

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GEOGRAFIA
CURSO DE ENGENHARIA DE AGRIMENSURA E CARTOGRÁFICA

ALICE PEDRO BOM PAES

USO DE SÉRIES TEMPORAIS ESPECTRAIS PARA ESTIMAR A PRODUTIVIDADE
DO CAFÉ

MONTE CARMELO
2022-1

ALICE PEDRO BOM PAES

USO DE SÉRIES TEMPORAIS ESPECTRAIS PARA ESTIMAR A PRODUTIVIDADE
DO CAFÉ

Projeto de pesquisa apresentado como requisito para aprovação na disciplina de Trabalho de Conclusão de Curso 2 no curso de Engenharia de Agrimensura e Cartográfica da Universidade Federal de Uberlândia campus Monte Carmelo.

Orientador: Prof. Dr. George Deroco Martins.

MONTE CARMELO
2022-1

AGRADECIMENTOS

Agradeço acima de tudo aos meus familiares por todo apoio, em especial a minha mãe, que sempre esteve muito presente mesmo à distância, e sempre fez de tudo para acalmar meu coração. A minha irmã, Júlia, por sempre me incentivar e apoiar em tudo. Ao meu pai, por todo auxílio e apoio.

Ao meu professor orientador de TCC e de outros três projetos de iniciação científica, Dr. George Deroco Martins, pela confiança para a execução deste trabalho, por todo conhecimento transmitido e por tudo vivido até então neste período de graduação.

À Laura Moura Xavier por toda atenção, orientação e tempo disponibilizado para a realização deste trabalho.

Aos meus colegas de turma, em especial às amigas Isadora, Letícia e Mariana por estarem sempre comigo.

Ao meu namorado Rodrigo, por acreditar no meu potencial e por todo companheirismo.

E por fim, a Universidade Federal de Uberlândia pela oportunidade de estudo.

Somos o que pensamos. Tudo o que somos surge com nossos pensamentos. Com nossos pensamentos, fazemos o nosso mundo.

Buda

RESUMO

Aplicações de técnicas de Sensoriamento Remoto voltados à Agricultura de Precisão (AP) vêm se tornando cada vez mais crescente, devido às necessidades de demanda oriunda do crescimento populacional, a crescente competitividade no mercado e também, a complexidade da agricultura. Com isso, o conhecimento sobre a lavoura tornou-se ainda mais necessário, onde o uso da AP aliada às imagens multiespectrais, torna-se possível a elaboração de mapas de estimativa de produtividade de uma região, fazendo com que os agricultores possam identificar as áreas da lavoura sob estresse e facilitando a aplicação de insumos em locais corretos e com suas respectivas quantidades necessárias. Até então, diversos trabalhos e estudos vêm fazendo uso de imagens multiespectrais coletadas referente à data de colheita da lavoura, e este trabalho tem como objetivo a quebra (ou não) deste paradigma. Sendo assim, esta pesquisa teve como propósito a determinação do melhor período para se estimar a produtividade, dentro do período fenológico do café, além da definição do melhor algoritmo classificador. Para isto, o estudo foi realizado em uma parcela situada no município de Monte Carmelo, no estado de Minas Gerais (MG), e teve como metodologia a extração da reflectância de alguns pontos do talhão cuja produtividade foi mensurada em campo. Com isto, concluiu-se que março de 18 foi o mês mais assertivo do período fenológico cafeeiro perante os demais meses. Este mês corresponde à segunda fase da fenologia do café, referente ao último mês de seu período vegetativo. Além disso, concluiu-se que o algoritmo que obteve melhores resultados foi o *Support Vector Machine*, por possuir maiores correlações, menores erros e valores menos tendenciosos comparados aos resultados obtidos com os outros algoritmos.

Palavras chaves: Cafeicultura, predição da produtividade, modelos espectrais, mapas de distribuição da produtividade.

LISTA DE FIGURAS

- Figura 1** Esquematização das fases fenológicas do cafeeiro
- Figura 2** Curva espectral de uma folha verde sadia
- Figura 3** Mapa de Localização da Área de Estudo
- Figura 4** Fluxograma da Metodologia do Trabalho
- Figura 5** Aquisição de dados do cafeeiro
- Figura 6** Imagens do Satélite *Planet*
- Figura 7** Configuração do arquivo *weka*
- Figura 8** Comportamento Espectral do cafeeiro no espectro do visível
- Figura 9** Comportamento Espectral do cafeeiro na banda do IVP
- Figura 10** Comportamento Espectral do cafeeiro referente ao índice NDVI
- Figura 11** Correlação de cada banda de satélite com a produtividade
- Figura 12** Comportamento da Correlação com os algoritmos testados
- Figura 13** Comportamento do RMSE com os algoritmos testados
- Figura 14** Comportamento do MAPE com os algoritmos testados
- Figura 15** Representação da estimativa da produtividade cafeeira

LISTA DE TABELAS

- Tabela 1** Resolução espectral satélite *Planet*
- Tabela 2** Correlação de cada banda de satélite e NDVI com a produtividade mensurada em campo

LISTA DE EQUAÇÕES

- Equação 1** Equação do Índice de Vegetação da Diferença Normalizada (NDVI)
- Equação 2** RMSE (%)
- Equação 3** MAPE

LISTA DE SIGLAS E ABREVIATURAS

AM	Aprendizado de Máquina
AP	Agricultura de Precisão
ARP	Aeronave Remotamente Pilotada
ART	Anotação de Responsabilidade Técnica
BA	Bahia
IA	Inteligência Artificial
IAF	Índice de Área Foliar
IBGE	Instituto Brasileiro de Geografia e Estatística
IV	Infravermelho
IVP	Infravermelho Próximo
MG	Minas Gerais
MLP	<i>Multilayer Perceptron</i>
NDVI	<i>Normalized Difference Vegetation Index</i>
NIR	<i>Near Infrared Spectroscopy</i>
PPP	Posicionamento por Ponto Preciso
PR	Paraná
REM	Radiação Eletromagnética
RGB	<i>Red, Green, Blue</i>
RMSE	<i>Root Mean Squared Error</i>
RNA	Rede Neural Artificial
RTK	<i>Real Time Kinematic</i>
SIG	Sistema de Informação Geográfica
SVM	<i>Support Vector Machine</i>

SUMÁRIO

1 INTRODUÇÃO.....	09
2 OBJETIVOS.....	11
3 JUSTIFICATIVA.....	12
4 FUNDAMENTAÇÃO TEÓRICA.....	13
4.1 A Fenologia do Café.....	13
4.1.1 Comportamento Espectral do Café.....	14
4.2 Conceitos relacionados ao SR e a AP.....	15
4.2.1 Uso de imagens multiespectrais para estimativa da produtividade	17
4.3 Métodos de aprendizado de máquina para a estimativa de produtividade.....	18
5 MATERIAL E MÉTODOS.....	20
5.1 Área de estudo.....	20
5.2 Material.....	21
5.3 Método.....	21
5.3.1 Etapa 1: Aquisição de Dados.....	22
5.3.2 Etapa 2: Organização dos Dados.....	26
5.3.2.1 Extração dos Valores de Reflectância das Bandas de Satélite..	26
5.3.2.2 Cálculo de Índice de Vegetação por Diferença Normalizada..	26
5.3.2.3 Arquivo para o <i>weka</i>	27
5.3.3 Etapa 3: Processamento computacionais.....	28
5.3.3.1 Geração do Modelo.....	28
5.3.4 Etapa 4: Análise dos resultados.....	31
6 RESULTADOS E DISCUSSÕES.....	32
7 CONCLUSÃO.....	44
REFERÊNCIAS.....	45

1 INTRODUÇÃO

O estudo e desenvolvimento de novas técnicas de Sensoriamento Remoto voltados à Agricultura de Precisão (AP), para uma melhor gestão e manejo agrícola, vêm se tornando bastante crescente, devido à complexidade da agricultura, ao crescimento da competitividade no mercado, e ao aumento das necessidades de demanda, oriunda do crescimento populacional e o aumento do consumo per capita (VIEIRA FILHO, 2017).

Segundo Roza (2000), a AP leva em consideração que cada parcela da lavoura possui propriedades distintas, e, com o conhecimento de cada uma dessas particularidades é possível identificar quais áreas da cultura estão sob estresse (CALERA et al., 2017).

Sabendo das particularidades de cada parcela, torna-se possível aplicar insumos nos locais corretos, nos momentos adequados e com quantidades requeridas (TSCHIEDEL *et al.*, 2002), solucionando ou minimizando problemas de controle de doenças e pragas (BESTELMEYER *et al.*, 2020), aumentando então sua eficiência na produção, com uma redução de custos (ARTUZO *et al.*, 2017) e tempo de trabalho.

O Brasil vem atuando em uma posição importante no comércio mundial do agronegócio nos últimos anos, e sua participação no abastecimento de grãos deve aumentar ainda mais nos próximos anos (EMBRAPA). Sendo assim, as imagens de satélite vinculadas às técnicas de AP vêm desempenhando um papel fundamental, fornecendo um grande suporte para o mapeamento e monitoramento da dinâmica da vegetação em grande escala, devido à resolução temporal e espacial oferecida pelos sensores dos satélites (TEIXEIRA, 2005).

De acordo com RAMIREZ e JÚNIOR (2010), o Brasil pode ser considerado um dos maiores usuários de imagens e dados remotos do mundo, e por isso pesquisadores vêm desenvolvendo e aprimorando técnicas para que os agricultores possam ter uma melhor tomada de decisão e um controle maior sobre as áreas a serem manejadas, a fim de alcançarem uma maior rentabilidade (TEIXEIRA, 2005).

Como consequência da necessidade do aumento de produção, detecção de pragas e doenças nas lavouras de café, a fim de manejá-las de maneira adequada, requer-se encontrar o período ideal para se estimar a produtividade, com intuito de antecipá-la. No atual cenário científico, trabalhos desenvolvidos abordam o uso de imagens multiespectrais tomadas na mesma data da coleta de dados da produtividade em campo, cujos pontos do talhão possuem produtividade desconhecida. Em suma, a estimativa é feita em um momento tardio, impossibilitando quaisquer intervenções agrícolas para melhoramentos no cultivo, que poderia acarretar em possíveis melhorias na produtividade cafeeira.

Sendo assim, sabendo que o uso de imagens multiespectrais vinculadas à estimativa de produtividade de culturas perenes em geral vem trazendo bons resultados, é questionado se culturas em estádios fenológicos anteriores ao da colheita já apresentam indícios fisiológicos sensíveis às bandas de satélite, proporcionando um modelo de produtividade precoce, e possibilitando a previsão com antecedência.

2 OBJETIVOS

Levando em consideração que até então, os modelos espectrais de estimativa da produtividade do cafeeiro limitam-se pela utilização de imagens tomadas no dia da colheita, esta pesquisa tem como objetivo principal a quebra deste paradigma, e determinar qual o melhor período fenológico para aquisição de dados multiespectrais para predição da produção da cultura cafeeira. Assim sendo, os objetivos específicos são:

- Verificar o melhor período para se estimar produtividade e;
- Definir o melhor algoritmo estatístico, dentre os três testados, para se estimar a produtividade cafeeira.

3 JUSTIFICATIVA

A produtividade é definida como sendo a quantidade de produção por unidade de área, que está relacionada diretamente com as quantidades de seus produtos e seus insumos utilizados durante a produção. Estes ganhos na eficiência da produção estão associados a fatores externos, como o desenvolvimento científico e tecnológico. Sendo assim, com a globalização e a expansão dos negócios, ocasionou-se a preocupação com a eficiência de produtividade da agricultura brasileira, que foi impulsionada em meados dos anos 80 (GASQUES E VILLA VERDE, 1990), com as novas políticas públicas para o setor agrícola (PEROBELLI, 2007).

A partir da década de 90, o ambiente do agronegócio tornou-se mais competitivo, e o sistema agroindustrial brasileiro passou a participar mais do comércio internacional, que foi ocasionado por algumas mudanças estruturais ocorridas na economia. O sistema agroindustrial passou a adotar novas formas de organização, através de novos mecanismos e técnicas de cultivo, visando uma maior obtenção de produtividade (possível através da melhor gerência dos recursos produtivos) (PEROBELLI, 2007).

O bom desempenho e o crescente desenvolvimento do setor agrícola estão diretamente atrelados à disponibilidade de terras, ao avanço na pesquisa tecnológica e as recentes técnicas de manejo de lavouras (PEROBELLI, 2007). Assim sendo, o surgimento e o desenvolvimento da AP está diretamente interligada às melhorias do setor agroindustrial brasileiro, pois com ela houve o desenvolvimento de equipamentos automatizados, disponibilização de satélites para a localização de pontos, surgimento de ferramentas de Sistemas de Informações Geográficas (SIG), sensores, monitores de colheita e outras tecnologias que podem auxiliar na identificação e em um melhor manejo das variabilidades agrícolas (PIRES, 2004).

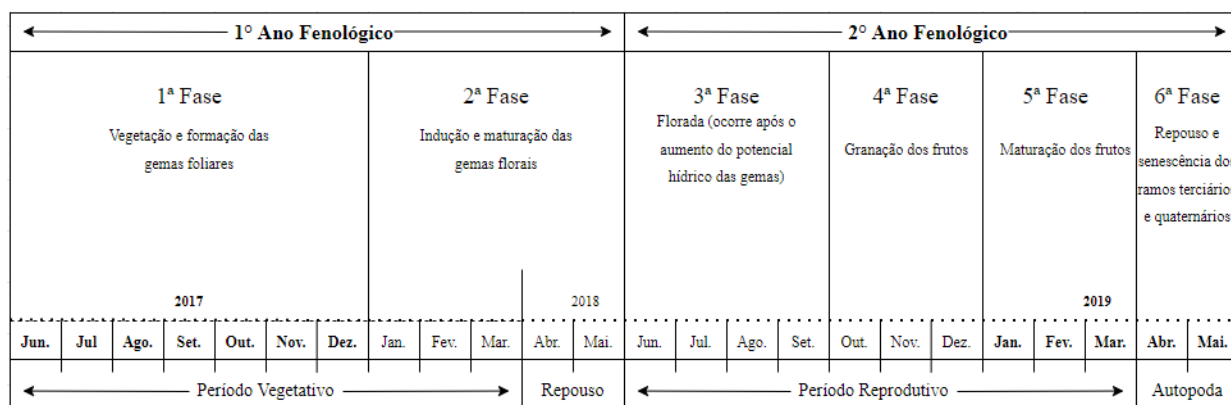
Levando em consideração que as pesquisas vêm fazendo uso de imagens multiespectrais coletadas na mesma data que a colheita em campo, como citado no artigo publicado por Sartori *et al.* na Revista Brasileira de Cartografia, em 2009, cujos autores fizeram um estudo baseado no mapeamento de plantas daninhas com a utilização de Redes Neurais Artificiais (RNA), este trabalho se mostra importante pelo questionamento deste padrão, a fim de se encontrar a melhor fase do período fenológico cafeeiro para se obter uma predição de produtividade mais assertiva possível.

4 FUNDAMENTAÇÃO TEÓRICA

4.1 A Fenologia do café

O cafeeiro leva dois anos para completar seu ciclo fenológico, que inclui desde a fase do período vegetativo até a autopoda (CAMARGO, 2001), subdividindo-se em seis fases, sendo elas duas vegetativas e quatro reprodutivas (DE OLIVEIRA, 2012), conforme representado na esquematização da Figura 1.

Figura 1- Esquematização das fases fenológicas do cafeeiro



Fonte: Adaptado de Cargo e Camargo, 2001

Esta esquematização é importante para fundamentar e facilitar as pesquisas e observações do plantio do café, possibilitando a identificação dos períodos mais críticos do cultivo, a fim de se aplicar de tratamentos específicos e até mesmo identificar épocas propensas ao estresse hídrico (CAMARGO, 2001).

Sendo assim, neste esquema, apresentado conforme a Figura 1, nota-se que a primeira fase fenológica perdura por 7 meses, cujo período consiste na vegetação e formação das gemas foliares. Este período consiste em um período de luminosidade solar acima de 13 horas (CAMARGO, 1985), fazendo com que sua formação dependa desta condição fotoperiódica (DE OLIVEIRA, 2012).

A segunda fase, ainda dentro do primeiro ano fenológico, é caracterizada pela indução, desenvolvimento, maturação e dormência (cujas etapas são dadas de forma consecutiva) das gemas florais (CAMARGO e FRANCO, 1985). Ao final, ao entrarem em dormência, tornam-se aptas a se transformarem em botões florais, e florescerem após choque hídrico existente da próxima fase (GOUVEIA, 1984).

Então, a terceira fase, que consiste no segundo ano do período fenológico cafeeiro, é caracterizada pela florada e expansão dos frutos, na qual os botões florais florescem de

maneira rápida, após um choque hídrico, seja ela causada por chuva ou irrigação (DE OLIVEIRA, 2012).

Em seguida, a quarta fase consiste na granação dos frutos, quando há formação dos grãos através da solidificação dos líquidos internos (CAMARGO, 2001). Em caso de estresse hídrico, esta fase é responsável pelo desenvolvimento de frutos mal granados.

Na quinta fase ocorre a maturação dos frutos, fruto da precocidade da cultivar e da acumulação da energia solar (DE OLIVEIRA, 2012). Nesta etapa, a qualidade do produto está atrelada às deficiências hídricas moderadas (CAMARGO, 2001).

Por último, a sexta fase constitui a senescência do cafeeiro, período no qual os ramos produtivos não primários secam e morrem, ocasionando a autopoda (CAMARGO, 2001).

4.1.1 Comportamento espectral do café

Em decorrência às fases fenológicas do café, esquematizado conforme Cargo e Camargo (2001), o comportamento espectral do dossel sofre variações em relação a sua reflectância. Segundo Martins (2016), a banda do vermelho e a do infravermelho próximo são responsáveis por melhor discriminar uma vegetação sadia e não sadia. Além disso, a ausência da clorofila faz com que haja uma maior reflectância, devido sua menor absorção de energia nas bandas do visível (CARTER, 1993).

Waisten Resende Carrijo *et al.* (2021) desenvolveram um trabalho que diz respeito à resposta espectral do índice NDVI em cada fase fenológica do café, na qual os autores fizeram uso de imagens do Sentinel 2A e 2B, e concluíram que a fase responsável por emitir maior e melhor resultado do NDVI foi a fase 3, período referente a etapa reprodutiva do café, cujo mesmo acontece na florada, época em que há um aumento na biomassa da planta. Por outro lado, os autores observaram que a sexta fase apresentou um pior resultado, devido a diminuição da biomassa, acarretado pela autopoda.

Moreira *et al.* (2004) fizeram um estudo da análise espectral e temporal da cultura cafeeira em imagens devidamente retificadas do satélite Landsat 5 e 7 (imagens melhoradas radiometricamente), no município de Aguanil, centro-sul de Minas Gerais (MG), com imagens referentes aos anos 1999 e 2001, e concluíram que a variação da reflectância das lavouras pode estar atrelada ao fato do ambiente estar mais seco, propiciando uma menor absorção e conseqüentemente, maior reflectância dos cafezais na banda 3, referente ao espectro do vermelho, fazendo com que épocas de seca sejam melhores meses para mapeamento das lavouras através de imagens multiespectrais. Além disso, concluíram que

com a análise espectral e temporal da reflectância de lavouras do cafeeiro, foi possível identificar se houve ou não a redução da fitomassa.

4.2 Conceitos relacionados ao Sensoriamento Remoto e a Agricultura de Precisão

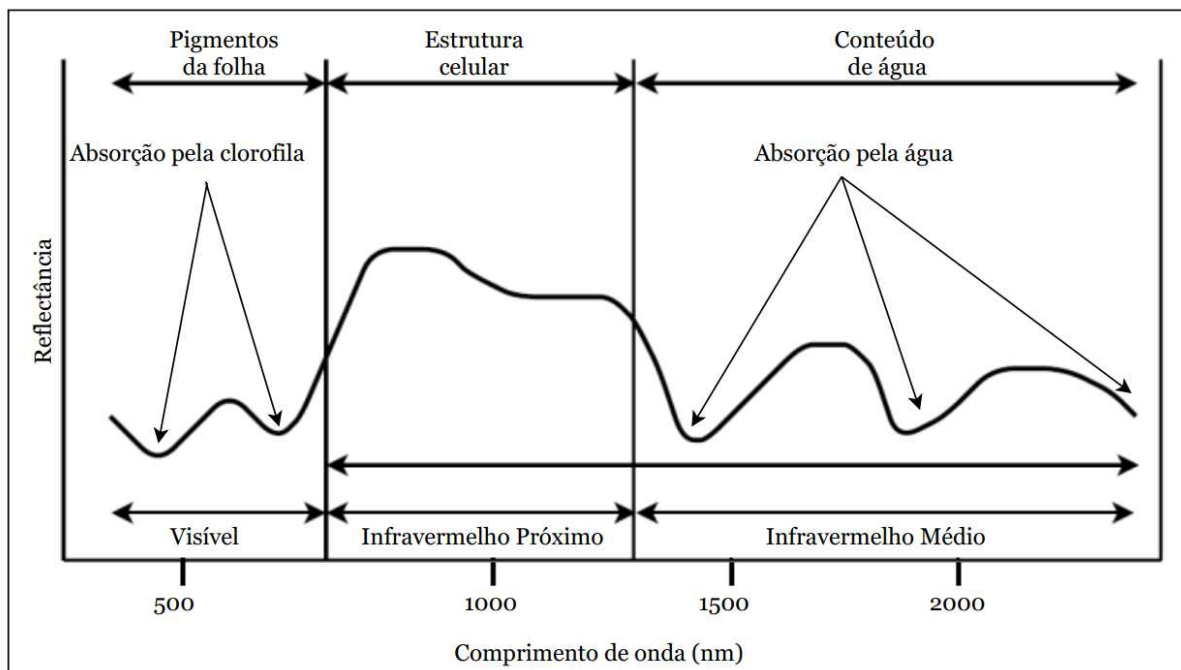
O termo “Sensoriamento Remoto” surgiu pela primeira vez na literatura em meados dos anos 60, e foi caracterizado por adquirir informações de forma remota, sem que houvesse contato físico com os objetos (NOVO, 2010). A obtenção de dados remotamente é realizada através da interação de sensores acoplados em satélites (plataformas orbitais) ou Aeronaves Remotamente Pilotadas (ARPs), que respondem aos estímulos físico/químico da reflectância oriunda das plantas e suas respectivas superfícies de fundo (FORMAGGIO, 2017; MOLIN, 2015), cuja informação é transferida por meio da Radiação Eletromagnética (REM) (NOVO, 2010).

Os dados de reflectância captados por estes sensores podem ser aliados no estudo do manejo de praga e água, por exemplo, pois cada faixa do espectro eletromagnético pode ser associados a diferentes características da cultura, como o Índice de Área Foliar (IAF) e a biomassa (ATZBERGER, 2013), e suas respostas espectrais podem servir para análise quanto ao estado nutricional e sanidade das culturas (FURTADO, 2020).

No final dos anos 90, quando o uso do sensoriamento remoto aplicado às melhorias de gestão em cultivos agrícolas ainda não era tão popular no Brasil, Johann *et al.*, que tiveram o artigo publicado na Revista Brasileira de Engenharia Agrícola e Ambiental em 2004, realizaram um estudo visando um manejo localizado em Cascavel (PR), através da análise da variabilidade espacial da produtividade das culturas de soja e trigo e suas respectivas propriedades físico-químicas.

A produtividade de uma cultura pode estar diretamente associada com a quantidade de radiação captada pelos sensores, oriunda da reflectância dos alvos, e por isso o conhecimento do comportamento espectral dos alvos é importante para a extração de informações quanto à sanidade de uma planta (NOVO, 2010). A Figura 2 apresenta a curva espectral de uma folha verde sadia, em cada faixa do espectro eletromagnético.

Figura 2 – Curva espectral de uma folha verde sadia



Fonte: adaptado de Novo (2010)

A curva espectral de uma vegetação fotossinteticamente ativa, representada conforme a Figura 2, pode ser decomposta em três diferentes regiões espectrais: i) Baixa reflectância da vegetação, para ondas com comprimentos até 700 nm (referente ao espectro do Visível); ii) Alta reflectância da vegetação, para ondas de comprimento inseridas entre o intervalo de 700 nm e 1300 nm (referente ao espectro do Infravermelho Próximo) e iii) Reflectância que depende do conteúdo de águas das folhas (referente ao espectro do Infravermelho Médio). Esta região é composta por dois pontos máximos de absorção pela água (em 1400 nm e 1950 nm) (NOVO, 2010).

A AP se desenvolveu pela necessidade da elaboração de estratégias para o gerenciamento da desuniformidade das lavouras (MOLIN, 2015), fazendo uso de um conjunto de técnicas para geração de mapas que podem auxiliar na gestão de produção e na tomada de decisão sobre a área a ser manejada, com o uso de sensores e tecnologias (EMBRAPA, 2017), que permitem detectar doenças, pragas ou outros distúrbios de cada cultura, tornando-se possível o ajuste na previsão de safras agrícolas (FORMAGGIO, 2017).

É importante ressaltar que a AP nem sempre visa a uniformização de cada parcela da lavoura, pois existem situações em que a baixa produtividade pode estar atrelada a fatores que estão fora do controle humano, como por exemplo a variabilidade da textura do solo (NOVO, 2010). Neste caso, estas regiões são tratadas de acordo com seu baixo potencial,

com um menor aporte de insumos. Por outro lado, regiões com maior potencial produtivo recebem um maior aporte de insumos, visando explorar seu limite econômico. Assim sendo, essa variabilidade natural presente nas lavouras faz com que haja a necessidade de geri-las da forma mais conveniente possível (MOLIN, 2015), otimizando a quantidade de corretivos, fertilizantes, herbicidas e água irrigada a serem aplicados, maximizando os lucros e reduzindo o uso inconsequente de insumos, protegendo assim, o meio ambiente (Johann *et al.*, 2004).

Algumas aplicações do uso da AP para gestão de lavouras podem ser observadas por meio de alguns trabalhos como MACHADO (2004), que desenvolveu mapas de produção cujos dados foram obtidos por meio de coletas de amostras em diferentes épocas do ano, para análise da fertilidade do solo em Carambeí (PR), ou então pelo trabalho de Bertolin, publicado na Revista Brasileira de Agricultura Irrigada em 2017, que detalhou seu experimento realizado no Oeste da Bahia para a predição da produtividade de milho irrigado com auxílio de uma imagem de satélite Landsat 8, coletada em uma data que mais se aproximava do máximo estágio vegetativo do milho, cujo resultado mostrou-se satisfatório.

4.2.1 Uso de imagens multiespectrais para estimativa da produtividade

O sensoriamento remoto aplicado no estudo e na análise de imagens multiespectrais pode ser um grande aliado para a identificação de áreas da lavoura sob estresse, principalmente estresse hídrico (Jackson *et al.*, 1981), distinguido através do comportamento espectral do objeto. A chamada “assinatura espectral” da planta pode sofrer modificações em decorrência as alterações naturais da vegetação, ou então modificações impostas pelo homem, através das práticas de irrigação e adubação por exemplo (NOVO, 2010), e por isso o Infravermelho Próximo (IVP) desempenha um papel fundamental para o conhecimento quanto a sanidade da lavoura.

Existem diversos estudos que abordam o uso de imagem de satélite para predição e previsão da produtividade (ou outras variáveis agrícolas) de diferentes culturas, como por exemplo, o trabalho publicado em 2011 por BRANDÃO *et al.* no Congresso Brasileiro do Algodão, cujos autores detalharam os procedimentos do experimento realizado em 2007 no município de Bom Jesus da Lapa (BA), em que o resultado obtido para a estimativa da produtividade do algodoeiro por meio de imagens de satélite Landsat 5 mostrou-se satisfatório em todas as imagens coletadas em suas diferentes fases fenológicas, com a obtenção IAF;

O trabalho de MARTINS *et al.* (2014) publicado na Revista Brasileira de Cartografia, teve como objetivo a investigação do potencial de imagens multiespectrais Rapideye, com a utilização de cinco bandas (azul, verde, vermelho, vermelho limítrofe e infravermelho próximo), para a detecção de áreas infestadas por nematoides e *migdolus fryanus* em cultura canavieira, cujo mesmo foi atingido com sucesso.

4.3 Métodos de aprendizado de máquina para a estimativa de produtividade

Muitas pesquisas vêm fazendo uso de imagens multiespectrais para controle de doenças e pragas ou para estimativa de produtividade de culturas perenes (café, laranja, limão, banana, etc), a fim de se obter um maior conhecimento sob a área de interesse e desenvolver técnicas de manejo especializado (FORMAGGIO; SANCHES, 2017). Rosa *et al.* (2010), por exemplo, fizeram um estudo com utilização do sensor MODIS, presente no satélite Terra e Aqua, abordando a estimativa da produtividade de café através de um modelo agrometeorológico-espectral, cujo resultado se mostrou satisfatório.

Além disso, diversos outros algoritmos de Aprendizado de Máquina (AM) vêm sendo utilizados, como por exemplo o algoritmo *Support Vector Machine* (SVM), que possui um aprendizado supervisionado, cujo mesmo necessita de dados de entrada para que seja possível a extração de conhecimentos, a fim de representar o que o usuário deseja (HAYKIN, 1999), de modo mais correto possível. As SVMs são utilizadas para tarefas de reconhecimento de padrões e têm como vantagem a eficácia em espaços de alta dimensão e em termos de memória. Ela é bastante versátil, sendo possível variar as funções Kernel, por exemplo.

Outro algoritmo classificador que vem sendo bastante utilizado nas pesquisas são as Redes *Multilayer Perceptrons* (MLP). O *perceptron* é um tipo de Rede Neural Artificial (RNA) de algoritmo simples que é construído em torno de um neurônio não-linear, cujo objetivo é classificar corretamente um conjunto de estímulos aplicados.

Os *Perceptrons* de Múltiplas Camadas representam uma generalização do *perceptron* de camada única. Esta rede consiste em um conjunto de unidades sensoriais (nós de fonte) que compõem: uma camada de entrada (encarregada de receber o sinal); uma ou mais camadas ocultas de nós computacionais e uma camada de saída (encarregada de fazer previsões a partir dos dados de entrada). Sendo assim, o sinal de entrada se propaga para frente através da rede, camada por camada (HAYKIN, 2000). Este método de MLP vem sendo muito utilizado para a resolução de problemas complexos, através de um treinamento

feito com um algoritmo denominado retropropagação de erro, cujo mesmo é baseado em uma regra de aprendizagem por correção de erro.

Árvore de decisão é outro método de aprendizado de máquina bastante utilizado e conhecido pela sua simplicidade e interpretabilidade, onde cada nó de decisão contém um teste para algum atributo; cada ramo descendente corresponde a um possível valor do mesmo; cada folha têm associação a uma classe e cada percurso corresponde a uma regra de classificação (Gama, 2004).

Já as Florestas Aleatórias (*Random Forest*), que é um algoritmo composto por diversos classificadores do tipo “árvore de decisões” (Breiman, 2001), faz com que seja oferecido uma maior gama de predições. Além disso, é considerado bastante flexível e de fácil uso, trazendo ótimos resultados sem que haja a necessidade dos ajustes de seus hiperparâmetros. Assim sendo, é uma boa opção por poder ser utilizada em tarefas de classificação e regressão, tornando-se uma boa escolha de algoritmo para predição de classes.

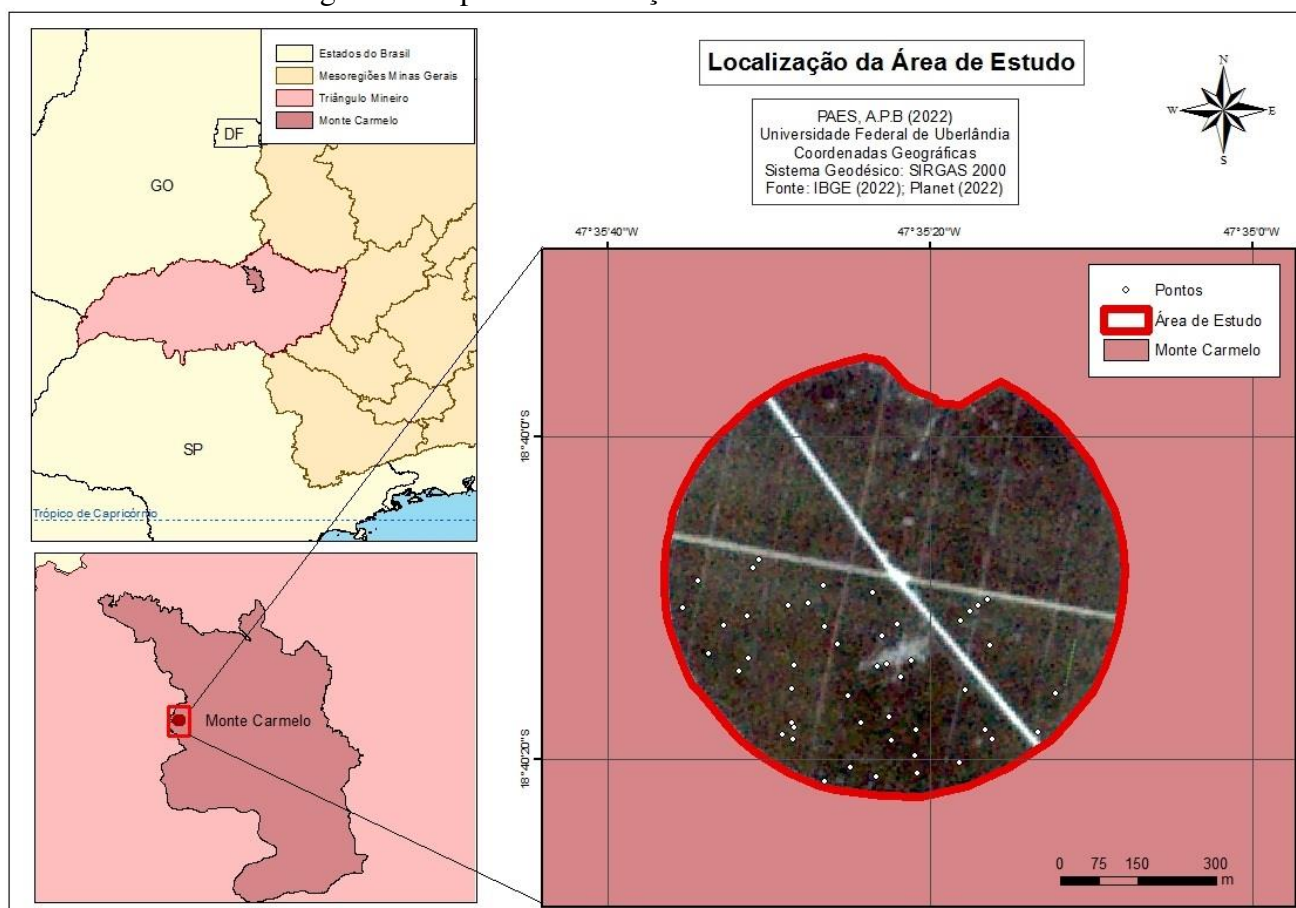
5 MATERIAIS E MÉTODOS

5.1 Área de estudo

O estudo foi realizado em uma parcela situada no município de Monte Carmelo, no estado de Minas Gerais (MG), região sudeste do país, e faz parte da microrregião de Uberlândia, e mesorregião do Triângulo Mineiro. O município se localiza nas coordenadas geográficas: Latitude: $18^{\circ} 44' 5''$ S, Longitude: $47^{\circ} 29' 47''$ O, e está situado a uma altitude média de 890 metros. O município possui clima tropical, com invernos secos e verões quentes, de acordo com a classificação de Köppen, e é formada por um relevo constituído por planaltos, morros e montes. A parcela de estudo é denominada “Terra Rica” e é caracterizada por ser um pivô central de irrigação, possuindo uma área total de aproximadamente 50 hectares.

O estudo consiste em 48 pontos distribuídos de forma aleatória em uma parcela desta área de estudo, com valores de produtividade mensurados. O mapa de localização juntamente com estes pontos se encontra na Figura 3.

Figura 3: Mapa de Localização da área de estudo



Elaboração: A Autora (2022).

5.2 Material

Para este trabalho foi necessário uma planilha eletrônica contendo de dados de produtividade cafeeira com suas respectivas coordenadas, além de *softwares* para organização, manipulação e processamento dos dados.

Sendo assim, o material necessário para que se tornassem possíveis a predição de produtividade foi:

- Planilha eletrônica contendo informações de produtividades da cultura cafeeira;
- Bandas de imagem de satélite que compoem a área de interesse (imagens multiespectrais orbitais) e;
- *Shapefile* da delimitação da parcela de estudo.

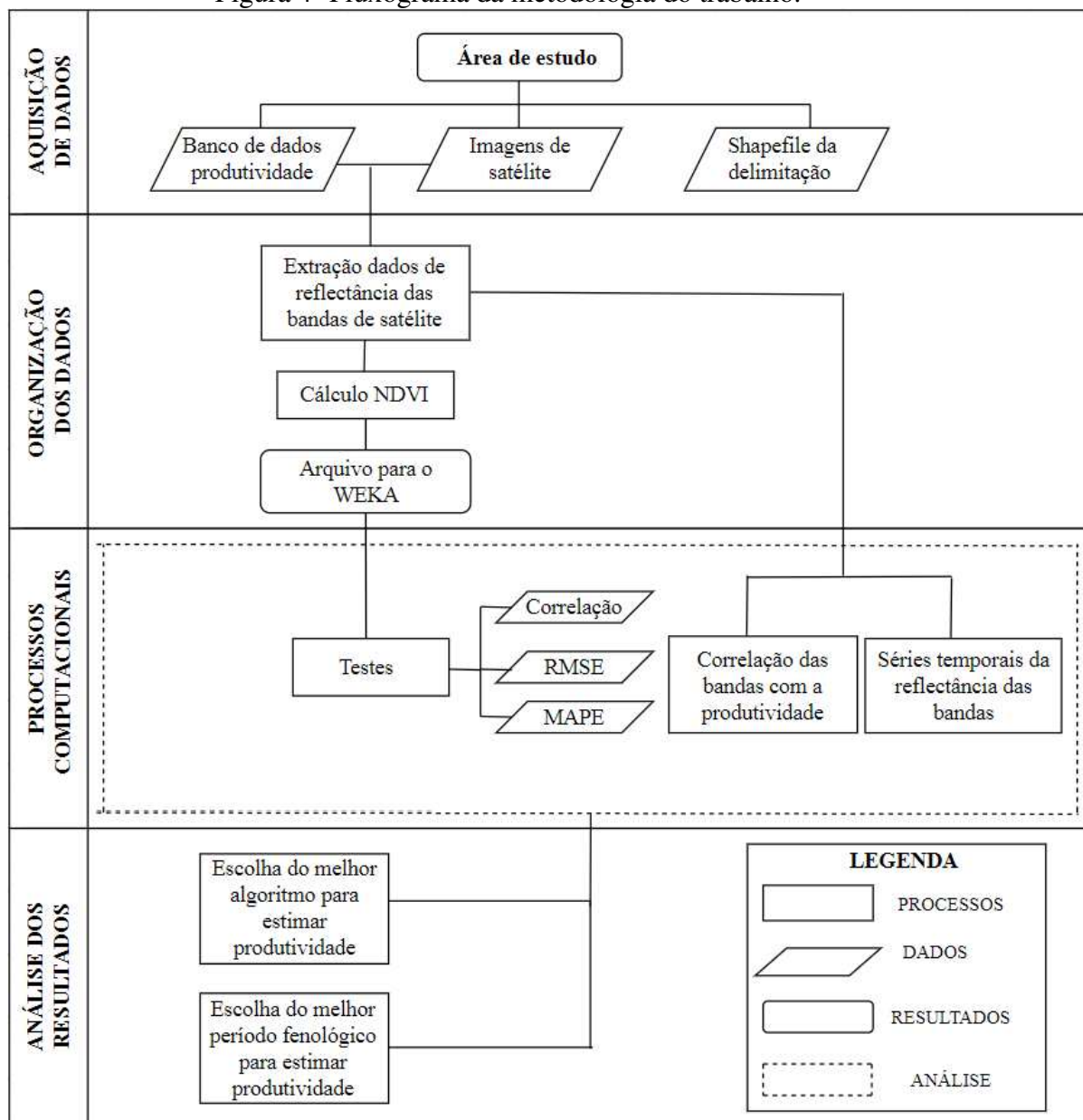
Os *softwares* utilizados durante todas as etapas de manipulação de dados até a predição de produtividade de fato, foram:

- *Excel*, utilizado para organização dos dados;
- *Envi Classic* versão 5.1, para extração dos valores de reflectância dos pontos mensurados em campo;
- *Weka* versão 3.8.5, para estimar a produtividade cafeeira;
- *Notepad ++*, para extração das informações geradas pelo *weka* e;
- *Minitab*, para a extração de algumas informações;
- *ArcMap* versão 10.4.1 para representação gráfica do resultado

5.3 Método

O trabalho teve como execução os procedimentos descritos conforme o fluxograma representado na Figura 4, no qual subdivide-se em quatro etapas: (I) Aquisição de dados; (II) Organização dos dados; (III) Processos Computacionais e (IV) Análise dos Resultados, explicados posteriormente.

Figura 4- Fluxograma da metodologia do trabalho.



Elaboração: A Autora (2022).

5.3.1 Etapa 1: Aquisição de dados

Após a definição da área de estudo, foram adquiridos todos os dados essenciais para a procedência deste projeto (planilha eletrônica contendo valores de produtividade de cada ponto amostrado; imagens de satélite de todos os meses do período fonológico do cafeeiro e o *shapefile* da delimitação da área de estudo).

Os dados foram adquiridos em Maio de 2019, durante a realização de uma das etapas de um projeto de Iniciação Científica vinculada com o curso de Engenharia de Agrimensura e Cartográfica e Agronomia.

A coleta e mensuração de dados teve como marco inicial o planejamento de campo, que serviu para se determinar os locais para a realização da colheita de café. Esta determinação se deu de maneira aleatória através de uma ferramenta de geração de pontos no *software* QuantumGis versão 2.18. Posteriormente, estes lugares foram demarcados em cada um dos pés de café com fitas coloridas.

A aquisição e coleta de dados totalizaram dois dias intensos de serviços, com auxílio de uma equipe contendo cerca de sete pessoas, sendo que duas pessoas experientes foram responsáveis pela colheita do café e o restante para a mensuração da coordenada de cada ponto; pesagem (para obtenção da variável “peso total”) e maturação do cafeeiro (classificação quanto às classes do café: verde, cana, cereja e seco). A Figura 5 são fotos tiradas na época da coleta de dados (em maio de 2019, período da autopoda).

Figura 5- Aquisição de dados do cafeeiro



Elaboração: A Autora (2019).

Em cada ponto amostrado, foram colhidos frutos referentes a cinco pés de café, sendo elas duas plantas à esquerda, e duas plantas à direita da planta marcada na etapa de planejamento de campo.

Para a mensuração das coordenadas dos pontos amostrados, foi utilizado um receptor GNSS – RTK (*Global Navigation Satellite System - Real Time Kinematic*), cuja base foi instalada em um ponto estratégico (ponto de maior elevação e sem presença de interferências externas). O rover foi utilizado para a mensuração das coordenadas de cada ponto, através do método de posicionamento relativo.

Posteriormente, estes pontos passaram pelo ajuste online para pós-processamento de dados GNSS pelo método de Posicionamento por Ponto Preciso (PPP), no *site* do Instituto Brasileiro de Geografia e Estatística (IBGE).

Por fim, a produtividade foi estimada em laboratório, após a conclusão das etapas de secagem e beneficiamento dos frutos. A produtividade foi dada em sacas por hectares.

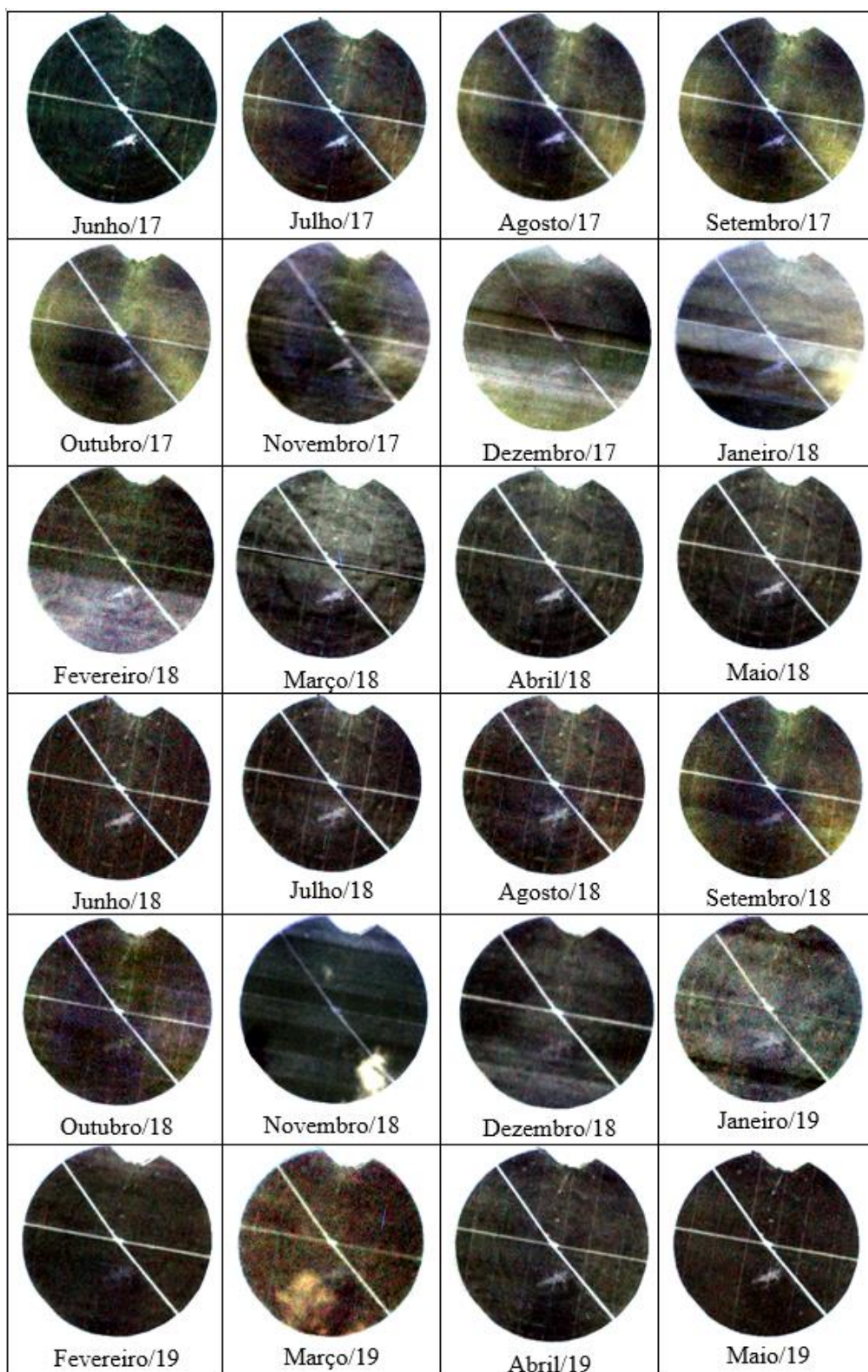
As imagens foram adquiridas de forma gratuita através do *site* do *Planet*, disponível no endereço eletrônico <<https://www.planet.com/>>. Este sensor oferece uma resolução espacial de 3,125m e uma resolução espectral apresentada conforme a Tabela 1.

Tabela 1- Resolução espectral do satélite *Planet*

Nome da banda	Resolução espectral (nm)
B1 (azul)	455–515
B2 (verde)	500–590
B3 (vermelho)	590–670
B4 (infravermelho próximo)	780–860

Fonte: TecTerra (2022).

É importante ressaltar que estas imagens não haja nuvens e/ou ruídos. A Figura 6 ilustra as imagens utilizadas neste estudo, referentes aos 24 meses das fases fenológicas do cafeeiro.

Figura 6- Imagens do satélite *Planet*

Elaboração: A Autora (2022).

A metodologia empregada neste trabalho não foi considerada etapas de tratamentos manuais das imagens de satélite (tais como normalização radiométrica, remoção de possíveis

ruídos, realces e correção de distorções geométricas), pois as mesmas já vêm inseridas durante seu *download*, por meio da plataforma *Planet*.

5.3.2 Etapa 2: Organização dos dados

Para a elaboração do arquivo utilizado para a definição do melhor algoritmo e melhor fase fenológica a se estimar produtividade, foram necessários a execução dos seguintes procedimentos:

5.3.2.1 Extração dos Valores de Reflectância das Bandas de Satélite:

Feito o *download* de todas as imagens de todos os meses pertencentes ao período fenológico do café, foram extraídos os valores de reflectância de todos os pontos amostrais com valores de produtividade mensurados em campo no *software* Envi Classic. Neste caso, por se tratar do satélite *Planet*, foram extraídos valores de reflectância das bandas do espectro do visível (RGB- *Red*, *Green*, *Blue*) e Infravermelho Próximo (IVP). Neste processo, para a extração da reflectância, o arquivo de entrada, contendo os pontos georreferenciados (em formato *shapefile*) se converteu por um arquivo ROI, compatível com o *software* Envi (efv).

5.3.2.2 Cálculo do Índice de Vegetação por Diferença Normalizada:

Alguns Índices de Vegetação podem ser calculados através dos dados obtidos por meio das leituras radiométricas, e podem servir para a identificação das variabilidades que uma cultura pode apresentar (SANTOS JÚNIOR, 2001). O Índice de vegetação por Diferença Normalizada, do inglês *Normalized Difference Vegetation Index* (NDVI) foi inicialmente proposto por Rouse *et al.* (1973) e é considerado uma normalização oriunda de uma razão simples, que resulta em valores entre -1 a 1. Alvos terrestres apresentam valores contidos entre zero e um (quanto mais alto o valor, mais densa/saudável é a vegetação), e a água, com valores resultantes negativos. O NDVI é o índice mais utilizado no sensoriamento remoto e vem sendo bastante empregado devido a sua correlação com as variáveis climáticas da região (BARBOSA *et al.*, 2006), com a capacidade fotossintética, índice de área foliar, biomassa e produtividade primária (PEÑUELAS, 1995). Desta maneira, foi crucial o cálculo do NDVI para compor o arquivo de estudo, cuja fórmula é dada por meio da Equação 1:

$$NDVI = \frac{IVP - VERMELHO}{IVP + VERMELHO} \quad (1)$$

Onde:

- IVP: Valor de reflectância referente ao Espectro do Infravermelho Próximo;
- VERMELHO: Valor de reflectância referente ao Vermelho do Espectro do Visível

5.3.2.3 Arquivo para o *weka*

Dado isto, foi elaborado um arquivo contendo as coordenadas dos pontos mensurados em campo, juntamente com os dados de produtividade com seus respectivos valores de reflectâncias e NDVI, para treinamento e teste dos algoritmos. Os dados foram salvos em um arquivo do tipo ARFF, com uma sintaxe própria, para que posteriormente pudesse ser aberto no *software weka*, cujo resultado destes arquivos está representado conforme a Figura 7.

Figura 7- Configuração do arquivo para o *software weka*

```
Arquivo Editar Formatar Exibir Ajuda
@relation produtividadedados.symbolic
@attribute MAPX numeric
@attribute MAPY numeric
@attribute B1 numeric
@attribute B2 numeric
@attribute B3 numeric
@attribute B4 numeric
@attribute NDVI numeric
@attribute PRODUTIVIDADE numeric
@data
226599,7933839,479,652,699,3652,0.678694553,32.47292
226587,7933824,503,650,686,3592,0.679289388,34.06443
226488,7933797,479,648,692,3637,0.68029568,25.13769
226716,7933791,500,657,702,3691,0.680400637,35.66247
226806,7933779,494,642,694,3666,0.681651376,21.131
227016,7933770,487,648,706,3688,0.678652708,19.45707
226689,7933758,511,634,691,3724,0.686976217,32.96521
226998,7933758,504,661,721,3598,0.666126418,13.82174
226653,7933752,489,651,701,3655,0.678145087,29.80076
226983,7933746,505,653,720,3660,0.671232877,16.80679
```

Elaboração: A Autora (2022).

5.3.3 Etapa 3: Processamento computacionais

Esta etapa foi responsável pela mensuração das correlações de cada uma das bandas compostas pelo satélite *Planet* com a variável produtividade, por meio do *software Minitab*, a fim de se obter resultados plausíveis para a determinação dos intervalos espectrais mais sensíveis a este parâmetro agrônomo.

Além disso, também foram elaboradas séries temporais da reflectância das bandas compostas pelo satélite de estudo, com intuito de se avaliar o comportamento espectral do cafeeiro nos 24 meses do período fenológico do café.

Ademais, também foram realizados testes para a descoberta do melhor algoritmo para se estimar produtividade. Estes testes serviram para analisar o índice da correlação entre a produtividade estimada e a mensurada, além da mensuração do erro médio e da tendência do mesmo, cujas medidas estatísticas são explicadas posteriormente.

5.3.3.1 Geração do Modelo

Para a escolha do melhor modelo de predição da produtividade cafeeira, foram analisados três algoritmos classificadores no *software weka* versão 3.8.5, cujo teste de efetividade de cada um destes algoritmos foi dado por meio do método “divisão por porcentagem”, do inglês “*Percentage Split*”, fixado em 80%, que consta na utilização de 80% dos dados para treinamento dos modelos e 20% dos dados para testes de validação dos valores preditos.

Os algoritmos testados foram: *Support Vector Machine*, *Multilayer Perceptron* e *Random Forest*, algoritmos explicados conforme a seção 4.3 deste presente trabalho (Métodos de aprendizado de máquina para a estimativa de produtividade).

Em todos os algoritmos testados foram utilizados os parâmetros no modo *default* (padrão) do programa. Toda RNA têm como parâmetros principais as camadas ocultas/escondidas, também conhecidas como *Hidden Layers* e a taxa denominada “*Momentum*”. A camada oculta é definida como sendo um conjunto de neurônios representando uma transformação matemática aplicada à camada anterior (*Input Layer*). Recomenda-se não utilizar um grande número de camadas escondidas. Duas camadas são suficientes para a classificação de padrões pertencentes a quaisquer tipos de regiões geométricas que apresentem aprendizados descontínuos, já para funções contínuas, apenas uma camada torna-se suficiente (LIPPMANN, 1987). Muito relevante também, é o parâmetro

Momentum, que deve ser adotado valor compreendido entre 0 e 1, sendo que se adotado 0, a taxa adotada tende a afetar positivamente o processo de convergência da rede. Quanto menor o seu valor, mais lento se torna o treinamento da rede, assim sendo, costuma-se adotar valor próximo a 0,3 (SILVA, 2016; HAYKIN, 2001). No algoritmo *Multilayer Perceptron* foi utilizado *Hidden Layer* = a e *Momentum* = 0,2.

Já no algoritmo *Support Vector Machine*, o parâmetro C é possui uma maior importância dentre os demais. Ele é denominado como sendo o parâmetro de complexidade (ou regularização), e controla o grau de flexibilidade do processo para desenhar a linha para ajustar os dados. O padrão normalmente adotado é 1. O *Kernel* também é de muita relevância, ele pode ser dividido em *Kernel* simples e *Kernel* Polinomial. O *Kernel* simples é linear e separa os dados com uma linha reta, já o *Kernel* Polinomial (o padrão em *weka*), ajusta os dados utilizando uma curva, quanto maior o valor do expoente, mais contorcida é a curva.

No algoritmo *Random Forest*, o parâmetro principal são os números de interações, que nada mais é que o número de interações de uma árvore com a outra, adotou-se como 100 o número de interações.

Para se determinar a qualidade dos arquivos testados por este método, algumas medidas estatísticas puderam desempenhar um papel crucial na decisão da acurácia dos dados, podendo dimensionar a precisão de previsões, verificando a efetividade dos resultados. Sendo assim, a escolha do melhor modelo estatístico se deu pela análise do Coeficiente de Correlação e de duas medidas de erros: a Raiz Quadrada do Erro-Médio (RMSE, do inglês *Root Mean Squared Error*) e o Erro Percentual Absoluto Médio (MAPE, do inglês *Mean Absolute Percentage Error*).

COEFICIENTE DE CORRELAÇÃO

A correlação é uma medida de associação bivariada referente ao nível de associação entre duas variáveis quantitativas, na qual mensura sua direção e relação linear (GARSON, 2009; MOORE, 2007). Assim sendo, o Coeficiente de Correlação (do inglês *Correlation Coefficient*) é uma medida que varia de -1 a 1, e quanto mais próximo de seus extremos, maior o grau de relacionamento entre as variáveis (perfeita correlação). Esta medida está atrelada a força entre duas variáveis testadas, que neste caso, são os valores de reflectâncias e a produtividade mensurada em campo.

RMSE

A Raiz do Erro Quadrático Médio (RMSE, do inglês *Root Mean Squared Error*) traz informação a respeito à dispersão dos dados, cuja medida é expressa pelo erro médio entre os valores observados (reais) e preditos (hipóteses), através do arquivo inserido pelo usuário. Ou seja, o RMSE é utilizado para expressar a acurácia dos resultados numéricos nas mesmas dimensões da variável analisada. Neste caso, os RMSEs foram reproduzidos em porcentagem (%), para maior compreensão quanto ao dimensionamento do erro (representado na Equação 2). Nesta métrica, quanto menor o valor, melhor é o desempenho do modelo.

$$RMSE(\%) = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2}}{n} * n * 100 \quad (2)$$

Onde:

- $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$: Valores preditos;
- y_1, y_2, \dots, y_n : Valores observados e;
- n : Número de observações

MAPE:

O Erro Percentual Absoluto Médio (MAPE, do inglês *Mean Absolute Percentage Error*) é uma escala sensível, não indicada para utilização em caso de baixo volume de dados, cujo mesmo indica a tendência dos dados, designando o quão preciso foi a previsão, e está indicado conforme a Equação 3.

$$M = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \quad (3)$$

Onde:

- M : erro percentual absoluto médio
- N : número de vezes que a iteração da soma acontece
- A_t : Valor real
- F_t : Valor previsto

5.3.4 Etapa 4: Análise dos resultados

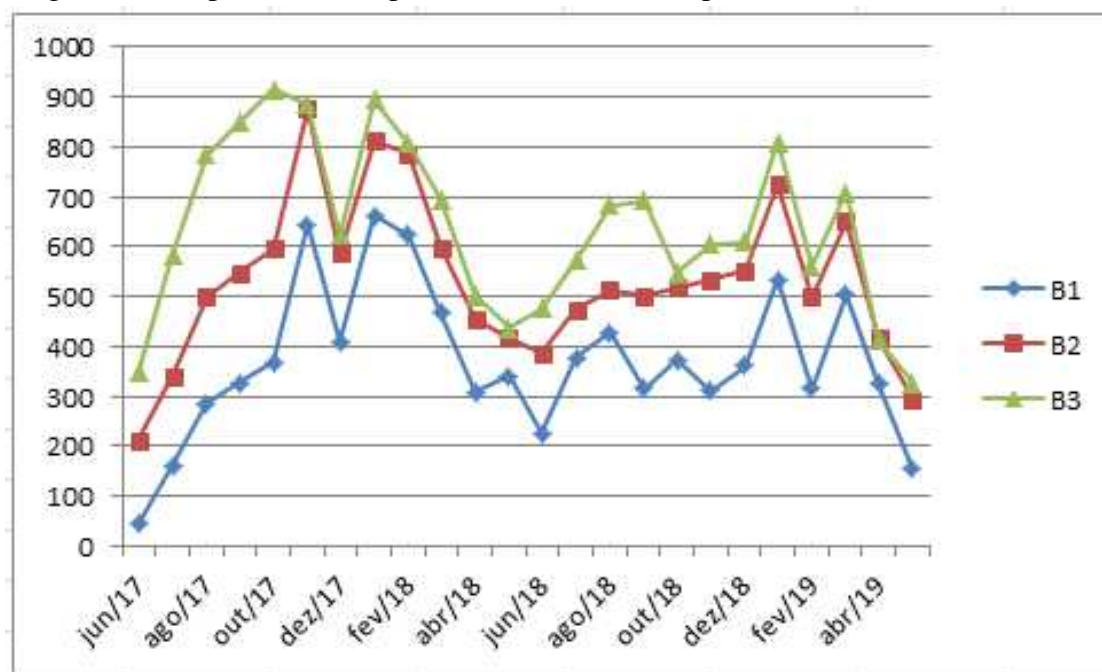
Assim sendo, foi possível determinar a melhor fase do cafeeiro a se estimar produtividade, além da escolha do melhor algoritmo classificador, a partir de gráficos gerados representando o comportamento espectral do cafeeiro oriundo dos 48 pontos mensurados em campo, durante todo o período fenológico do cafeeiro, referentes às bandas do espectro do visível (B1, B2 e B3), do IVP e do NDVI; gráficos indicando a correlação de cada banda de satélite com a variável produtividade e por fim, gráficos contendo os valores de Correlação, RMSE e MAPE obtidos, referentes a todos os meses que compõem o período fenológico do café.

6 RESULTADOS E DISCUSSÕES

Após a conclusão de todos os tópicos descritos conforme o Fluxograma apresentado na seção anterior, foram elaborados alguns gráficos e tabelas a fim de auxiliarem na tomada de decisões, que diz respeito ao melhor algoritmo e período fenológico a se estimar produtividade.

O gráfico presente na Figura 8 representa o comportamento espectral do cafeeiro oriundo dos 48 pontos mensurados em campo, durante todo o período fenológico do cafeeiro, referentes às bandas do espectro do visível (B1, B2 e B3). Para a representação do comportamento espectral do cafeeiro na B4 e do índice NDVI, foram elaborados outros dois gráficos, presentes na Figura 9 e Figura 10, respectivamente. A B4 e o índice NDV foram representados em gráficos distintos devido ao fato do IVP possuir um comportamento espectral com elevado índice de reflectância comparado às bandas do espectro do visível e ao índice NDVI possuir valores incomparáveis aos demais.

Figura 8- Comportamento espectral do cafeeiro no espectro do visível (B1, B2 e B3)



Elaboração: A autora (2022).

De modo geral a vegetação absorve uma elevada quantidade de radiação no espectro do visível, cuja energia é utilizada durante seu processo de fotossíntese (PONZONI, 2001).

Desta forma, a banda do azul (B1) e a banda do vermelho (B3) são caracterizadas por possuírem uma alta sensibilidade com esta taxa de fotossíntese da planta e então quanto menor sua resposta espectral, maior a taxa fotossintética da lavoura.

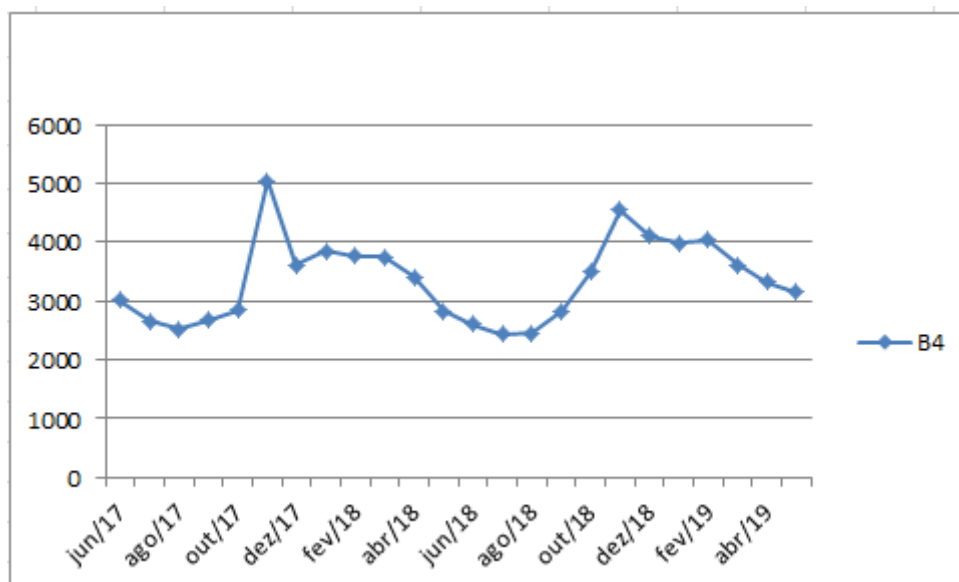
A banda do verde (B2) diz respeito à pigmentação do vegetal, o que explica a coloração verde das plantas (FIGUEIREDO, 2005). Então, quanto maior a resposta espectral, mais vigorosa é a planta, e conseqüentemente, quanto menor a resposta espectral, menos vigorosa ela é.

Em relação à B1, observa-se que nov./17 e jan./18 são os meses que possuem maior índice de reflectância, períodos nos quais são pertencentes ao estágio vegetativo do cafeeiro, e são caracterizados por possuírem as menores taxas de fotossíntese das plantas. Já jun./17 e mai./19 períodos de menor resposta espectral, cujas fases pertencem justamente ao primeiro mês da primeira fase fenológica do café (referente ao período vegetativo) e ao último mês da última fase fenológica (autopoda), e correspondem a períodos com maior taxa fotossintética.

Quanto a B2, nota-se uma similaridade com o comportamento espectral do cafeeiro na banda do azul (B1), pois nov./17 e jan./18 (meses pertencentes ao estágio vegetativo do cafeeiro) foram os meses que apresentaram uma maior resposta espectral, acrescido do período de jan./18 a mar./18 (meses pertencentes ao final do período vegetativo do café, referente à 2ª fase do período fenológico do cafeeiro), que também possuem elevado nível de reflectância, comparado com os demais meses. Já o mês com menor resposta espectral foi referente à jun.17, estágio inicial do cultivo do cafeeiro, cujo mês é caracterizado por possuir um dossel menos vigoroso.

Em relação a B3, nota-se que os meses com maior reflectância foram detectados em out./17, nov./17, períodos pertencentes à 1ª fase da fenologia do café, referente ao período vegetativo, e também a jan./18, primeiro mês oriundo da 2ª fase fenológica do cafeeiro, ainda pertencente ao estágio vegetativo. Os meses de jun./17 e mai./18 tiveram um baixo nível de reflectância comparado aos demais meses.

Figura 09- Comportamento espectral do cafeeiro na banda do infravermelho próximo (IVP)

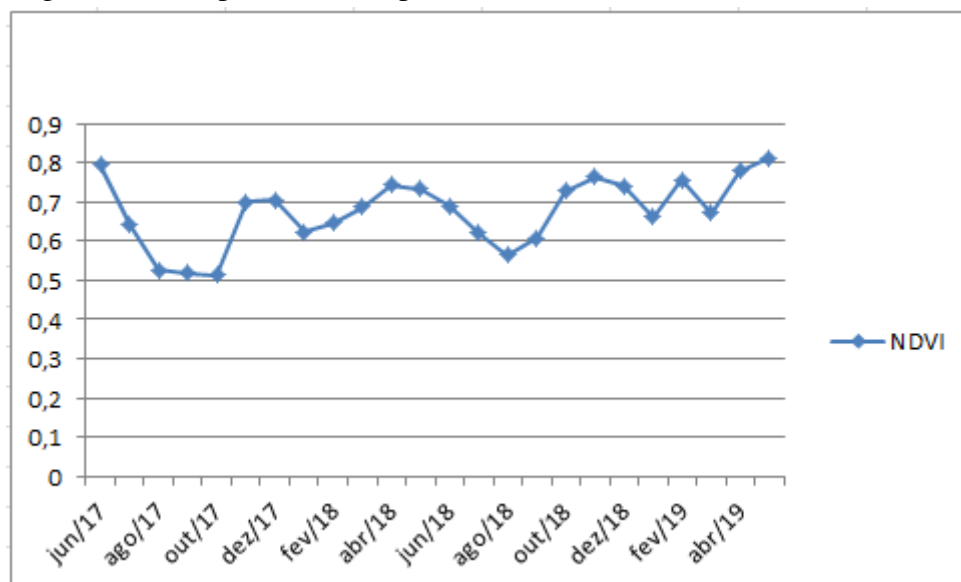


Elaboração: A autora (2022).

O infravermelho próximo é caracterizado por possuir um comportamento espectral com elevado índice de reflectância e está atrelado à estrutura celular interna da folha, sendo assim, possui sensibilidade com a estimativa indireta da biomassa (PEÑUELAS, 1998) e com o índice de área foliar. Segundo Ponzoni (2001), quanto maior for a estrutura interna foliar, maior será o índice de espalhamento da radiação interna incidente, ocasionando maior reflectância. Sendo assim, quanto maior a resposta espectral do IVP, maiores são estes índices mencionados.

Com a Figura 09 percebe-se que o IVP teve uma maior reflectância em nov./17 (1ª fase do período fenológico do café, pertencente ao estágio vegetativo cafeeiro), além dos períodos de jan./18 a mar./18 (período relativo à 2ª fase, pertencente ao final do período vegetativo do café) e nov./18 a jan./19. Já o período de mai./18 a set./18 (início do período reprodutivo do café) é caracterizado por ter tido uma queda considerável na resposta espectral do IVP, indicando um menor índice de biomassa e área foliar.

Figura 10- Comportamento espectral do cafeeiro referente ao índice NDVI



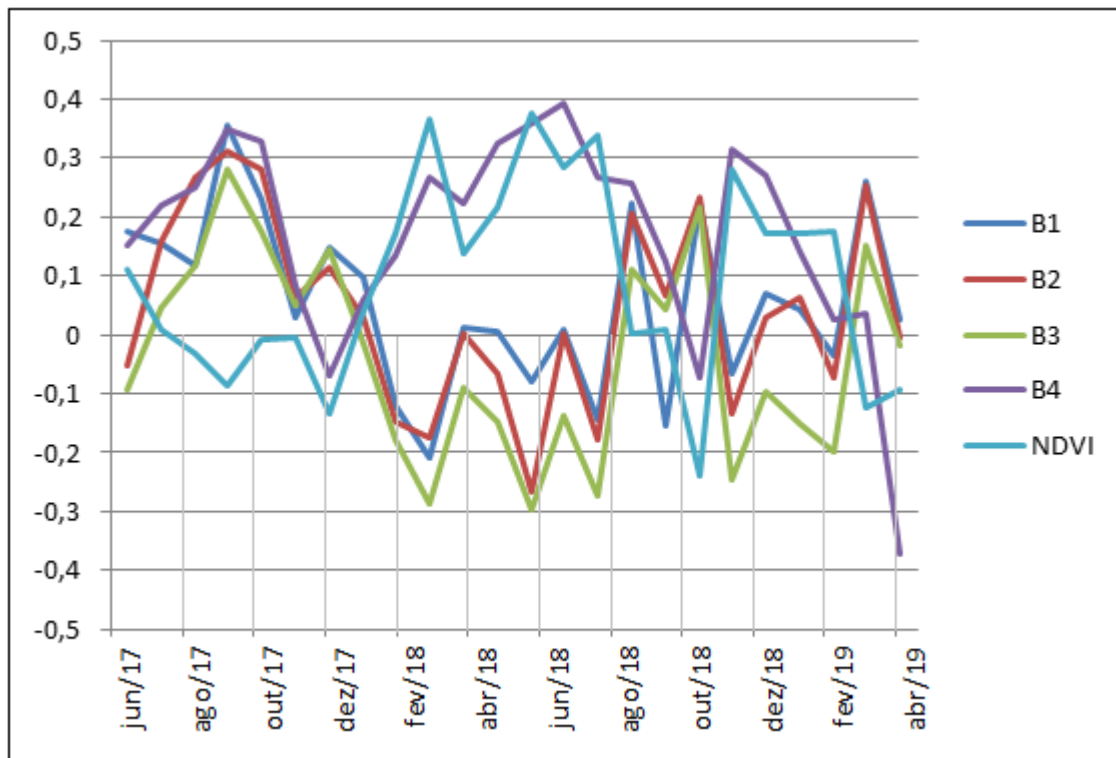
Elaboração: A autora (2022).

O índice NDVI, calculado em função da banda do vermelho e infravermelho próximo, também diz respeito à biomassa e ao índice de área foliar, e conseqüentemente, segundo Molin (2001) e Figueiredo (2005), reflete o grau de desenvolvimento do plantio, cujo mesmo expressa a expectativa quanto ao seu rendimento.

Na Figura 10, pode-se observar que os meses com maior resposta espectral do índice NDVI foram jun./17 (primeiro mês do período fenológico do cafeeiro, referente ao estágio vegetativo da planta) e mai./19 (último mês do período fenológico do cafeeiro, referente à autopoda).

Além do comportamento espectral do cafeeiro ao longo das seis fases de seu período fenológico, foi mensurada a correlação de cada banda de satélite com a variável produtividade, cujos resultados são apresentados conforme a Figura 11.

Figura 11- Correlação de cada banda de satélite e NDVI com a produtividade mensurada em campo



Elaboração: A autora (2022).

A representação em tabela destas informações a respeito da correlação de cada banda de satélite e índice NDVI com a produtividade mensurada em campo está presente na Tabela 2.

Tabela 2- Correlação de cada banda de satélite e NDVI com a produtividade mensurada em campo

Período Fenológico	BANDAS DE SATÉLITE				
	B1	B2	B3	B4	NDVI
Junho/2017	0,174	-0,053	-0,093	0,152	0,11
Julho/2017	0,155	0,158	0,046	0,22	0,008
Agosto/2017	0,118	0,266	0,117	0,250	-0,032
Setembro/2017	0,357	0,310	0,282	0,348	-0,085
Outubro/2017	0,230	0,282	0,176	0,328	-0,007
Novembro/2017	0,030	0,063	0,050	0,084	-0,005
Dezembro/2017	0,148	0,115	0,146	-0,068	-0,135
Janeiro/2018	0,099	0,029	-0,014	0,061	0,039
Fevereiro/2018	-0,121	-0,147	-0,177	0,135	0,172
Março/2018	-0,208	-0,176	-0,286	0,266	0,365
Abril/2018	0,011	0,003	-0,090	0,224	0,137
Maio/2018	0,004	-0,067	-0,149	0,324	0,217
Junho/2018	-0,081	-0,266	-0,298	0,359	0,375
Julho/2018	0,010	0,002	-0,138	0,393	0,285
Agosto/2018	-0,145	-0,177	-0,272	0,266	0,338
Setembro/2018	0,223	0,207	0,111	0,258	0,001
Outubro/2018	-0,155	0,066	0,042	0,124	0,010
Novembro/2018	0,224	0,234	0,216	-0,074	-0,24
Dezembro/2018	-0,066	-0,134	-0,245	0,314	0,282
Janeiro/2019	0,069	0,029	-0,097	0,271	0,172
Fevereiro/2019	0,043	0,062	-0,150	0,140	0,173
Março/2019	-0,036	-0,071	-0,199	0,026	0,177
Abril/2019	0,262	0,254	0,153	0,035	-0,124
Maio/2019	0,027	-0,003	-0,017	-0,373	-0,093

Elaboração: A Autora (2022).

Dado o gráfico e a tabela acima, em geral, observa-se que set./17, out./17 e mar./18 são meses com valores de reflectância das bandas de satélite (B1, B2, B3 e B4) mais correlatos com a variável produtividade. Além disso, é possível observar que o IVP (B4),

banda responsável pela detecção de possíveis problemas não visíveis a olho nu, possui uma maior correlação entre os períodos referente a set./17, out./17 e mar./18 a set./18.

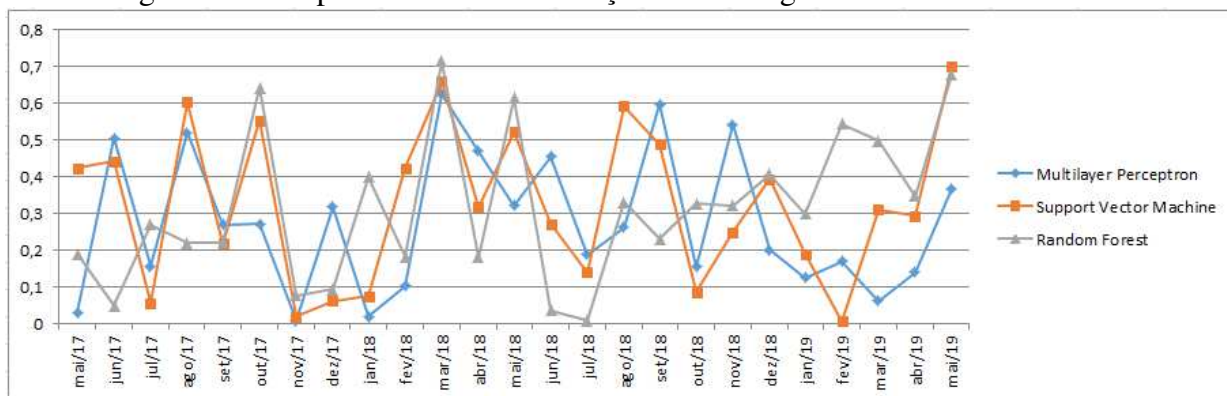
Quanto ao índice NDVI, obteve-se uma maior correlação com a produtividade nos meses de mar./18 e jun./18 a ago./18, período referente ao final do estágio vegetativo e início do período reprodutivo do ciclo do café.

Já nov./17, jan./18, abr./18 a jul./18 (referente ao período do repouso e início do período reprodutivo) e jan./19 a mar./19 (referente à 5ª fase da fenologia do café, estágio da maturação dos frutos) foram os meses menos correlatos.

Por último, foram testados três algoritmos classificadores para predição de produtividade, no *software weka*, sendo eles: *Multilayer Perceptron*; *Support Vector Machine* e *Random Forest*. Nesta etapa de testes foram mensuradas três medidas estatísticas, sendo elas a correlação do algoritmo com a variável a ser testada; RMSE e MAPE, já mencionadas anteriormente, que serviram como fatores adicionais para tomada de decisões da melhor época e algoritmo para se estimar a produtividade cafeeira.

A Figura 12, Figura 13 e Figura 14 apresenta o índice de Correlação, RMSE e MAPE dos três algoritmos testados, consecutivamente.

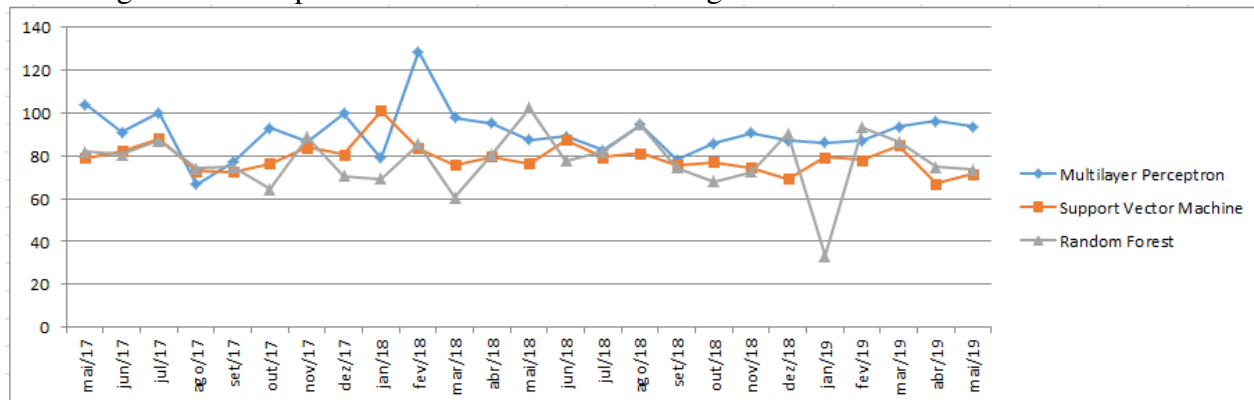
Figura 12- Comportamento da Correlação com os algoritmos testados



Elaboração: A autora (2022).

Mar./2018 destaca-se como sendo o mês mais correlato dentre os três algoritmos testados, além de jun./17, ago./17, set./18 e nov./18 no algoritmo *Multilayer Perceptron*, ago./17, out./17, ago./18 e mai./19 no algoritmo *Support Vector Machine* e out./17 e mai./18 em *Random Forest*. Em geral, o mês que possuiu uma menor correlação em ambos os algoritmos foi nov./17, acrescido de mai./17 e jan./18 em *Multilayer Perceptron*, jul./17 e fev./19 em *Support Vector Machine* e jun./17, jun./18 e jul./18 em *Random Forest*.

Figura 13- Comportamento do RMSE com os algoritmos testados



Elaboração: A autora (2022).

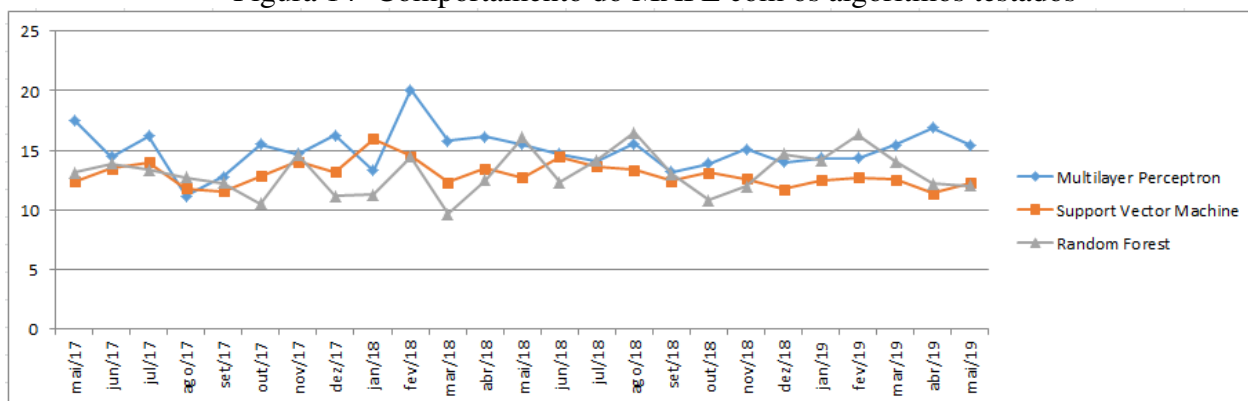
Esta medida expressa o percentual de erro do algoritmo ao estimar a produtividade de pontos desconhecidos através de outros já mensurados em campo, então, quanto menor o valor, mais preciso se torna o algoritmo.

Em relação ao algoritmo *Multilayer Perceptron* nota-se que o mês de ago./17 foi o que respondeu melhor a esta medida, em contrapartida, fev./18 foi o mês que obteve mais erros para a estimativa da produtividade do cultivo do cafeeiro.

No que diz respeito ao algoritmo *Support Vector Machine*, mês com maior percentual de erro foi referente à segunda fase da fenologia do café (jan./18), e os demais meses obtiveram um erro mais homogêneo.

Por último, quanto ao *Random Forest*, jan./19 destacou-se por possuir um menor erro e em seguida mar./18. Por outro lado, mai./18 foi o mês com maior percentual médio de erro registrado durante a etapa de estimativa de produtividade.

Figura 14- Comportamento do MAPE com os algoritmos testados



Elaboração: A autora (2022)

O MAPE diz respeito ao quão tendencioso o algoritmo é, então, assim como o RMSE, quanto menor o MAPE, melhor.

Quanto ao algoritmo *Multilayer Perceptron*, ago./17 destaca-se por ser o mês com menor índice do MAPE, e fev./18 o mês com resultados preditos mais tendenciosos.

Com a utilização do algoritmo *Support Vector Machine* observou-se que em jan./18 foi o mês que se obteve uma predição mais tendenciosa, e jul./17 e ago./17 menos tendenciosos. Os demais meses se mantiveram mais constantes, ou seja, com menores oscilações do valor do MAPE.

Random Forest destacou-se mar./18 como sendo o mês com menor valor do MAPE relatado, e em contrapartida, mai./18 e ago./18 os meses com valores mais tendenciosos.

Por último, foi gerado um arquivo para teste contendo a junção dos dados de todos os meses cuja reflectância das bandas foram mais correlatas com a variável produtividade (apresentado conforme a Figura 12). Sendo assim, este modelo consistiu nas informações extraídas de set./17, out./17, mar./18, jun./18, set./18, nov./18 e abr./19. Os resultados da correlação, RMSE e MAPE estão representados conforme a Tabela 3.

Tabela 3- Valores estatísticos com modelo gerado contendo os meses mais correlatos do período fenológico do cafeeiro.

	<i>Multilayer Perceptron</i>	<i>Support Vector Machine</i>	<i>Random Forest</i>
Correlação	0,1455	-0,02	0,063
RMSE	68,6582	67,3612	70,0210
MAPE	11,9236	11,6561	12,533

Elaboração: A Autora (2022).

Em relação à Tabela 3, nota-se que os resultados obtidos com o modelo gerado com os meses mais correlatos do período fenológico do cafeeiro não superaram aos modelos criados com meses isolados.

Sendo assim, em suma, percebe-se que mar./18 foi um dos meses mais correlatos e assertivos em todos os três algoritmos testados. Março de 2018 pertence ao último mês do período vegetativo do café, contido na segunda fase fenológica do cafeeiro e Janeiro de 2019 é o primeiro mês da 5ª da fenologia do café, referente à fase de maturação dos frutos.

Março é o mês que está na transição do verão para o outono, cuja época é caracterizada pela passagem de temperaturas elevadas, com alto índice de chuva para uma estação marcada pela redução de temperatura e umidade do ar. Esta transição, de um período

chuvoso para um clima seco, propicia uma menor absorção da planta, e conseqüentemente um maior índice de reflectância (MOREIRA, 2004).

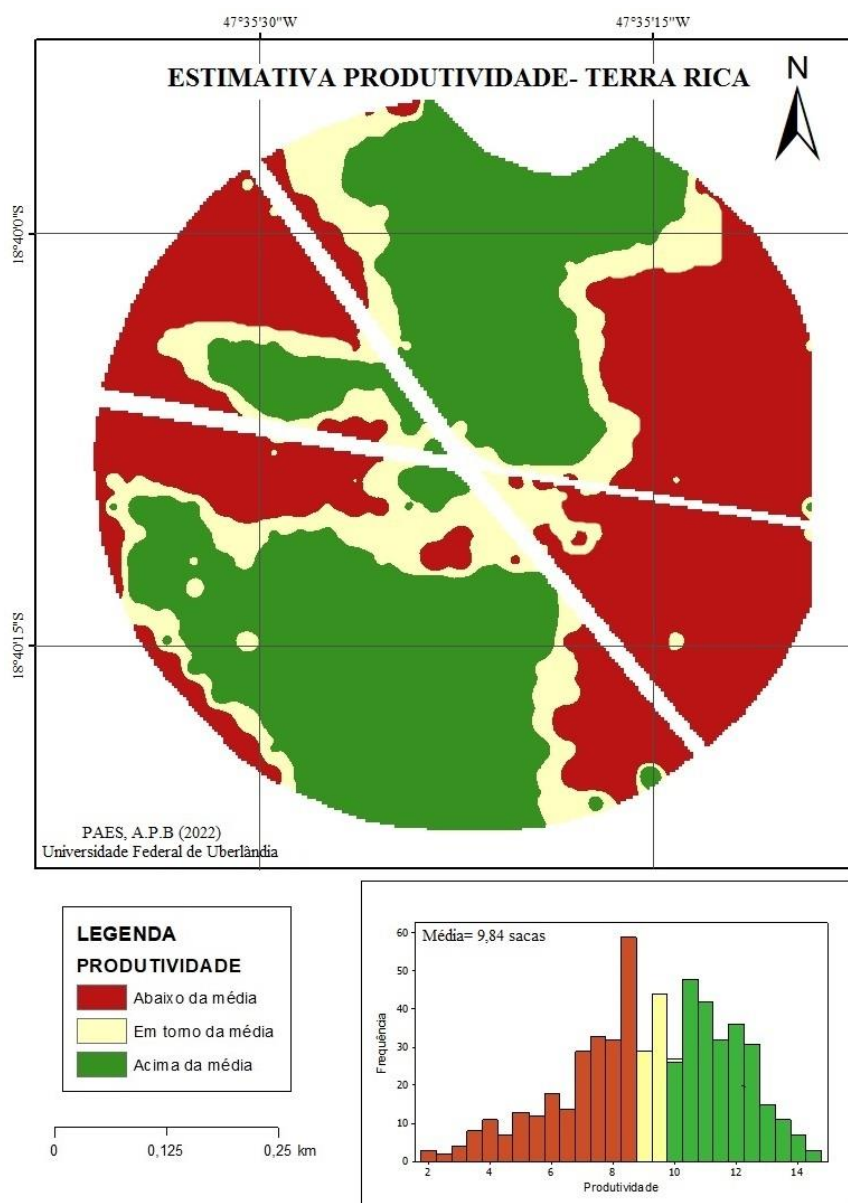
Além disso, nota-se que mar./18 foi o mês que obteve reflectância com maior correlação com a variável produtividade, conforme já mencionado na análise da Tabela 1. Em acréscimo, a B2 e B4 tiveram um alto índice de reflectância neste período, comparado aos demais meses.

Quanto aos algoritmos, percebe-se que o *Support Vector Machine* respondeu melhor aos testes feitos em relação às medidas estatísticas. Foi por meio dele que obteve-se as melhores correlações, menores erros e valores tendenciosos.

Desta forma, conclui-se que o algoritmo *Support Vector Machine* pertencente ao mês de março de 2018, referente ao final do período vegetativo do cafeeiro, respondeu melhor às análises realizadas.

Com isto, na Figura 15 podemos analisar a representação gráfica do resultado gerado pelo *software weka*, com a estimativa de valores de produtividade de pontos não mensurados, a partir de valores mensurados em campo.

Figura 15 - Representação da estimativa da produtividade cafeeira



Elaboração: A Autora (2022).

A Figura 15 é uma das possíveis formas de representação dos valores de saída disponibilizados pelo *software weka*, contendo a estimativa da produtividade cafeeira em todo o talhão de interesse.

É interessante haver essa representação em forma de mapa da estimativa da produtividade na região cafeeira, para que o cafeicultor possa identificar áreas que necessitam de mais cuidados, ou seja, de áreas que necessitam de uma melhoria na gestão, a fim de economizar insumos agrícolas, para que futuramente haja uma melhoria na produtividade e consequentemente, lucro para o produtor.

Os mapas apresentados na Figura 22 foram divididos em três classes:

- Vermelha- representando a Baixa Produtividade;
- Amarelo- representando a Média Produtividade;
- Verde- representando a Alta Produtividade.

Vale ressaltar que os limites de cada intervalo podem ser alterados de acordo com o desejo do produtor, podendo assim produzir um mapa de acordo com as necessidades do contratante.

7 CONCLUSÃO

Neste estudo foram testados três algoritmos matemáticos no *software weka* (*multilayer perceptron*, *support vector machine* e *random forest*), em relação a três medidas estatísticas, sendo elas a correlação, o RMSE e o MAPE.

É importante ressaltar a importância da alta correlação entre o valor de reflectância das bandas de satélite referente aos pontos mensurados em campo e a variável produtividade, pois esta medida estatística diz respeito à eficiência assertiva do modelo. Ou seja, o algoritmo consegue discriminar o local que possui maior e menor produtividade do talhão, o que torna um elemento fundamental para o produtor rural, possibilitando-o realizar uma aplicação de insumos nos locais corretos e nas quantidades requeridas, reduzindo custos e minimizando horas de serviço. Ressalta-se também que os valores obtidos de RMSE não são considerados bons.

Sendo assim, com base em todos os gráficos e tabelas apresentados na seção anterior, detectou-se que o melhor período a se estimar a produtividade (período no qual houve maior correlação, e foi mais assertivo e menos tendencioso) foi referente ao último mês do período vegetativo do café, referente à segunda fase fenológica do cafeeiro. Este período, neste caso, referente a março de 18, é caracterizado pela transição do verão para o outono, cujo mesmo é caracterizado por um clima ameno e seco, propício para melhor mapeamento das lavouras através de imagens multiespectrais (devido a baixa absorção e conseqüentemente uma maior reflectância da lavoura para os sensores acoplados nos satélites).

Março do 2018 obteve -0,208 de correlação entre a variável “produtividade” e a B1; -0,176 com a B2; -0,286 com a B3; 0,266 com a B4 e 0,365 com o índice NDVI. Além disso, obteve maiores correlações em todos os algoritmos testados (correlação acima de 0,5 em todos os casos).

Além disso, conclui-se que para este trabalho, e para esta área de estudo, o algoritmo *Support Vector Machine* foi o modelo estatístico que obteve as melhores respostas em relação às medidas estatísticas, além de ser responsável por obter o maior número de meses correlatos com a produtividade, e valores mais assertivos.

REFERÊNCIAS

- ARTUZO, F. D.; SOARES, C.; WEISS, C. R. Inovação de processo: O impacto ambiental e econômico da adoção da agricultura de precisão. **Espacios**, v. 38, n. 2, p. 1-6, 2017.
- ATZBERGER, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. **Remote sensing**, v. 5, n. 2, p. 949-981, 2013.
- BARBOSA, H. A.; HUETE, A. R.; BAETHGEN, W. E. A 20-year study of NDVI variability over the Northeast Region of Brazil. **Journal of arid environments**, v. 67, n. 2, p. 288-307, 2006.
- BERTOLIN, N. O.; FILGUEIRAS, R.; VENANCIO, L. P.; MANTOVANI, E. C. Predição da produtividade de milho irrigado com auxílio de imagens de satélite. **Revista Brasileira de Agricultura Irrigada**, v. 11, n. 4, p. 1627, 2017.
- BESTELMEYER, B. T.; MARCILLO, G.; McCORD, S. E.; MIRSKY, S.; MOGLEN, G.; NEVEN, L. G.; PETERS, D.; SOHOULANDE, C.; WAKIE, T. Scaling up agricultural research with artificial intelligence. **IT Professional**, v. 22, n. 3, p. 33-38, 2020.
- BRANDÃO, Z. N.; SOFIATTI, V.; BEZERRA, J. R.; FERREIRA, G. B. Estimativa da produtividade do algodoeiro através de imagens de satélite. In: **CONGRESSO BRASILEIRO DO ALGODÃO**, 8.; COTTON EXPO, 1., 2011, São Paulo. Evolução da cadeia para construção de um setor forte: Anais. Campina Grande, PB: Embrapa Algodão, 2011.
- BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- CALERA, A.; CAMPOS, I.; OSANN, A.; URSO, G.; MENENTI, M. *Remote sensing for crop water management: from ET modelling to services for the end users*. **Sensors**, v. 17, n. 5, p. 1104, 2017.
- CAMARGO, Ângelo Paes de; CAMARGO, Marcelo Bento Paes de. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, v. 60, p. 65-68, 2001.
- CAMARGO, AP de; FRANCO, C. F. Clima e fenologia do cafeeiro. **Cultura de café no Brasil: manual de recomendações**, v. 5, p. 19-50, 1985.
- DE OLIVEIRA, Itamar Pereira; OLIVEIRA, Luana Carvalho; DE MO, Camila Stéffane Fernandes Teixeira. Cultura de café: histórico, classificação botânica e fases de crescimento. **Revista Eletrônica Faculdade Montes Belos**, v. 5, n. 4, 2012.

FIGUEIREDO, Divino. Conceitos básicos de sensoriamento remoto. **São Paulo**, 2005.

FORMAGGIO, Antonio Roberto; SANCHES, Ieda Del'Arco. **Sensoriamento remoto em agricultura**. Oficina de Textos, 2017.

FORMAGGIO, A. R. et al. O sensoriamento remoto na agricultura: conceitos básicos, metodologia e aplicações. **São José dos Campos: INPE**, p. 217, 1989.

FURTADO, G. F.; SILVA, E. V.; FURTADO, B. E. F.; PALMEIRA, A. C. R. AS GEOTECNOLOGIAS E O AGRONEGÓCIO BRASILEIRO: UM ESTUDO DE CASO NA PRODUÇÃO DE SOJA. **Revista Brasileira De Agrotecnologia**, v. 10, n. 3, p. 88-96, 2020.

BRUNET, James R.; GARSON, G. David. Dimensionality analysis as a computerized tool for strategic planning in policing and security. **Social science computer review**, v. 27, n. 2, p. 228-242, 2009.

GAMA, J. a. Functional trees. **Machine Learning**, v. 55, p. 219–250, 2004. 22, 31, 32, 50

GASQUES, José Garcia et al. **Gastos públicos em agricultura: retrospectiva e prioridades**. IPEA, Instituto de Pesquisa Econômica Aplicada, 2006.

GASQUES, José Garcia; VERDE, Carlos M. Villa. Novas fontes de recursos, propostas e experiências de financiamento rural. **Revista de Economia e Sociologia Rural**, v. 34, n. 3, p. 39-80, 2019.

GARSON, G. David. Statnotes: Topics in multivariate analysis. **North Carolina State University**. <https://faculty.chass.ncsu.edu/garson/PA765/statnote.htm> (accessed Feb. 02, 2021), 2009.

GOUVEIA, N. M. Estudo da diferenciação e crescimento das gemas florais de *Coffea arabica* L.: observações sobre antes e maturação dos frutos. 1984.

HAYKIN, S. **Redes Neurais: princípios e práticos**. 2. ed. Porto Alegre: Bookman, 2001. 899

JACKSON, R. D.; IDSO, S. B.; REGINATO, R. J.; PINTER JUNIOR., P. J. Canopy temperature as a crop water stress indicator. **Water resources research**, v. 17, n. 4, p. 1133-1138, 1981.

JOHANN, J. A.; OPAZO, M. A. U.; SOUZA, E. G.; ROCHA, J. V. Variabilidade espacial dos atributos físicos do solo e da produtividade em um Latossolo Bruno distrófico da região de Cascavel, PR. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 8, p. 212-219, 2004.

LIPPMANN, R. An introduction to computing with neural nets. **IEEE Assp magazine**, v. 4, n. 2, p. 4-22, 1987.

MARTINS, G. D.; GALO, M. L. B. T. Detecção de áreas infestadas por nematoides e *Migdolus fryanus* em cultura canavieira a partir de imagens multiespectrais Rapideye. **Revista Brasileira de Cartografia**, v. 66, n. 2, 2014.

MACHADO, P. L. O. A.; BERNARDI, A. C. C.; SILVA, C. A.; CARMO, C. A. F. S.; MEIRELLES, M. S. P.; MANZATTO, C. V. Estudo de caso em agricultura de precisão: manejo de lavoura de soja na região de campos gerais, PR. **Agricultura de precisão para o manejo da fertilidade do solo em sistema plantio direto**, p. 93, 2004.

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

MOORE, David S.; KIRKLAND, Stephane. **The basic practice of statistics**. New York: WH Freeman, 2007.

NOVO, E. M.L. **Sensoriamento Remoto: princípios e aplicações**. Editora Blucher, 2010

PEÑUELAS, Josep.; FILELLA, Iolanda. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. **Trends in Plant Science**, London, v.3, n.4, p.151-156, 1998.

PEÑUELAS, Josep; FILELLA, Iolanda; GAMON, John A. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. **New Phytologist**, v. 131, n. 3, p. 291-296, 1995.

PEROBELLI, Fernando Salgueiro et al. Produtividade do setor agrícola brasileiro (1991-2003): uma análise espacial. **Nova economia**, v. 17, p. 65-91, 2007.

PIRES, João Leonardo Fernandes et al. Discutindo agricultura de precisão-aspectos gerais. **Embrapa Trigo-Docmentos (INFOTECA-E)**, 2004.

Ponzoni, F.J. Comportamento espectral da vegetação. In: MENESES, P. R.; NETTO, J. da S. M. (Org.). **Sensoriamento Remoto: reflectância dos alvos naturais**. Brasília, UnB, 2001.

RAMIREZ, G. M.; ZULLO JÚNIOR, J. Estimativa de parâmetros biofísicos de plantios de café a partir de imagens orbitais de alta resolução espacial. **Engenharia Agrícola**, v. 30, p. 468-479, 2010.

ROSA, Viviane Gomes Cardoso da et al. Estimativa da produtividade de café com base em um modelo agrometeorológico-espectral. **Pesquisa agropecuária brasileira**, v. 45, n. 12, p. 1478-1488, 2010.

ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the great plains with ERTS. In: **Third ERTS Symposium**, Proceedings, NASA SP-351, NASA, Washington, DC, v. 1, p. 309-317, 1973.

ROZA, D. Novidade no campo: Geotecnologias renovam a agricultura. **Revista InfoGEO**, n 11, jan/fev. 2000. Disponível em: <http://www.infogeo.com.br/Revista/materia_11.htm>. Acessado em: 01 de agosto de 2021.

SANTOS JUNIOR, R. F.; SANTOS, J. M.; RUDORFF, B. F. T.; MARCHIOORATO, I. A. Detecção de *Heterodera glycines* em plantio de soja mediante espectrorradiometria no visível e infravermelho próximo. **Fitopatologia Brasileira**, v. 27, p. 355-360, 2002.

TEIXEIRA, J. C. Modernização da agricultura no Brasil: impactos econômicos, sociais e ambientais. **Revista Eletrônica da Associação dos Geógrafos Brasileiros Seção Três Lagoas**, p. 21-42, 2005.

TSCHIEDEL, M.; FERREIRA, M. F. Introdução à agricultura de precisão: conceitos e vantagens. **Ciência Rural**, v. 32, p. 159-163, 2002.

VIEIRA FILHO, J. E. R.; FISHLOW, A. Agricultura e indústria no Brasil: inovação e competitividade. 2017.