária Brasileira**, v.37, 769-773, 2002.

FIX, E.; HODGES, J. L. Discriminatory analysis, nonparametric discrimination. 1951.

FREITAS, C. H. de et al. Smart Coffee: Machine Learning Techniques for Estimating Arabica Coffee Yield. **AgriEngineering**, v. 6, p. 4925–4942, 2024.

HAYKIN, S. Redes neurais: princípios e prática. 2. ed. Porto Alegre: Bookman, 2001.

HOLZMAN, M.E., et al. Early assessment of crop yield from remotely sensed water stress and solar radiation data. **ISPRS journal of photogrammetry and remote sensing**, v. 145, p. 297-308, 2018.

JIANG, Z. et al. Development of a two-band enhanced vegetation index without a blue band. **Remote Sensing of Environment**, v. 112, n. 10, p. 3833-3845, 2008.

KITTICHOTSATSAWAT, Y.; TIPPAYAWONG, N.; TIPPAYAWONG, K. Y. Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques. **Scientific Reports**, v. 12, n. 1, p. 14488, 2022.

KOUADIO, L. et al. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. **Computers and Electronics in Agriculture**, v. 155, p. 324-338, 2018.

KOUADIO, L. et al. Probabilistic yield forecasting of robusta coffee at the farm scale using agroclimatic and remote sensing derived indices. **Agricultural and Forest Meteorology**, v. 306, 108449, 2021.

LORENÇONE, J. A.; DE OLIVEIRA APARECIDO, L. E.; LORENÇONE, P. A. Previsão da produtividade do café com base em dados agroclimáticos e aprendizagem de máquina. **International Journal of Environmental Resilience Research and Science**, v. 3, n. 1.

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.

MIRANDA, J. M.; REINATO, R. A. O.; SILVA, A. B. da. Modelo matemático para previsão da produtividade do cafeeiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 18, p. 353-361, 2014.

QUINLAN, J.R. Induction of decision trees. **Machine learning**, v. 1, p. 81-106, 1986.

RODRÍGUEZ, J. P. et al. A Non-Destructive Time Series Model for the Estimation of Cherry Coffee Production. **Computers, Materials & Continua**, v. 70, n. 3, p. 4725–4743, 2022.

ROUSE, J.W., et al. Monitoring the vernal advancement of retrogradation of natural vegetation. **Greenbelt: National Aerospace Spatial Administration**, 1973. 371p.

SESSO, P. P.; SESSO FILHO, U. A.; PEREIRA, L. F. P. Dimensionamento do agronegócio do café no Brasil. **Cadernos de Ciência & Tecnologia, Brasília**, v. 38, n. 2, p. 26901, 2021.

SILVA, A. D. et al. Produtividade, rendimento de grãos e comportamento hídrico foliar em função da época, parcelamento e do método de adubação do cafeeiro Catuaí. **Engenharia Agrícola**, v. 23, n. 3, p. 434-440, 2003.

SILVA, C. A.; TEODORO, R. E. F.; MELO, B. Productivity and yield of coffee plant under irrigation levels. **Pesquisa Agropecuária Brasileira**, v. 43, p. 387-394, 2008.

SILVA, P. A. D. A. et al. Estimativa da produtividade do café por imagens Landsat-8 considerando os efeitos de sombreamento da orientação da linha de plantio no pivô central. **Sensoriamento Remoto Aplicado à Sociedade e Meio Ambiente**, v. 24, 100613, 2021.

SOARES, L. dos S. et al. Export behavior of the Brazilian coffee agribusiness and interactions with production elements. *Research, Society and Development*, v. 10, n. 3, p. e39210313503, 2021.

SCHOLANDER P. F., et al. Sap pressure in vascular plants. **Science**, v.148, 339, 1965.

TAKANO, A. L. R.; CABRERA, L. C.; CALDARELLI, C. E. Cadeia produtiva e mercado cafeeiro no Brasil: desafios e potencialidades. *Revista Economia Ensaios*, v. 36, n. 1, p. 128-145, 2021.

WHETTON, R. et al. Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI. **Computers and Electronics in Agriculture**, v. 138, p. 127-136, 2017.

### **TERCEIRA PARTE – CONSIDERAÇÕES FINAIS**

Este trabalho trouxe uma visão integrada e prática sobre como as tecnologias da Agricultura de Precisão (AP) e Agricultura Digital (AD) podem ser aplicadas na gestão da cafeicultura. Ao longo dos três artigos que compõem esta tese, foi possível explorar diferentes etapas do manejo da lavoura, sempre com o objetivo de entender melhor a produção e buscar formas mais eficientes e sustentáveis de conduzir o cultivo.

No primeiro artigo, observou-se que a produtividade do café pode variar, mesmo em uma área pequena. Ao analisar três safras consecutivas, usando métodos geoestatísticos, foi possível observar fatores como a bienalidade e o manejo convencional uniforme, influenciam diretamente na produtividade das plantas. Esses dados mostraram a importância de conhecer bem a lavoura para tomar decisões mais acertadas.

Já no segundo artigo, o foco foi nos índices de vegetação, como o NDVI e na clorofila, coletados por diferentes tipos de sensores. A boa correlação entre os dados dos sensores mostrou que essas ferramentas funcionam bem para monitorar o estado das plantas. O uso de Aeronave Remotamente Pilotada se destacou por permitir a coleta de informações rápidas, precisas e em alta resolução, o que é um grande avanço para quem quer acompanhar de perto a saúde da lavoura.

Por fim, no terceiro artigo, a proposta foi construir modelos preditivos de produtividade usando algoritmos de aprendizado de máquina. Com base em dados agronômicos, ambientais e imagens, foram testados diferentes algoritmos para prever o rendimento da lavoura. Mesmo com algumas limitações como número reduzido de amostras do banco de dados, disponibilidade limitada de amostras temporais e heterogeneidade espacial dos dados, os resultados foram promissores e mostram que a tecnologia pode, sim, ajudar o produtor a se planejar melhor e tomar decisões mais estratégicas.

De forma geral, os três capítulos se complementam e mostram que a integração entre dados de campo, sensoriamento remoto e inteligência artificial pode transformar a maneira como o café é cultivado. O estudo reforça que investir em tecnologias digitais não é apenas uma tendência, mas uma necessidade para quem busca mais eficiência, produtividade e sustentabilidade na cafeicultura. Com isso, esta tese contribui não só para a pesquisa acadêmica, mas também para práticas mais modernas e conectadas no campo.