

CARLOS ALBERTO MATIAS DE ABREU JÚNIOR

ESTIMATIVA DE PRODUTIVIDADE DA CULTURA DO MILHO A PARTIR DE
IMAGENS MULTIESPECTRAIS

Dissertação apresentada ao Programa de Pós-Graduação em
Agricultura e Informações Geoespaciais da Universidade
Federal de Uberlândia, Campus Monte Carmelo, como
parte das exigências para obtenção do título de “Mestre”.

Orientador

Prof. Dr. George Deroco Martins

Monte Carmelo
Minas Gerais - Brasil
2022

CARLOS ALBERTO MATIAS DE ABREU JÚNIOR

ESTIMATIVA DE PRODUTIVIDADE DA CULTURA DO MILHO A PARTIR DE
IMAGENS MULTIESPECTRAIS

Dissertação apresentada ao Programa de Pós-Graduação em
Agricultura e Informações Geoespaciais da Universidade
Federal de Uberlândia, Campus Monte Carmelo, como
parte das exigências para obtenção do título de “Mestre”.

APROVADA em 26 de maio de 2022

Banca examinadora

Prof. Dr. George Deroco Martins

Prof. Dr. Douglas José Marques – Universidade Federal de Uberlândia

Prof^a. Dr^a. Maria de Lourdes Bueno Trindade Galo – Universidade Estadual Paulista

Prof. Dr. George Deroco Martins
Faculdade de Engenharia Civil

MONTE CARMELO
MINAS GERAIS – BRASIL
2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

A162 Abreu Jøenior, Carlos Alberto Matias de, 1996-
2022 Estimativa de produtividade da cultura do milho a
partir de imagens multiespectrais [recurso eletrñico] /
Carlos Alberto Matias de Abreu Jøenior. - 2022.

Orientador: George Deroco Martins.
Dissertaçãoo (Mestrado) - Universidade Federal de
Uberlândia, Pã-graduaçãoo em Agricultura e Informaões
Geoespaciais.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.di.2022.264>

Inclui bibliografia.

1. Agronomia. I. Martins, George Deroco, 1987-,
(Orient.). II. Universidade Federal de Uberlândia. Pã-
graduaçãoo em Agricultura e Informaões Geoespaciais.
III. Título.

CDU: 631



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
 Coordenação do Programa de Pós-Graduação em Agricultura e Informações
 Geoespaciais

Rodovia LMG 746, Km 01, s/nº, Bloco 1AMC, Sala 1A202, Monte Carmelo-MG, CEP 38.500-000
 Telefone: (34) 3810-1033 - ppgaig@iciag.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Agricultura e Informações Geoespaciais				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	26/05/2022	Hora de início:	13:30	Hora de encerramento:	17:00
Matrícula do Discente:	32012AIG005				
Nome do Discente:	Carlos Alberto Matias de Abreu Júnior				
Título do Trabalho:	Estimativa de produtividade da cultura do milho a partir de imagens multiespectrais				
Área de concentração:	Informações geoespaciais e tecnologias aplicadas à produção agrícola				
Linha de pesquisa:	Aplicações e desenvolvimento de métodos em informações espaciais				

Reuniu-se na sala virtual <https://conferenciaweb.rnp.br/webconf/george-deroco-martins> a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agricultura e Informações Geoespaciais, assim composta: Douglas José Marques (UFU), Dra. Maria de Lourdes Bueno Trindade Galo (Unesp – Faculdade de Ciência e Tecnologia – Departamento de Cartografia), Dr. George Deroco Martins (UFU) – orientador do candidato.

Iniciando os trabalhos o presidente da mesa, Dr. George Deroco Martins, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado com louvor

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **George Deroco Martins, Professor(a) do Magistério Superior**, em 01/06/2022, às 09:11, conforme horário oficial de Brasília, com fundamento no art. 6º,



§ 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Maria de Lourdes Bueno Trindade Galo, Usuário Externo**, em 01/06/2022, às 09:42, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Douglas José Marques, Professor(a) do Magistério Superior**, em 01/06/2022, às 23:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3641931** e o código CRC **2C50BEAB**.

Dedico

Aos meus pais Carlos e Josina

AGRADECIMENTOS

O desenvolvimento desta pesquisa foi possível devido ao esforço e à participação de diversas pessoas. Dentre contribuições diretas e indiretas, gostaria de agradecer profundamente:

Primeiramente a Deus, detentor de toda sabedoria e bondade, por ter me permitido completar mais essa jornada e pelas pessoas que tem colocado em meu caminho.

Aos meus pais, Carlos Alberto Matias de Abreu e Josina Geralda Moreira de Souza Abreu, pelos esforços e apoio incansáveis, imensuráveis e incondicionais, que me fizeram mais forte. Esse caminho teria sido exponencialmente mais difícil sem o apoio de vocês. Talvez fosse impossível. Por isso, agradeço-os profundamente, pois devo a vocês tudo o que sou.

Aos meus irmãos Jhonathan Souza de Abreu e Eduarda Souza de Abreu, por terem me dado apoio mesmo à distância. E, acima de tudo, por terem brindado nossas vidas com as princesas Helena e Aurora.

Ao meu sobrinho Miguel Eduardo, pelas jogatinas e momentos que tivemos juntos.

À minha namorada Pâmela Inês de Souza Castro, por todo o companheirismo e apoio imensuráveis. Por ter sido o raio de sol que iluminou os dias mais difíceis, tornando-os mais leves e sutis.

Ao meu orientador, prof. Dr. George Deroco Martins, por todo conhecimento e confiança depositados na minha formação, desde os tempos de graduação. Agradeço também por ter me ensinado lições que vão além da ciência e, sobretudo, por ter se tornado um amigo.

À minha amiga e parceira de pesquisa, Laura Cristina Moura Xavier, pela ajuda e conhecimentos transmitidos desde a época da graduação.

À Universidade Federal de Uberlândia e ao Programa de Pós-Graduação em Agricultura e Informações Geoespeciais, por terem aberto a porta do conhecimento e da profissionalização para mim.

À Lallemand Plant Care, pelo fomento da pesquisa e pelos dados cedidos, que tornaram todo o estudo aqui desenvolvido mais viável.

A todos aqueles, que mesmo não sendo citados aqui, contribuíram de alguma forma para que eu chegasse nesse momento, deixo o meu sincero agradecimento.

BIOGRAFIA

Carlos Alberto Matias de Abreu Júnior nasceu em Tupaciguara (MG), em 1 de março de 1996. Graduou-se em Engenharia de Agrimensura e Cartográfica pela Universidade Federal de Uberlândia em 2020, ingressando no Programa de Pós-Graduação em Agricultura e Informações Geoespaciais no mesmo ano. Seu interesse de pesquisa parte principalmente das linhas derivadas do Sensoriamento Remoto, tais como suas aplicações na área ambiental e na agronomia. Nos últimos anos, tem se dedicado a pesquisas voltadas para a detecção de doenças e para a estimativa da produtividade de culturas agrícolas.

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 CONTEXTUALIZAÇÃO DO TEMA.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Desenvolvimento Fenológico Da Cultura Do Milho.....	3
2.2 Algoritmos De Aprendizado De Máquina.....	5
2.3 Sensoriamento Remoto Da Vegetação.....	7
2.4 Índices De Vegetação	8
2.5 Métricas De Validação Utilizadas.....	12
3 CAPÍTULO 1: DEFINIÇÃO DO ESTÁDIO FENOLÓGICO IDEAL PARA ESTIMATIVA DA PRODUTIVIDADE DO MILHO A PARTIR DE IMAGENS MULTIESPECTRAIS	13
RESUMO	13
3.1 Introdução.....	15
3.2 Material e Métodos.....	16
3.2.1 Área de estudo	16
3.2.2 Metodologia utilizada para o processamento dos dados.....	17
3.2.3 Caracterização da área experimental e produtividade.....	18
3.2.4 Aquisição dos dados multiespectrais.....	19
3.2.5 Processamento dos dados multiespectrais.....	20
3.2.6 Geração dos modelos de previsão e controle de qualidade.....	21
3.2.7 Mapa de discrepância entre valores de produtividade observado em campo e estimado pelo modelo.....	24
3.3 Resultados E Discussões	25
3.3.1 Resultados da análise exploratória da variável produtividade	25
3.3.2 Métricas de validação entre produtividade e os estádios fenológicos e seleção do melhor modelo para estimativa.....	26
3.3.3 Espacialização da variável estimada	31
3.4 Considerações Finais.....	33
4 CAPÍTULO 2: ESTIMATIVA DE PRODUTIVIDADE NA CULTURA DO MILHO EM ESCALA LOCAL A PARTIR DE MODELOS MULTIESPECTRAIS BASEADOS EM ALGORITMOS DE APRENDIZADO DE MÁQUINA	34
RESUMO	34
4.1 Introdução.....	36
4.2 Material E Método	37
4.2.1 Área de estudo	37
4.2.2 Metodologia utilizada para o processamento dos dados	38
4.2.3 Delimitação do experimento	39
4.2.4 Filtragem e remoção de outliers	39

<i>4.2.5 Aquisição de dados multiespectrais</i>	<i>39</i>
<i>4.2.6 Processamento dos dados multiespectrais</i>	<i>40</i>
<i>4.2.7 Análise descritiva e exploratória da produtividade</i>	<i>40</i>
<i>4.2.8 Geração de modelos de predição e controle de qualidade</i>	<i>40</i>
<i>4.2.9 Comparação entre os valores de produtividade mensurados em campo e os estimados in situ</i>	<i>42</i>
4.3 Resultados E Discussões.....	42
<i>4.3.1 Resultados da Análise exploratória da variável produtividade.....</i>	<i>42</i>
<i>4.3.2 Índices/bandas escolhidos para geração dos modelos e análise dos modelos individuais.....</i>	<i>44</i>
<i>4.3.3 Análise da diferença entre os valores mensurados e estimados pelos modelos</i>	<i>48</i>
4.4 Considerações Finais.....	50
REFERÊNCIAS	51

RESUMO

ABREU JÚNIOR, CARLOS ALBERTO MATIAS DE. **Modelos espectrais baseados em algoritmos de aprendizado de máquina para estimativa da produtividade do milho**. 2022, 61p. Dissertação (Mestrado em Agricultura e Informações Geoespaciais) – Universidade Federal de Uberlândia, Campus Monte Carmelo, Minas Gerais, Brasil.¹

A estimativa de produtividade é um parâmetro agrônomo importante para auxiliar o mercado nacional e internacional no que tange a demanda, capacidade de transporte e armazenamento de produtos agrícolas. Nesse sentido, para que a produtividade continue a ser monitorada, torna-se necessário a criação de métodos que aprimorem o mapeamento dessa variável em campo. Nesse contexto, esta pesquisa apresenta como objetivo a definição do estágio fenológico ideal e geração de modelos para estimativa de produtividade em escala local, ou seja, utilizando dados de áreas de estudo distintas para treinamento e validação do modelo. Para isso, foram utilizadas imagens multiespectrais, advindas de diferentes sensores orbitais, juntamente com índices de vegetação derivados das bandas de cada satélite, e modelos gerados por meio de algoritmos de aprendizado de máquina, sendo eles, Máquinas de Vetores de Suporte (Support Vector Machine - SVM), Redes Neurais (Neural Net - NN), Florestas Aleatórias (Random Forest - RF) e Árvores Aleatórias (Random Trees - RT). No Capítulo 1, para a definição do estágio fenológico ideal, foram utilizadas imagens do satélite Planet. Já no Capítulo 2, para a geração de modelos de estimativa de produtividade, foi utilizada uma imagem do satélite Sentinel 2. Foram utilizadas como métricas avaliativas a raiz do erro médio quadrático em porcentagem (Root Mean Square Error porcentagem - RMSE%) e o Erro Absoluto Médio em porcentagem (Mean Absolute Percentagem Error - MAPE) para avaliar a acurácia e tendência das estimativas de produtividade. Pela análise da definição de um estágio fenológico ideal, concluiu-se que o modelo espectral baseado em imagem da fase fenológica reprodutiva R2 foi o que obteve o melhor RMSE% de 9,17% e o segundo melhor MAPE de 7,07%. Pela análise da geração de modelos de estimativa em escala local foi possível estimar a produtividade com um RMSE% e MAPE de 20,97% e 19,19%, respectivamente.

Palavras-Chave: Cultura do milho; produtividade; imagens multiespectrais; algoritmos de aprendizado de máquina; modelos espectrais de produtividade.

¹ Orientador: George Deroco Martins

ABSTRACT

ABREU JÚNIOR, CARLOS ALBERTO MATIAS DE. **Spectral models based on machine learning algorithms to estimate corn yield.** 2022, 61p. Dissertation (Master's Degree in Agriculture and Geospatial Information) – Federal University of Uberlândia, Campus Monte Carmelo, Minas Gerais, Brazil.¹

Estimated yield is an important agronomic parameter for the domestic and international market in the demand, transport capacity and storage of agricultural products. In this respect, continuous yield monitoring requires the creation of methods to enhance mapping of this variable in the field. As such, this study aimed to determine the ideal phenological stage to estimate yield and generate models for this purpose on a local scale, that is, using data from different study areas to train and validate the model. To that end, we used multispectral images from different orbital sensors combined with vegetation indices from the reflectance bands of each satellite and models generated by the following machine learning algorithms: support vector machine (SVM), neural network (NN), random forest (RF) and random trees (RT). In Chapter 1, Planet satellite images were used to determine the ideal phenological stage and in Chapter 2, yield estimate models were generated via a Sentinel 2 satellite image. The root mean square error percentage (RMSE%) and mean absolute percentage error (MAPE) were used as evaluation metrics to assess the accuracy and trend of the yield estimates. Based on analysis to determine the ideal phenological stage, it was concluded that the spectral model based on the image of phenological stage R2 obtained the best RMSE% (9.17%) and second best MAPE (7.07%). The local-scale estimation models generated made it possible to estimate yield with an RMSE% and MAPE of 20.97 and 19.19%, respectively.

Keywords: Corn crops; yield; multispectral images; machine learning algorithms; spectral yield models.

¹ Advisor: George Deroco Martins

1 CONTEXTUALIZAÇÃO DO TEMA

Considerando a importância do milho em contexto global, faz-se necessário estimar a sua produção, a fim de gerar informações sobre a demanda mundial (ERENSTEIN et al., 2021). A produtividade é um parâmetro agrônomico importante para auxiliar o mercado nacional e internacional no que tange a demanda, capacidade de transporte e armazenamento de produtos agrícolas. Por meio da previsão das safras das culturas, é possível a comercialização antes da colheita, além da estimativa de preços, considerando a quantidade de produto que será gerado (BERTOLIN et al., 2017). Outra característica que pode ser obtida por meio do mapeamento da produtividade é a identificação de áreas com menor potencial produtivo, o que pode eventualmente estar associado a deficiências nutricionais das plantas ou do solo para futuras lavouras.

Uma forma de estimar a produtividade é por meio da combinação dos parâmetros biofísicos amostrados em campo (FIGUEIRAS et al., 2017) com as informações obtidas por técnicas de sensoriamento remoto (SHANAHAN, 2001). Essas técnicas consistem na análise da resposta espectral da cultura e dos alvos presentes na superfície, que são obtidos por sensores multiespectrais acoplados em satélites, aeronaves remotamente pilotadas ou maquinários de agrícola (JENSEN, 2009).

A junção do sensoriamento remoto com a agricultura, no que se refere à estimativa de parâmetros agrônomicos, não é algo novo. Foster (1951) pode ser considerado o primeiro a tentar relacionar o sensoriamento remoto com a agronomia. Em seu estudo, o autor propõe a interpretação dos dados advindos de fotos aéreas por um geógrafo. Um pouco mais tarde, Brunnschweiler (1957) introduz efetivamente o conceito de fotointerpretação especificamente na agricultura. Em seu estudo, o autor relata como é possível identificar o desenvolvimento de uma cultura em função das mudanças sazonais sofridas. Posteriormente, Dill (1959) relata a possibilidade da mensuração da área total utilizada ao longo do tempo, um processo que poderia ser feito de forma rápida e relativamente simples, com o subsídio de poucos profissionais, e que poderia ser usado na criação de dados sobre o uso de terras agrícolas e as suas conseqüentes influências econômicas.

Goodman (1959) talvez tenha desenvolvido o estudo mais importante desse período, no qual o autor desenvolve um método baseado na análise de textura e tonalidade das imagens para o aumento da velocidade e acurácia do mapeamento das plantações. Os anos de 1959 a 1961 marcam a chegada das imagens coloridas e das imagens oriundas do infravermelho. É a partir

desse período que uma série de estudos visando análises sobre aspectos qualitativos e quantitativos das plantações surgem, tais como detecção de doenças, crescimento vegetativo e, finalmente, estimativa de produtividade (GATES, 1965; ESTES, 1966; PHILPOTTS; WALLEN, 1969; RYERSON, 1970). Nesse contexto, surgem os índices de vegetação (IV). O primeiro índice a se popularizar é o Normalized Difference Vegetation Index (NDVI), proposto por Rouse et al. (1974). Os IV são significativos para o estudo da vegetação, uma vez que a maioria desses modelos visa caracterizar a vegetação nas imagens. Os IV são comumente utilizados na geração de modelos para estimativa e produtividade, bem como na geração de outros parâmetros biofísicos em culturas agrícolas.

Diante do desenvolvimento histórico apresentado, ressalta-se que, nessa época, os modelos de estimativa de parâmetros agrônômicos utilizavam metodologias baseadas em regressão linear ou múltipla das variáveis estudadas. Embora essas metodologias tenham alcançado um sucesso inegável, com o passar do tempo e com o aumento da quantidade de dados para ser processado, começam a surgir os primeiros problemas relacionados à recolha de muitos dados. Diante das altas demandas de processamento, as técnicas de aprendizado de máquina começam a surgir como eventuais soluções para o processamento de volumes grandes de dados.

As primeiras aplicações voltadas a pequenos processos de automação na agricultura têm registro de mais de cinquenta anos (DOKIC et al., 2020). Entretanto, é inegável que a limitação do poder computacional, à época, era uma barreira difícil de ser transposta. Nesse contexto, a inteligência artificial começou a ser aplicada na agricultura com maior frequência no começo dos anos 2000 (DOKIC et al., 2020). Para entender a variabilidade por trás dos modelos de estimativa de produtividade, Lobell et al. (2005) utilizam imagens multiespectrais combinadas com variáveis obtidas *in situ* e árvores de regressão (*Regression Trees* - RT). No trabalho, os autores destacam que os melhores resultados são aqueles advindos dos modelos de RT, tendo em vista que foram capazes de identificar restrições de rendimento regionalmente significativas, que impactam na estimativa de produtividade. Panda, Ames e Panigrahi (2010) utilizam índices de vegetação combinados com técnicas de rede neural artificial (Artificial Neural Network - ANN) para estimativa de produtividade no milho. No trabalho, os autores testam diversos índices e técnicas de transformação de imagem, e conseguem gerar modelos com coeficiente de determinação de 0,72. Pantazi et al. (2016) utilizam técnicas de aprendizado de máquina e imagens multiespectrais para estimar produtividade numa cultura de trigo. No trabalho, os autores destacam que a Rede Supervisionada de Kohonen (*Supervised Kohonen Networks* - SVK) como o melhor resultado dentre os modelos analisados, conseguindo uma

acurácia média de 81,65%. Utilizando imagens do sensor MODIS, Sakamoto (2020) consegue estimar a produtividade na cultura do milho mediante um modelo que engloba variáveis ambientais e algoritmos de Florestas Aleatórias (*Random Forest* - RF). No trabalho, o autor consegue um erro médio quadrático de 0,539 t/ha.

Apesar do relativo sucesso na aplicação desses métodos na estimativa de produtividade na cultura do milho, ainda existem lacunas a serem preenchidas. Em linhas gerais, existe uma escassez de trabalhos que abordam: (1) aspectos relativos ao período ideal para aquisição de imagens multiespectrais, ou, ainda, (2) a aplicação dos modelos gerados em outras áreas de estudo, além daquela sobre o qual ele foi gerado. Essas questões se fazem pertinentes, uma vez que culturas anuais são caracterizadas pelo curto período de desenvolvimento em campo. Dessa forma, metodologias que auxiliem tomadas de decisão, de forma precoce, podem garantir um aumento significativo na produtividade.

Por conseguinte, esta dissertação é composta por dois capítulos, cada qual com experimentos diferentes. O Capítulo I analisa qual o estado fenológico mais propício para a obtenção de imagens multiespectrais para a estimativa de produtividade da cultura do milho, tendo em vista a correlação dos valores de produtividade com os períodos de desenvolvimento fenológico da cultura do milho. Os resultados encontrados nesse experimento foram utilizados como ponto de partida para o segundo capítulo. Portanto, a partir da obtenção de um estágio fenológico ideal para a estimativa de produtividade, buscou-se realizar testes em diferentes áreas de estudo. O Capítulo II, então, baseia-se na construção de um modelo mais robusto para a estimativa de produtividade da cultura do milho mediante imagens multiespectrais advindas do estágio fenológico determinado no primeiro capítulo.

2 REFERENCIAL TEÓRICO

2.1 DESENVOLVIMENTO FENOLÓGICO DA CULTURA DO MILHO

O ciclo fenológico da cultura do milho pode ser dividido em 11 estádios de desenvolvimento distintos (FANCELLI, 1986). Para elaboração da escala, o autor leva em consideração diversos elementos inerentes à cultura, tais como fisiológicos, entomológicos, fitotécnicos e fitopatológicos, além de elementos externos, como o climatológico. O desenvolvimento da planta é dividido em dois grupos, sendo eles os estádios vegetativos, indicados pela letra V, e os estádios reprodutivos, indicados pela letra R. Assim, o primeiro estágio fenológico é o de emergência (VE), que é originado mediante condições normais de

campo, quando a semente absorve água, incha e começa o seu desenvolvimento. Nesse estágio, ocorre o desenvolvimento inicial da radícula e posteriormente do coleóptilo, que é empurrado para superfície pelo mesocótilo durante o seu processo de alongação. Em condições normais, o período de emergência dura de 4 a 5 dias, variando principalmente em função de temperatura e umidade. Após o surgimento da plântula, tem-se duas folhas completamente expandidas, que permitirão o início do processo fotossintético (MAGALHÃES; DURÃES, 2006).

Passada em média duas semanas após o plantio, ocorrerá o estágio fisiológico V4, caracterizado por possuir 4 folhas completamente desenvolvidas. Nesse período, ocorre a preparação para o começo da diferenciação do meristema apical. Até esse momento, o desenvolvimento da planta sofre influência direta da temperatura. Em casos de geada, a produção será reduzida. Com o desenvolvimento contínuo da planta, entre 30 e 35 dias começa o estágio V8, quando a planta apresenta 8 folhas desenvolvidas. Além disso, ocorre o crescimento do colmo, tanto em comprimento quanto em diâmetro, que servirá de suporte para o armazenamento de sólidos solúveis, utilizados na etapa de formação dos grãos (SOLOGUREM, 2015).

Na sexta e sétima semana após o período de emergência, a planta normalmente avança para o estágio V12. Ao atingir esse nível de desenvolvimento, a planta já se encontra com aproximadamente 85% a 90% da área foliar definida. Outras características que definem o estágio são: elevada taxa de crescimento do colmo, pendão e espiga superior, perda de duas a quatro folhas basais. Durante o período de oito a nove semanas após a emergência, inicia-se o estágio pendramento, VT. Esse estágio de desenvolvimento é associado ao surgimento parcial do pendão e ao crescimento intenso do estilos-estigma (CIAMPITTI; ELMORE; LAUER, 2016).

Após o desenvolvimento dos estádios vegetativos, durante o período de nove a dez semanas após a emergência das plantas, inicia-se o desenvolvimento dos grãos, que são definidos pelos seus estádios reprodutivos. Em R1, são finalizados o crescimento do colmo e dos internódios. Nesse estágio, os estilo-estigma continua o seu crescimento, a fim de que eles sejam polinizados. Desse modo, o processo de fecundação do óvulo pode continuar. O vento é responsável por realizar a polinização do milho, na maioria dos casos, e pode superar distâncias de até 500 m (MAGALHÃES; DURÃES, 2006).

O estágio R2 tem seu período de incidência entre 12 a 18 dias após a fecundação. A principal característica desse estágio é o acúmulo de açúcares solúveis no endosperma do grão, o que auxilia no aumento da massa do grão. Nesse período, a espiga de milho fica próxima do tamanho final que irá atingir. No estágio R3, os grãos intensificam o ganho de peso e a

quantidade de amido cresce consideravelmente. O período começa em média 20 a 32 dias após o surgimento dos estilos-estigmas CIAMPITTI; ELMORE; LAUER, 2016). No estágio R4, inicia-se o processo de formação dos “dentes”, que são concavidades presente na parte superior de cada grão. Nesse estágio, os grãos estão em transição, entre o estado pastoso e o farináceo. No estágio R5, a umidade da planta é reduzida significativamente, caindo para até 55%. Além disso, ocorre uma queda no acúmulo de substâncias orgânicas pelo grão. Esse estágio é associado à maturação morfológica das sementes. Já o estágio fisiológico R6 é caracterizado pela maturidade fisiológica dos grãos. Nesse período, o processo de senescência nas folhas torna-se mais visível, uma vez que elas ganham a aparência de “mais seca”, o que acontece devido à degradação de clorofila. Esse estágio é comumente destinado à colheita, uma vez que os grãos se apresentam independentes da planta mãe. De forma geral, os principais estádios fenológicos são mostrados na Figura 1 (SOLOGUREM, 2015).

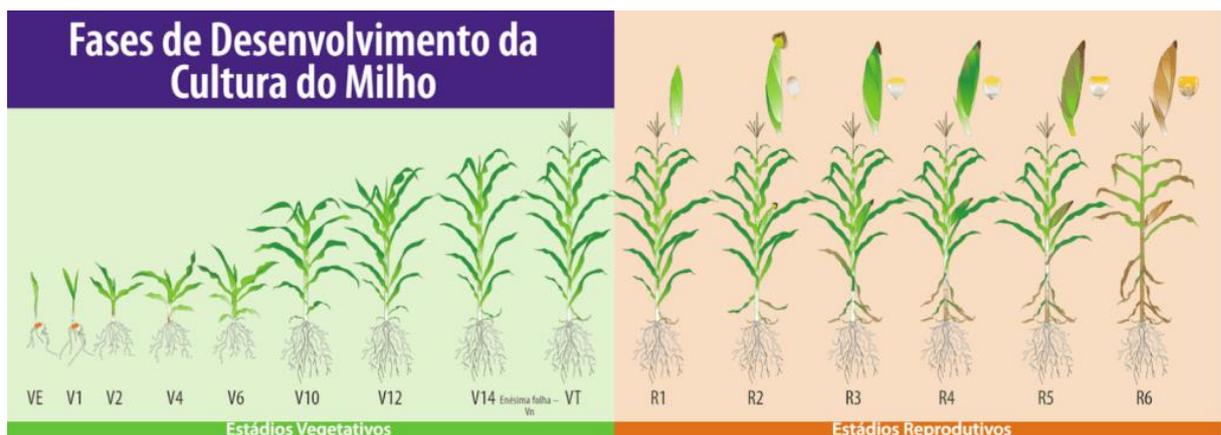


FIGURA 1. Desenvolvimento fenológico da cultura do milho por Ciampitti; Elmore; Lauer (2016)

2.2 ALGORÍTMOS DE APRENDIZADO DE MÁQUINA

Na definição dos algoritmos para a estimativa de produtividade, são utilizadas duas metodologias, as quais baseadas em métodos de regressão não paramétricos, como Random Tree (RT), Random Forest (RF), Neural Net e (NN) Support Vector Machine (SVM), e paramétrico, modelos de Linear Regression (LR). Essa etapa foi desenvolvida no software Weka 3.9.5.

De forma simples, uma rede neural artificial pode ser descrita como um conjunto de neurônios de entrada que serão processados por camadas ocultas dando resultado a neurônios

de saída. Esses dados são ligados por conexões, cada qual tendo um peso diferente. Segundo Wang (2003), a saída h_i de um neurônio i numa camada oculta pode ser definida como:

$$h_i = \sigma(\sum_{j=1}^N V_{ij} x_j + T_i^{hid}) \quad (1)$$

No qual $\sigma()$ é a função de ativação, N é o número de neurônios de entrada, V_{ij} são os pesos, x_j entradas dos neurônios de entrada e T_i^{hid} o limiar das camadas ocultas. Normalmente, as redes utilizam a função sigmoide para fazer a ativação, entretanto, existem outras funções que podem desempenhar o mesmo papel, tais como a tangente e tangente hiperbólica. O software Weka se baseia tradicionalmente na função sigmoide. Segundo Wang (2003), a função sigmoide pode ser definida como:

$$\sigma(u) = \frac{1}{1 + \exp(-u)} \quad (2)$$

No processamento por Redes Neurais Artificiais, é feito, inicialmente, o treinamento do algoritmo utilizando os dados de entrada, o que se dá por meio de um processo repetitivo de iterações. Após o treinamento da rede, a classificação de todos os padrões usando o conjunto de pesos ajustados é feita em associação com um dado padrão como pertencente a uma determinada classe, quando a saída para essa classe for alta e baixa para as demais classes (GONZALEZ; WOODS, 1993). Em linhas gerais, o critério de decisão adotado estabelece que, para um dado padrão de entrada, o elemento de processamento da camada de saída (associado a uma classe específica) que produzir o maior resultado numérico definirá a classe a qual o padrão será atribuído. Dessa forma, será avaliada a acurácia da metodologia para a geração de parâmetros agronômicos.

O software Weka utiliza diversas variáveis de entrada, dentre as quais destacam-se as seguintes:

- Taxa de aprendizado (LearningRate): A quantidade de pesos que serão atualizados;
- Momento (Momentum): O momento em que os pesos serão atualizados;
- Número de camadas ocultas (HiddenLayers): O número de camadas ocultas utilizadas na rede.

Nas Máquinas de Vetor Suporte, a estratégia utilizada concentra-se no uso intrínseco de um classificador binário. A priori, são construídos hiperplanos de separação linear para classificar instâncias dadas. Dessa forma, o algoritmo pode ser gradativamente melhorado à medida que são feitas transformações dos espaços característicos originais em espaços

característicos de dimensão maior. Desse modo, o algoritmo pode ser amplamente usado para a regressão de dados em larga escala. No software Weka, foram priorizadas as seguintes variáveis de entrada:

- Kernel: A função kernel utilizada;
- Filter type: Determina o como e se os dados serão transformados.

As RF são um conjunto de técnicas de aprendizado que utilizam múltiplas árvores de decisão. Assim, é facilitada a manipulação de dados de diferentes origens, tais como aqueles advindos de imagens e os mensurados em campo, uma vez que o algoritmo não exige que os dados apresentem distribuição estatística normal. Devido a essa facilidade, o algoritmo vem sendo amplamente utilizado na agricultura para a resolução de diversos problemas, como é o caso da estimativa de produtividade (ARAB et al, 2021; ISLAM et al., 2021; KHANAL et al., 2021).

Desse modo, as Florestas Aleatórias são um conjunto de técnicas de aprendizado que utilizam múltiplas árvores de decisão. Estas são usadas para fazer um conjunto de validação, a partir do qual serão geradas previsões estatísticas que são baseadas em conjuntos de variáveis independentes. Por conseguinte, torna-se extremamente fácil o manejo de dados multiplataforma, tais como aqueles advindos de imagens e os mensurados em campo, uma vez que o algoritmo não exige que os dados apresentem distribuição estatística normal.

2.3 SENSORIAMENTO REMOTO DA VEGETAÇÃO

O sensoriamento remoto tem como uma de suas aplicações a vegetação. Nessa subdivisão ele estuda a propagação da REM nesses ambientes, que podem ser desde florestas e matas fechadas, a diversas culturas de plantio. Embora existam inúmeras espécies e tipos de folhas, em todas elas a propagação da REM é dada através do seu comprimento de onda. Isso acontece devido a alguns fatores relacionados à fisiologia das folhas (PONZONI; SHIMABUKURO; KUPLICH, 2012).

Os pigmentos fotossintetizantes possuem uma capacidade maior de absorção dos comprimentos de onda do visível. Isso faz com que a REM dessa região seja refletida com maior intensidade, garantindo assim uma melhor reflectância da vegetação nessa região do espectro eletromagnético. Entretanto, existem alguns fatores químicos que podem alterar a absorção de REM numa folha. Um deles é a presença de água, que tem facilidade para absorver

radiações e, conseqüentemente, reduzir a fluxo de energia refletida (PONZONI; SHIMABUKURO; KUPLICH, 2012). Na Figura 2, é apresentada a curva de reflectância de uma folha.

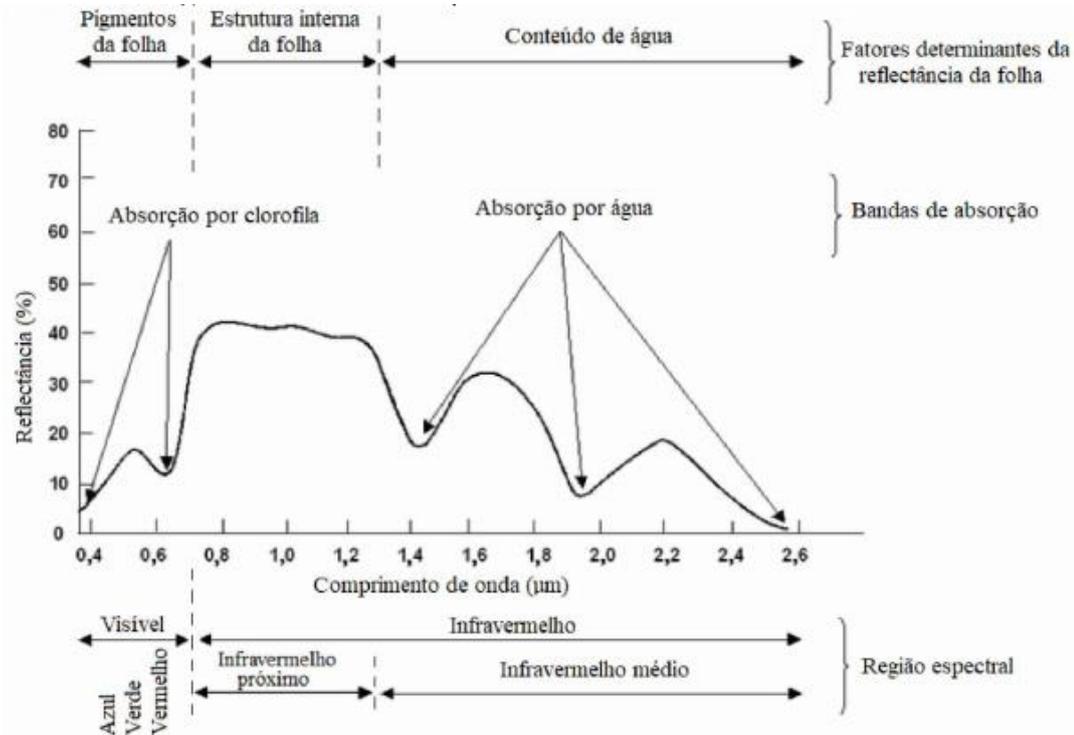


FIGURA 2. Reflectância de uma folha saudável

Os componentes internos da folha responsáveis pela absorção na região do visível são encontrados nos cloroplastos. Segundo Ponzoni, Shimabukuro, Kuplich (2012), a absorção é dividida da seguinte forma: clorofila (65%), carotenos (6%) e xantofilas (29%). Esses valores podem se diferenciar entre espécies. A clorofila é o único componente capaz de absorver a radiação proveniente da região do vermelho. No infravermelho próximo, a absorção é pequena. Parte da REM é espalhada internamente, variando de acordo com as características fisiológicas da folha.

2.4 ÍNDICES DE VEGETAÇÃO

Na literatura, observa-se que os índices de multiespectrais de vegetação compõem – em uma parte considerável – os modelos espectrais de estimativa de produtividade (RAMOS et al., 2020; AGHIGHI et al., 2018). Os índices de vegetação são normalmente utilizados para realçar características específicas da vegetação. A sua variabilidade também é associada às bandas das quais aquele índice é derivado.

Nesta pesquisa, calculou-se os índices de vegetação (Tabela 1) a partir das bandas originais do satélite Planet e Sentinel 2. O cálculo dos IV foram feitos no software ENVI 5.1. Vale ressaltar que os índices de vegetação dispostos na Tabela 1 foram selecionados porque ressaltam características específicas da vegetação, tais como biomassa, vigor vegetativo, pigmentação e índice de área foliar.

TABELA. 1 Equações e referências para cálculos dos índices de vegetação derivados das bandas originais dos satélites Planet e Sentinel 2

Índice	Equação	Referência	Contribuição
Normalized Difference Vegetation Index - NDVI	$\frac{(\rho_{ni\ r} - \rho_{red})}{(\rho_{ni\ r} + \rho_{red})}$	Rouse et al. (1974)	Alta correlação com a produtividade.
Green Normalized Difference Vegetation Index - GNDVI	$\frac{(\rho_{ni\ r} - \rho_{green})}{(\rho_{ni\ r} + \rho_{green})}$	Gitelson et al. (1996)	Sensibilidade à concentração de clorofila.
Ratio vegetation index - RVI	$\frac{\rho_{red}}{\rho_{ni\ r}}$	Jordan (1969); Pearson, Miller (1972); Huete (1988)	Correlação com a densidade foliar da cultura.
Soil adjusted vegetation index - SAVI	$\frac{(\rho_{ni\ r} - \rho_{red})}{(\rho_{ni\ r} + \rho_{red} + C) * (1 + C)}$	Huete (1988)	Minimiza o efeito de brilho do solo.
Modified soil adjusted vegetation index - MSAVI	$\frac{2 * \rho_{ni\ r} + 1 - \sqrt{(2 * \rho_{ni\ r})^2 - 8 * (\rho_{ni\ r} - \rho_{red})}}{2}$	Qi et al. (1994)	Minimiza o efeito de solo do índice SAVI.
Optimized soil adjusted vegetation index - OSAVI	$\frac{(\rho_{ni\ r} - \rho_{red})}{(\rho_{ni\ r} + \rho_{red} + 0,16)}$	Rondeaux et al. (1996)	Análise sobre estádios vegetativos.
Enhanced vegetation index - EVI	$\frac{\rho_{green} * (\rho_{ni\ r} - \rho_{red})}{(\rho_{ni\ r} + C_1 * \rho_{red} * C_2 * \rho_{blue} + X)}$	Huete et al. (2002)	Realça a vegetação com menor influência atmosférica.
Triangular vegetation index - TVI	$\frac{\rho_{green} * (\rho_{ni\ r} - \rho_{red})}{\sqrt{NDVI + 0,5}}$	Broge, Leblanc (2000)	Sensibilidade ao índice de área foliar da cultura.
Second modified triangular vegetation index - MTVI2	$\frac{1,5[2,5(\rho_{ni\ r} - \rho_{blue}) - 2,5(\rho_{red} - \rho_{green})]}{\sqrt{[(2\rho_{ni\ r} + 1)^2 - 6\rho_{ni\ r} - 5\sqrt{\rho_{ni\ r} - 0,5}]}}$	Haboudane et al. (2004)	Sensibilidade ao índice de área foliar da cultura.
Chlorophyll vegetation index - CVI	$\frac{\rho_{ni\ r} * \rho_{red}}{\rho_{blue}^2}$	Vincini et al. (2008)	Maior sensibilidade a clorofila.

Chlorophyll index CI-G	$\frac{\rho_{nir}}{\rho_{green}} - 1$	Gitelson et al. (2003)	Auxilia na estimativa total de clorofila da planta.
Green leaf index - GLI	$\frac{(2 * \rho_{green} - \rho_{red} - \rho_{blue})}{(2 * \rho_{green} + \rho_{red} + \rho_{blue})}$	Louhaichi et al. (2001)	Intensidade de área foliar da cultura.
Triangular greenness index - TGI	$-0,5[(\lambda_{red} - \lambda_{blue})(\rho_{red} - \rho_{green}) - (\lambda_{red} - \lambda_{green})(\rho_{red} - \rho_{blue})]$	Hunt et al. (2011)	Realça a vegetação com baixa sensibilidade a efeitos atmosféricos.
Normalized green red difference index - NGRDI	$\frac{(\rho_{green} - \rho_{red})}{(\rho_{green} + \rho_{red})}$	Tucker et al. (1979)	Correlação com biomassa da cultura.
Normalized Difference Red Edge Index - NDRE	$\frac{(\rho_{nir} - \rho_{re1})}{(\rho_{nir} + \rho_{re1})}$	Birth, McVey (1968)	Auxilia na estimativa total de clorofila da planta.
Chlorophyll index - red edge - CI-RE	$\frac{\rho_{nir}}{\rho_{re1}} - 1$	Gitelson et al. (2003)	Auxilia na estimativa total de clorofila da planta.
MERIS terrestrial chlorophyll index - MTCI	$\frac{(\rho_{nir} - \rho_{730})}{(\rho_{re2} - \rho_{red})}$	Dash, Curran (2004)	Auxilia na estimativa total de clorofila da planta.
Modified Chlorophyll Absorption in Reflectance Index - MCARI	$\left[(\rho_{re1} - \rho_{red}) - 0,2 * (\rho_{re1} - \rho_{green}) \right] * \left(\frac{\rho_{re1}}{\rho_{red}} \right)$	Daughtry et al. (2000)	Auxilia na estimativa da concentração de clorofila da planta.
Transformed Chlorophyll Absorption in Reflectance Index - TCARI	$3 * \left[(\rho_{re1} - \rho_{red}) - 0,2 * (\rho_{re1} - \rho_{green}) \right] * \left(\frac{\rho_{re1}}{\rho_{red}} \right)$	Haboudane et al. (2002)	Auxilia na estimativa da concentração de clorofila da planta.

2.5 MÉTRICAS DE VALIDAÇÃO UTILIZADAS

Para validação e análise da acurácia dos modelos, foram utilizadas duas métricas: Mean Absolute Percentagem Error (Mape) (Equação 3) e Root Mean Square Error percentagem (RMSE%) (Equação 4).

$$MAPE = \frac{\sum |\hat{y}_i - y_i|}{n} * 100 \quad (3)$$

Na Equação 1, tem-se: \hat{y}_i é o valor predito, y_i é o valor observado em campo, n é o número total de elementos.

$$RMSE (\%) = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} * 100} \quad (4)$$

Na Equação 2, tem-se: \hat{y}_i são os valores preditos, y_i são os valores mensurados em campo, n o número total de observações.

3 CAPÍTULO 1: Definição do Estádio Fenológico Ideal para Estimativa da Produtividade do Milho a partir de Imagens Multiespectrais

RESUMO

Os modelos espectrais baseados em imagens são capazes de ajudar a estimar a produtividade do milho. As lavouras de milho durante as fases fenológicas vegetativa e reprodutiva passam por alterações causadas por estresse biótico e abiótico. Essas variações podem ser quantificadas por modelos espectrais que são ferramentas que auxiliam os produtores no manejo da lavoura. Entretanto, definir o momento correto para a tomada das imagens ainda permanece um desafio. Dessa forma, a hipótese desta pesquisa baseia-se na possibilidade de estimar a produtividade do milho a partir de imagens multiespectrais, considerando-se a definição do momento mais propício para a detecção das variações causadas pelos diferentes estádios fenológicos. Nesta pesquisa, foram criados modelos espectrais de estimativa de produtividade, em diferentes estádios fenológicos do milho, baseados em algoritmo de aprendizado de máquina (Neural Net, Support Vector Machine, Random Forest, Random Tree). Como dados de entrada dos modelos, considerou-se como variáveis preditoras bandas multiespectrais e índices de vegetação derivados do satélite Planet. Utilizou-se Raiz do erro médio quadrático em porcentagem (Root Mean Square Error porcentagem - RMSE%) e o Erro médio absoluto em porcentagem (Mean Absolute Percentagem Error - MAPE) para avaliar a acurácia e a tendência das estimativas de produtividade. Concluiu-se que, para determinar modelos espectrais baseados em imagens, a fase fenológica reprodutiva R2 foi o que obteve o melhor RMSE%, de 9,17%, e o segundo melhor MAPE, de 7,07%.

Palavras-chave: Cultura do milho; estádios fenológicos; algoritmos de aprendizado de máquina; modelos espectrais.

3 CHAPTER 1: Defining the Ideal Phenological Stage for Estimation of Corn Yield from Multispectral Images

ABSTRACT

Image-based spectral models can help estimate corn yield. During the vegetative and reproductive stages, corn crops undergo changes caused by biotic and abiotic stress. These variations can be quantified by spectral models, tools that help producers manage crops. However, determining the ideal time to perform spectral imaging remains a challenge. As such, the hypothesis of this study is based on the possibility of estimating corn yield using spectral images by determining the optimal time to detect the variations caused by the different phenological stages. Spectral models were created to estimate yield at different corn phenological stages based on machine learning algorithms (neural network, support vector machine, random forest, random tree). Multispectral bands and vegetation indices derived from the Planet satellite were considered predictive variables and used as input data for the models. The root mean square error percentage (RMSE%) and mean absolute percentage error (MAPE) were used to assess the accuracy and trend of the yield estimates. It was concluded that, in order to generate image-based spectral models, phenological stage R2 obtained the best RMSE% (9.17%) and second best MAPE (7.07%).

Keywords: corn crops; phenological stages; machine learning algorithms; spectral models

3.1 INTRODUÇÃO

O uso de imagens multiespectrais para estimar produtividade cumpre uma série de requisitos e critérios metodológicos. O período de tomada de imagens e a definição de bandas espectrais e índices de vegetação ideais para a construção dos modelos de predição estão condicionados às características espectrais de distintos estádios fenológicos da cultura. Isso ocorre porque cada lavoura apresenta particularidades, variando em tipo de manejo aplicado em campo e mudanças estruturais inerentes a cada tipo de cultura, como o tempo de cultivo exigido por cada cultura (PÁDUA et al., 2018). As anuais e perenes, por exemplo, apresentam distintos estádios fenológicos e respostas espectrais completamente desiguais, decorrentes de fatores como geometria da copa do dossel, biomassa, índice de área foliar e senescência das folhas (JENSEN, 2009).

No período entre a semeadura e a colheita, a cultura do milho apresenta diferenças espectrais significativas em função de propriedades nutricionais e estruturais características de cada um dos seus estádios fenológicos de desenvolvimento vegetal. Nas fases específicas, o comportamento espectral apresenta variações no intervalo do comprimento do azul ao infravermelho médio. Em contrapartida, nas culturas perenes, os modelos espectrais de estimativa da produção da lavoura não podem ser baseados em imagens multiespectrais tomadas no dia da colheita, uma vez que o alto grau de senescência da folha do milho nesse período, caracterizada por uma queda natural de clorofila, torna-se um fator que pode comprometer o registro da variabilidade espectral de bandas espectrais que terão a função de predição da produtividade.

Em trabalhos recentes, modelos espectrais baseados em aprendizado de máquina e imagens MODIS (MA et al., (2021) e Landsat 8 (AHMAD et al., 2018), tomadas por sensores embarcados em aeronaves (WAHAB et al. 2018), para a estimativa de produtividade apresentaram uma maior acurácia com uso de imagens tomadas no estágio fenológico entre R2 e R3, ou seja, no período de 12 a 18 dias após a fecundação, quando a cultura responde a um ganho significativo de biomassa. As limitações de tais modelos muitas vezes não estão relacionadas propriamente com a acurácia e a precisão da estimativa do valor absoluto da produtividade, pois décadas de estudos de sensoriamento remoto agrícola mostram que, para qualquer cultura, as bandas espectrais sempre possuem alta sensibilidade a parâmetros agrônômicos correlacionados com fatores de produção, tornando-as potenciais variáveis preditoras. Para tanto, com base em recentes pesquisas disponíveis, considera-se que os atuais modelos espectrais de predição apresentam limitações quanto aos benefícios da aplicação, uma

vez que as metodologias descritas só recomendam estimavas em estádios fenológicos avançados do desenvolvimento do milho, ou seja, fases em que intervenções e manejos na cultura surtirão pouca diferença na produtividade final.

Diante do exposto, neste trabalho, tem-se que, além do potencial de acerto do valor absoluto da produção, a real contribuição de modelos espectrais para a estimativa da produtividade está relacionada com o quão útil são para práticas de precisão no manejo da lavoura do milho, isto é, se também podem ser utilizados com antecedência, para embasar tomadas de decisões do produtor.

Neste trabalho, são apresentados modelos espectrais baseados em aprendizado de máquina capazes de estimar a produtividade do milho com graus de acurácia e precisão satisfatórios, já em estádios fenológicos iniciais. Com a aquisição de imagens ao longo de todos os estádios fenológicos, os modelos a seguir foram mais assertivos com o uso de imagens multiespectrais de alta resolução espacial (Planet) tomadas no estágio fenológico R2 da cultura, ou seja, no período de 10 a 12 semanas após a semeadura. Porém, observa-se a possibilidade de se estimar a produtividade com acurácia e precisão satisfatórias já no estágio de VE, ou seja, uma semana após a semeadura.

3.2 MATERIAL E MÉTODOS

3.2.1 Área de estudo

Para o desenvolvimento desta pesquisa, foi utilizada uma área experimental que está localizada no município de Patos de Minas. A área está inserida na mesorregião do Triângulo Mineiro e Alto Paranaíba, Minas Gerais, Brasil (Figura 3). A área de estudo (21,56 ha) possui latossolo vermelho, altitude média de 938 m, relevo plano e ondulado e precipitação anual média de 950 mm. A mesorregião do Triângulo Mineiro e Alto Paranaíba é uma das principais produtoras de grãos do estado de Minas Gerais. O destaque das produções são as culturas de milho e soja, devido ao alto desenvolvimento tecnológico dessa região (BORGHI et al., 2016). Um dos fatores que impulsiona a produção é a temperatura (BAUM et al., 2019). Segundo Oliveira et al. (2020) a mesorregião do Triângulo Mineiro e do Alto Paranaíba tem clima tropical, com estações de inverno amenos e secos. As temperaturas médias da região variam de 23 °C a 28 °C durante o verão e de 16 °C a 21 °C durante o inverno. Essas temperaturas tendem

a ser favoráveis para o desenvolvimento da cultura do milho em segunda safra (BORGHI et al., 2016).

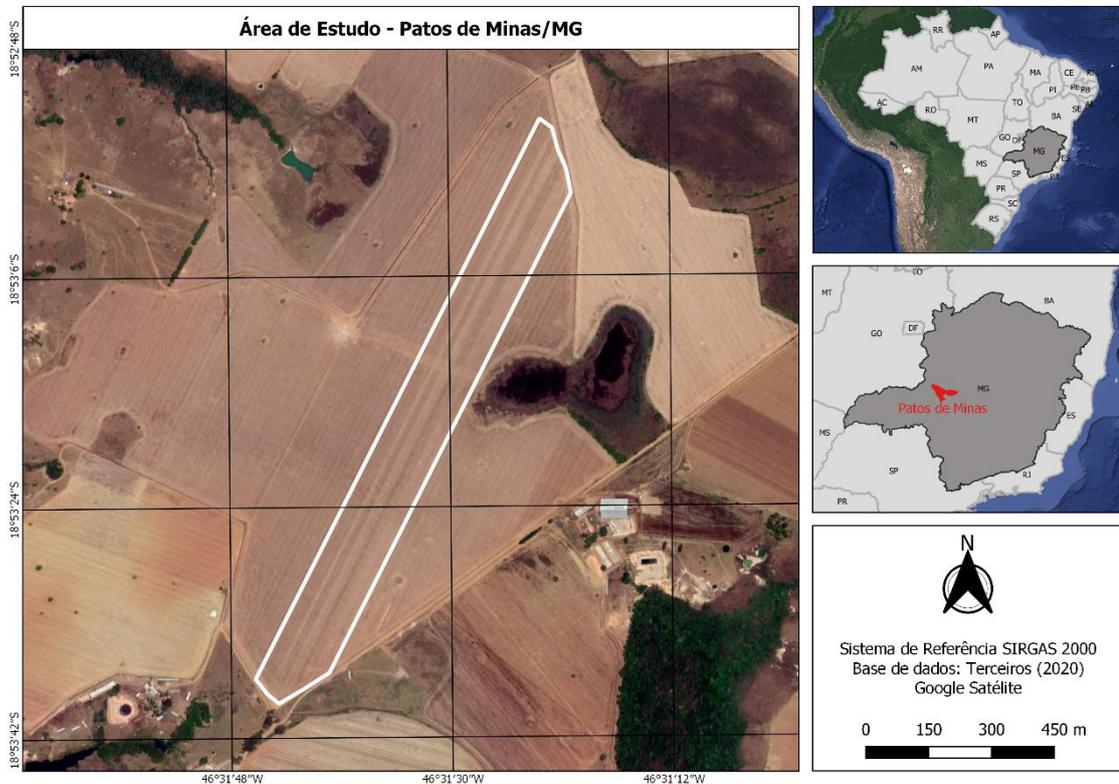


FIGURA 3. Imagem da área usada na pesquisa: contexto do município de Patos de Minas, estado de Minas Gerais, Brasil

3.2.2 Metodologia utilizada para o processamento dos dados

Inicialmente, teve-se por base Ji et al. (2021) e Sun et al. (2020) para a criar os modelos de estimativa e suas respectivas variáveis preditoras, a saber, as bandas das imagens e os índices multiespectrais, os algoritmos de regressão (aprendizado de máquina) e a variável estimada (produtividade). Em linhas gerais, a metodologia deste trabalho (Figura 4) foi desenvolvida passando pelas etapas de aquisição de dados de produtividade, tomada de imagens em diferentes estádios fenológicos, definição de algoritmos para a estimativa de produtividade, validação do método e mapeamento da distribuição espacial da produtividade. Assim, genericamente, as etapas de pesquisa foram: (I) Aquisição de dados; (II) Pré-processamento; (III) Amostragem e filtragem dos dados; (IV) Geração de modelos de predição; (V) Validação dos modelos; (VI) Confeção de mapas temáticos.

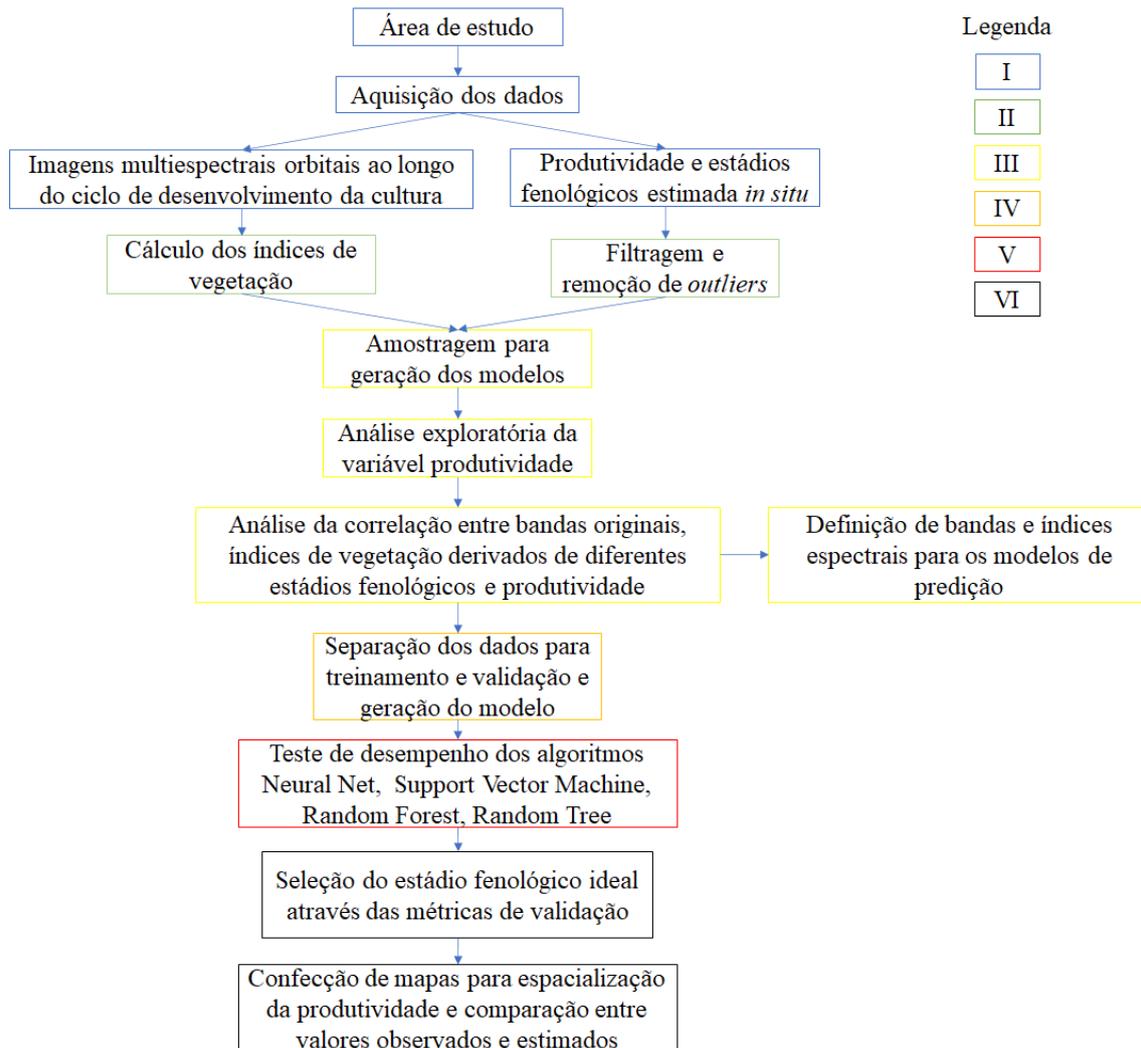


FIGURA 4. Fluxograma de atividades

3.2.3 Caracterização da área experimental e produtividade

Para análise dos dados de produtividade, foi padronizada no município de Patos de Minas (Figura 3) a área de estudo. As colheitas foram realizadas no dia 8 de agosto de 2020, no período de segunda safra, considerando o mercado nacional da cultura no Brasil. O experimento agrícola foi montado utilizando espaçamento entre linhas de 0,5 m, com distribuição de 3 plantas a cada metro. Dessa forma, o espaçamento vertical foi de aproximadamente 0,33 m. Ao todo, foram dispostas 24 linhas em um polígono irregular de área igual a 21,56 ha. O cultivar de milho utilizado foi o AG8088 PRO2. Para a estimativa de produtividade, foi considerada o índice da colhedora, sendo esta o modelo New Holland CR 9060. Para dados dessa natureza, a produtividade é apresentada em toneladas por hectare em um espaçamento um m². Os arquivos são disponíveis em formato vetorial georreferenciado (*shapefile*).

Para tratamento e filtragem dos dados de produtividade, foram definidos dois critérios de remoção de amostras, sendo: 1- Definição de um intervalo de confiança para remoção de outliers; 2- Remoção de ruídos provenientes do processo de colheita automatizada.

- 1- Foi definido o intervalo de confiança de acordo com recomendações, com base na estatística clássica para remoção de outliers. Assim, foram utilizados três vezes o desvio padrão ao redor da média da produtividade. Por conseguinte, estabeleceu-se limiares superiores e inferiores com base na média populacional, definindo um intervalo de confiança de 99,7%.
- 2- Os ruídos foram excluídos conforme recomendado pelo manual dos fabricantes das colheitadeiras, ou seja, eliminou-se amostras, dados de produtividade nula e negativa.

Na condição supracitada, os valores inferiores têm limiar mínimo de zero, uma vez que não é possível atribuir produtividade negativa às plantas para exclusão de *outliers*. Dessa forma, os valores não pertencentes ao intervalo foram excluídos.

Para uma descrição detalhada da produção de cada área, foi feita uma análise utilizando estatística descritiva sobre a variável produtividade. A priori, calculou-se valores de média, desvio padrão e coeficiente de variação para essa variável. Ademais, ainda foi feita uma análise exploratória, a fim de estimar os valores de máximo, mínimo, quartil 1, quartil 2 e mediana e histograma da variável.

3.2.4 Aquisição dos dados multiespectrais

Modelos de estimativa de produtividade com o uso de imagens multiespectrais de alta resolução espacial são amplamente aplicados em estudos para culturas de ciclo curto, pois possibilitam a visualização em detalhes da distribuição espacial de áreas mais ou menos produtivas. As imagens que compõem os modelos puramente espectrais são registradas por aeronaves que podem ser ou não tripuladas (GUO et al.; 2020; FATHIPOOR et al., 2019) e por satélites com bandas espectrais sensíveis a variáveis agrícolas (JI et al., 2021; SUN et al., 2020).

Para definição das variáveis de predição dos modelos de estimativa deste estudo, foram utilizadas imagens multiespectrais de alta resolução do satélite Planet, pois os experimentos encontravam-se em áreas pequenas, havendo alta disponibilidade de imagens referentes ao período de desenvolvimento da cultura.

Especificamente para constelação de satélites Planet, as imagens são capturadas por uma câmara CCD equipada com o filtro Bayer Mask. O sensor trabalha com a filtragem de fótons, para que eles sejam convertidos em elétrons. Dessa forma, os valores de número digital de cada pixel em cada banda são obtidos mediante a amplificação dos elétrons. As imagens Planet possuem resolução espacial de 3 metros e resolução radiométrica que varia entre 12 bits para as imagens “analytic”. A configuração espectral do satélite é de 4 bandas, sendo elas: (comprimento de onda médio) azul (ρ_{485}), verde (ρ_{545}), vermelho (ρ_{630}) e infravermelho próximo (ρ_{820}). As bandas são capturadas pelo instrumento óptico Planet Scope 0, Planet Scope 1 e Planet Scope 2 (PLANET, 2015).

3.2.5 Processamento dos dados multiespectrais

As imagens advindas do satélite Planet podem ser obtidas em três níveis distintos de processamento, sendo eles: *PlanetScope Basic Scene Product*, *PlanetScope Ortho Scene Product* e *PlanetScene Ortho Tile Product*. As variações entre as imagens disponibilizadas oscilam entre a presença de correção geométrica (contida nos dois últimos tipos de imagem) e se a imagem está em projeção UTM ou não (apenas a última imagem). Em todos os casos, as imagens fornecidas são corrigidas atmosféricamente pela própria empresa, ou seja, os números digitais das bandas são disponibilizados em valores de reflectância de superfície.

A imagens foram tomadas ao longo dos principais estádios fenológicos de desenvolvimento da cultura de milho presente na área de estudo. Nessa fase, priorizou-se apenas imagens que apresentaram significativas mudanças espectrais nas bandas do visível e do infravermelho próximo, pois em muitos estádios consecutivos houve pouca variabilidade espectral. Ressalta-se que cada estádio fenológico possui uma característica espectral distinta, de modo que a variabilidade espectral depende de fatores abióticos e bióticos da cultura e da data de cultivo.

Na Figura 5, são apresentadas as datas correspondentes aos estádios fenológicos, com início no Plantio (P). Tais estádios são representados pela cor verde. As linhas vermelhas indicam os intervalos em que foram tomadas as imagens.

Imagem
Data da aquisição
Estádio Fenológico



FIGURA 5. Estados fenológicos e aquisição das imagens para a área de estudo

3.2.6 Geração dos modelos de predição e controle de qualidade

Para geração de modelos, foram selecionadas amostras da população total de dados de produtividade para área de estudo. Dividiu-se cada área em um grid, com 100 retângulos iguais. Os grids foram utilizados para selecionar um ponto por retângulo. Dessa forma, foram selecionados ao todo 100 pontos do conjunto total de dados, distribuídos de forma regular ao longo de cada talhão de estudo (Figura 6). Teve-se o cuidado de selecionar os pontos mais próximos do centroide de cada retângulo, ou seja, dentro de cada amostra de 100 pontos foi realizada a divisão entre os dados para treinamento e validação do modelo. Assim, de forma aleatória, selecionou-se 80 pontos para treinamento dos modelos e 20 pontos para a validação.

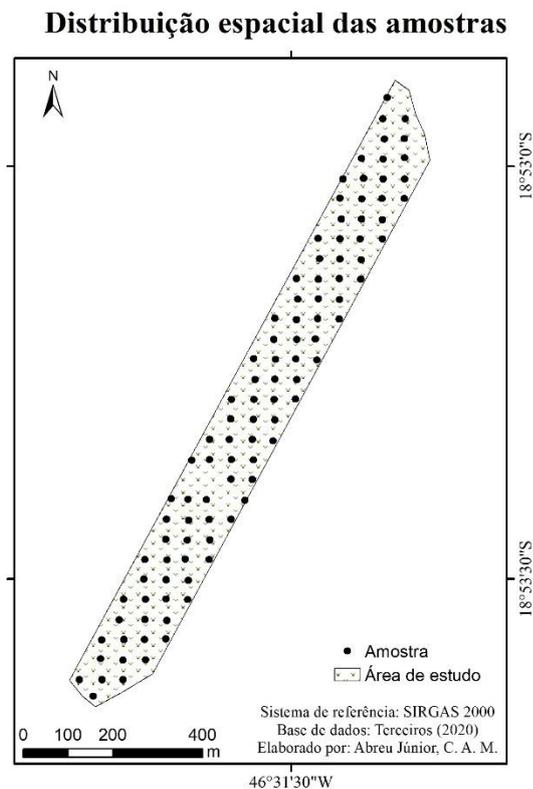


FIGURA 6. Distribuição espacial das amostras na área de estudo

Os valores de reflectância de superfície de bandas e de índices multiespectrais foram extraídos por meio da geração de regiões de interesse, utilizando parte dos pontos dos arquivos vetoriais. Isto posto, foram considerados o pixel correspondente ao centroide da grade regular. Dessa forma, cada valor de produtividade (extraído do centroide) georreferenciado foi relacionado com o respectivo número digital na imagem

Após a seleção dos pontos, calculou-se o modelo de Pearson entre a produtividade e os valores espectrais advindos das imagens, com respectivos IV derivados. Tendo em vista a alta quantidade de bandas e IVs disponíveis para estimar produtividade, definiu-se o critério de maior correlação com produtividade. Dessa forma, iriam compor o modelo as bandas e índices de vegetação que apresentassem maior valor de correlação, sendo selecionado os cinco maiores (Tabela 2). Os índices/bandas selecionadas variaram de acordo com cada estágio fenológico. A Tabela 3 mostram os valores dos parâmetros de entrada na constituição dos modelos não paramétricos para a área de estudo. Os dados que são omitidos, assim como os estágios fenológicos não apresentados na Tabela 3, são aqueles que utilizaram os parâmetros default do software Weka.

Para o processamento pelo método paramétrico, foi utilizado o algoritmo de regressão linear (Linear Regression - LR). Ele consiste na utilização de um ou mais parâmetros independentes que sejam capazes de descrever a variável de estudo. Dessa forma, é determinado a variável resposta mediante a geração de um modelo que melhor se ajuste aos dados utilizados (LEROUX et al., 2020; PAUL et al., 2020).

Foi utilizada a metodologia do algoritmo de regressão linear do software Weka, o qual, no cálculo do modelo, inclui a aplicação da técnica de regressão linear passo a passo (Linear Regression Stepwise - LRS) do tipo *forward* (passo à frente). A LRS considera, inicialmente, um modelo de regressão simples, usando como variável auxiliar (bandas e índices multiespectrais) a variável de maior coeficiente de correlação com a variável resposta (produtividade). O processo se sucede ao passo que uma nova variável auxiliar é incorporada ao modelo, sendo interrompido quando nenhuma nova variável é inserida no modelo.

TABELA 2. Correção entre as bandas e índices derivados das imagens do satélite Planet com a variável produtividade, para cada estádio fisiológico analisado na área de estudo

Bandas/ Índices	VE	V5	V8	V11	R1	R2	R3	R4	R5	R6
GLI	0,0742	-0,02578	-0,01331	0,0422	-0,0122	-0,0194	0,02668	-0,0429	-0,01571	0,15414
VARI	0,0556	-0,02149	-0,03241	-0,0359	-0,0353	-0,0273	0,05432	-0,015	-0,03898	0,17706
NGRDI	0,0472	-0,01906	-0,02675	-0,02	-0,0329	-0,0265	0,05732	-0,0183	-0,03848	0,17619
TGI	0,0428	0,01282	-0,00812	-0,0701	-0,0641	-0,0428	0,04966	-0,0292	-0,043	0,16807
Verde	0,0171	-0,11246	0,1172	-0,0458	-0,1221	-0,1075	0,01924	-0,1487	-0,02874	0,00905
Vermelho	-0,0168	-0,06566	0,11725	-0,0204	-0,0665	-0,0785	-0,01262	-0,0875	-0,00472	-0,17732
TVI	-0,0249	-0,0054	-0,044	0,0398	0,0892	0,1581	0,02206	0,0593	-0,03571	0,1952
EVI	-0,0285	0,02776	-0,0656	-0,0215	0,0315	0,026	0,03423	0,0483	-0,04358	0,20399
RVI	-0,0319	0,00901	-0,1079	0,0709	0,1045	0,1469	0,01167	0,0862	-0,06255	0,18717
Azul	-0,0325	-0,06253	0,08543	-0,0327	-0,0894	-0,0788	0,00814	-0,0359	-0,02077	-0,02483
NDVI	-0,0343	0,03579	-0,10665	0,034	0,1012	0,1462	0,02182	0,0969	-0,02766	0,20519
OSAVI	-0,0343	0,03579	-0,10665	0,034	0,1012	0,1462	0,02182	0,0969	-0,02767	0,20519
SAVI	-0,0343	0,03579	-0,10665	0,034	0,1009	0,1462	0,02182	0,0969	-0,02768	0,20519
MTVI	-0,0356	0,04037	-0,10684	0,0287	0,1008	0,146	0,02338	0,0984	-0,0234	0,20779
MSAVI	-0,0356	0,04036	-0,10683	0,0287	0,1008	0,146	0,02338	0,0984	-0,02342	0,20779
IVP	-0,0424	-0,04675	-0,00952	0,0499	0,0895	0,1719	0,01646	0,0512	-0,0324	0,18352
CVI	-0,0661	0,07884	-0,10852	0,1169	0,1258	0,118	-0,03953	0,1315	-0,00357	-0,056
CI_G	-0,0748	0,03678	-0,11577	0,1065	0,1574	0,1842	-0,01597	0,1361	-0,03304	0,14902
GNDVI	-0,0771	0,05503	-0,11681	0,0618	0,1396	0,1795	-0,00373	0,1383	-0,0081	0,16513

TABELA 3. Parâmetros de entrada para confecção dos modelos Support Vector Machine (A), Neural Net (B)

Estádio fenológico		Support Vector Machine		
		c	FilterType	Kernel
VE		1,0	Normalize	PolyKernel
R3		1,0	Normalize	PolyKernel
		Neural Net		
		HiddenLayer	LearningRare	Momentum
V8		3	0,3	0,2
V11		3	0,1	0,3
R1		1	0,1	0,2
R2		3	0,3	0,2

3.2.7 Mapa de discrepância entre valores de produtividade observados em campo e estimados pelo modelo

Para analisar os resultados obtidos pelos algoritmos de estimativa e verificar se houve uma diferença entre os dados observados em campo e os estimados pelos modelos, foram feitos mapas de diferença entre as superfícies geradas para ambos os casos. Portanto, foram subtraídos os valores das superfícies dos dados observados em campo e os valores estimados pelos modelos. Assim, foi possível analisar se a distribuição gerada pelo modelo é majoritariamente qualitativa ou quantitativa, bem como tendências no espaço. Com essa técnica, pode-se verificar se o modelo é capaz de registrar quais são as regiões de baixa e de alta produtividade do talhão, independentemente da acurácia dos valores absolutos.

Para isso, os mapas de diferença de superfície foram normalizados com valores entre -1 e 1. Para complementar a análise, nessa etapa, foram novamente calculadas as métricas avaliativas de RMSE% e MAPE, assim como os valores de média e desvio padrão oriundos do mapa. Dessa forma, também foi possível analisar as discrepâncias contidas entre o modelo e a sua aplicação (o mapa).

3.3 RESULTADOS E DISCUSSÕES

3.3.1 Resultados da análise exploratória da variável produtividade

A Tabela 4 apresenta os valores descritivos da variável produtividade (kg/ha) para a área de estudo. A Figura 7 apresenta os histogramas da área de estudo, sobreposto pela curva de distribuição normal, para a variável produtividade.

TABELA 4. Valores descritivos da variável produtividade: D.P. = Desvio padrão, Q1= Quartil1, Q3= Quartil3, C.V. = Coeficiente de variação

Variável	Média	D.P.	Mínimo	Máximo	Q1	Mediana	Q3	C. V. (%)
Produtividade	8871	1110	6345	11254	8115	8808	9675	12,52

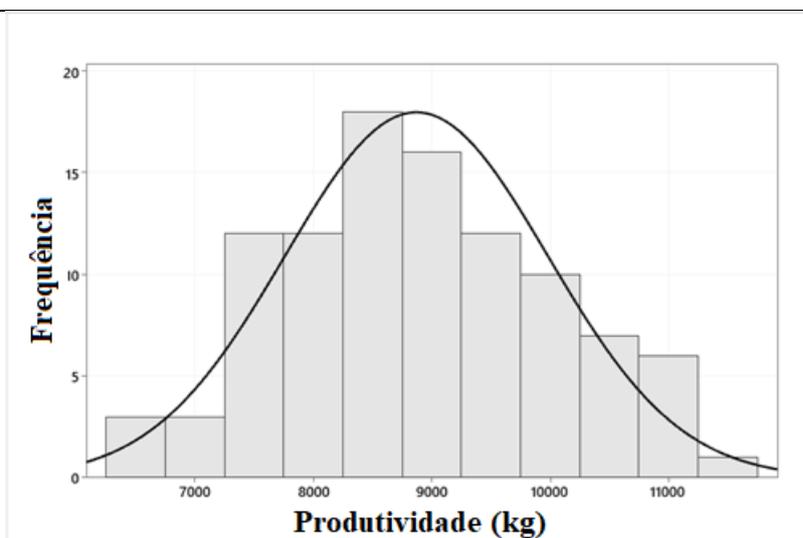


FIGURA 7. Histograma e curva de distribuição normal da área de estudo na lavoura do milho

Para área de estudo, os valores do limiar superior e inferior são respectivamente 11254 kg/ha e 6345 kg/ha. O primeiro quartil (Q1) apresenta valor de 8115 kg/ha, enquanto o terceiro quartil (Q3) possui o valor de 9675 kg/ha. Dessa forma, 50% dos valores de produtividade estão compreendidos no intervalo estabelecido entre Q1 e Q3, que possui amplitude interquartil de 1560 kg/ha. A produtividade média dessa área de cultivo do milho foi igual a 8871 kg/ha, ficando acima da produtividade média do estado de Minas Gerais na safra 2019/2020, que obteve valor de 5726 kg/ha (CONAB, 2020). O coeficiente de variação (CV) foi igual a 12,52%, o que não impede a variável de ser estimada. Tagarakis e Ketterings (2017), mesmo com um CV 26.8% no estágio vegetativo 6, conseguiram estimar a produtividade do milho com resultados significativos. Por meio da análise gráfica, é possível observar que os dados possuem

tendência à simetria. Esse resultado corrobora o menor CV encontrado pela variável produtividade, sendo este mais um indício do potencial da variável a ser estimada. Contudo, ainda é possível notar uma leve tendência à esquerda do gráfico, atribuindo características inerentes a uma curva assimétrica positiva.

3.3.2 Métricas de validação entre produtividade e os estádios fenológicos e seleção do melhor modelo para estimativa

A Tabela 5 apresenta os estádios fenológicos e as métricas de validação utilizadas, além dos índices/bandas que obtiveram melhor resposta com a variável produtividade para cada estágio fenológico.

TABELA 5. Análise do desempenho dos modelos por estágio fenológico para área de estudo: Neural Net (NN); Linear Regression (LR); Support Vector Machine (SVM); Random Forest (RF); Algoritmo (Alg.); Root Mean Square Error Percentagem (RMSE%); Mean Absolute Percentagem Error (MAPE)

Data	E. F.	Alg.	MAPE	RMSE	Bandas/Índices
20/02/20	VE	SVM	6,61	9,68	GLI, VARI, GNDVI, CI-G, CVI
13/03/20	V5	RT	7,09	9,62	R, B, G, CVI, GNDVI
31/03/20	V8	NN	7,39	9,7	R, G, GNDVI, CI-G, CVI
11/04/20	V11	NN	7,95	10,32	CVI, CI-G, RVI, TGI, GNDVI
25/04/20	R1	NN	7,25	9,25	G, RVI, CVI, GNDVI, CI-G
13/05/20	R2	NN	7,07	9,17	NIR, CI-G, GNDVI, TVI, RVI
22/05/20	R3	SVM	7,30	10,53	EVI, CVI, NGRDI, VARI, TGI
05/06/20	R4	RT	9,62	13,62	G, GNDVI, CI-G, CVI, MTVI
18/06/20	R5	RF	9,42	12,88	NGRDI, VARI, TGI, EVI, RVI
03/07/20	R6	RT	15,91	20,96	MSAVI, MTVI, SAVI, OSAVI, NDVI

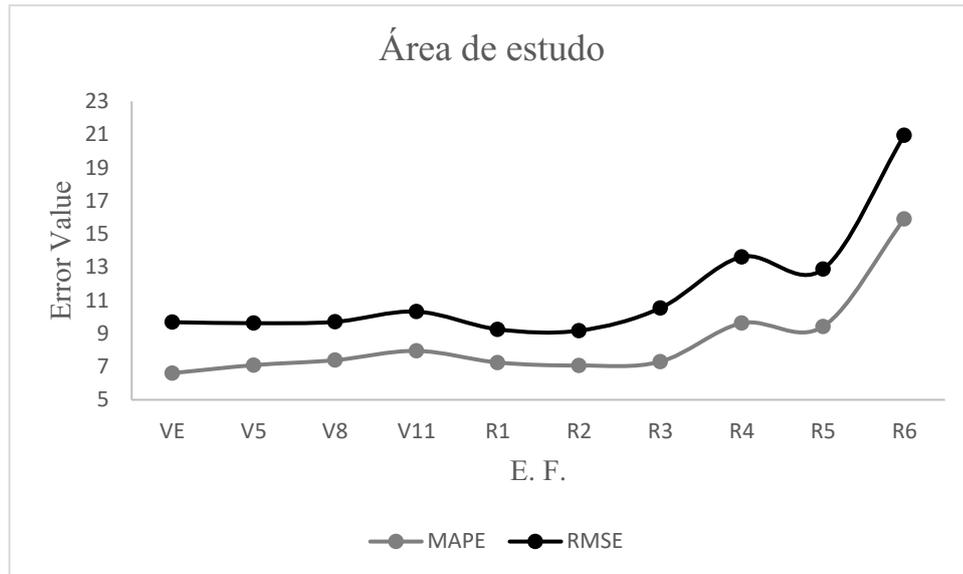


FIGURA 8. Curvas de Root Mean Square Error percentagem (RMSE%) e Mean Absolute Percentagem Error (MAPE) para a área de estudo

Observa-se que, para a área de estudo, já é possível estimar a produtividade no estágio vegetativo VE com RMSE% e MAPE, inferiores à 26% e 21%, respectivamente. Os modelos são mais assertivos no em R2 e VE, cujo os valores MAPE e RMSE% são inferiores a 25% e 23%, respectivamente.

A Figura 8 apresenta também o comportamento das métricas MAPE e RMSE%, gerada pelos modelos de estimativa de produtividade ao longo dos estádios fenológicos. O estágio vegetativo VE parte com valor próximo ao limiar inferior do gráfico. Ao longo do desenvolvimento dos estádios vegetativos iniciais, teve-se um crescimento dos erros, o que pode ser notado em V11. A partir desse momento, é possível observar uma queda no valor das métricas utilizadas, com a entrada dos estádios reprodutivos.

Dessa forma, o menor valor de MAPE e RMSE% é encontrado em R2 para área de estudo. Esse resultado aponta para uma possível tendência no estágio fenológico R2 ser o período ideal para estimativa de produtividade. Nesse estágio fenológico, a planta está em processo de polinização. Os estilo-estigma são visíveis, uma vez que eles tentam capturar os grãos de pólen (SOLOGUREN, 2015). A diferença de 4 dias indica que os estilos estigmas de cada planta já estavam polinizados ou em estágio avançado, uma vez que esse período de polinização costuma acontecer no período de 2 a 3 dias. (MAGALHÃES; DURÃES, 2006). Em um estudo realizado por Ma et al. (2021), os autores tentaram estimar a produtividade da cultura do milho por meio de imagens multiespectrais, utilizando algoritmo de RNA. No

referido trabalho, os autores constataam que os melhores resultados de coeficiente determinístico para os modelos de estimativa gerados são encontrados 2 meses antes do período de colheita.

De modo geral, os resultados evidenciam que os algoritmos com maior repetição na geração dos modelos dos estádios fenológicos foram SVM, NN e RT, os quais apresentaram melhor desempenho nos estádios fenológicos VE e R3 para SVM, V8, V11, R1 e R2 para NN e V5 R4 e R6 para RT. O algoritmo de SVM é amplamente utilizado na cultura do milho e mostra-se capaz de modelar diferentes parâmetros agronômicos, tais como doenças (ALEHEGN, 2019; ARAVIND et al., 2018) e transpiração (FAN et al., 2021; ZHENG et al., 2021). Já as NN são usadas com mais frequência em trabalhos ligados a estimativa de produtividade na cultura do milho, tanto por meio de imagens advindas de aeronaves remotamente pilotadas (GARCÍA-MARTÍNEZ et al., 2020) quanto de sensores orbitais (WANG et al., 2020). Para conseguir estimar a produtividade mediante imagens do satélite MODIS e GLASS, Wang et al. (2020) utilizaram NN. No estudo supracitado, os autores conseguiram estimar a produtividade com MAPE inferior a 10% e RMSE a 700 kg/ha na maioria dos casos.

Ao observar os resultados em VE e V5, é possível constatar a presença de bandas do visível e os índices de vegetação delas derivados. O comportamento da radiação eletromagnética no visível é determinado majoritariamente pela clorofila. Os pigmentos do mesófilo inclinam-se a absorver os comprimentos de onda no intervalo do azul e vermelho (LIESENBERG, 2007). Dessa forma, a radiação proveniente do intervalo espectral do verde é refletida em maior incidência, o que pode justificar a presença da banda verde e índices de vegetação como o GNDVI e CI-G nos modelos dos estádios fenológicos VE e V5.

Os índices de vegetação do visível também compõem, em maior número, os modelos dos estádios fenológicos V8. Para Garcia-Martines et al. (2020), os índices VARI e TGI apresentam potencial para a estimativa de produtividade na cultura do milho. No trabalho, os autores geraram modelos baseado no algoritmo de NN, utilizando seis IV como parâmetros de entrada, entre os quais estão TGI e VARI.

Os modelos originados dos estádios fenológicos reprodutivos marcam a presença do infravermelho próximo pela primeira vez entre os parâmetros de entrada utilizados. É possível verificar a sua presença nos estádios R1, R2, R3, R4, R5 e R6. A radiação proveniente do infravermelho próximo tende a ser sensível à vegetação devido à baixa absorção e retroespalhamento quando em contato com estruturas celulares internas da planta (LIESENBERG, 2007). Os estádios reprodutivos marcam o desenvolvimento da espiga do milho (SOLOGUREN, 2015). Em R1, são finalizados o crescimento do colmo e internódios.

Nesse estágio, os estílo-estigma continua o seu crescimento, a fim de que eles sejam polinizados. Desse modo, o processo de fecundação do óvulo pode continuar (MAGALHÃES; DURÃES, 2006). Já em R2, ocorre o acúmulo de açúcares solúveis no endosperma do grão, que auxilia no aumento da massa do grão. Nesse período, a espiga de milho fica próxima do tamanho final que irá atingir (SOLOGUREN, 2015). Essas mudanças internas podem ter favorecido a maior refletância proveniente do infravermelho próximo, assim como os índices de vegetação dele derivados.

O estágio fenológico R6 é caracterizado pela maturidade fisiológica dos grãos. Nesse período, o processo de senescência nas folhas torna-se mais visível, uma vez que elas ganham a aparência de “mais seca” devido à degradação da clorofila. Esse estágio é comumente destinado à colheita, uma vez que os grãos se apresentam independentes da planta mãe. Diante desse contexto, destaca-se a presença do índice SAVI, assim como as suas variações MSAVI e OSAVI, entre os selecionados para a constituição do modelo de R6.

Os resultados da Tabelas 5 ainda mostram uma oscilação entre as métricas de avaliação utilizadas. Em geral, a possibilidade de estimar a produtividade já nos estágios fenológicos iniciais é explicado pela sensibilidade de bandas e índices com macro e micronutrientes do solo, exposto devido dossel da cultura pouco desenvolvido. Estudos (RAY et al., 2019; BABU et al., 2020) mostram que a quantidade de macro e micronutrientes no momento do plantio geram influência e possuem alta correlação com a produtividade final da cultura (SELVI et al., 2019).

Para assumir melhor a representação da variabilidade entre os valores da produtividade observada e estimada, na Figura 9, são mostrados os gráficos de dispersão dos valores observados versus valores os valores estimados para a área de estudo.

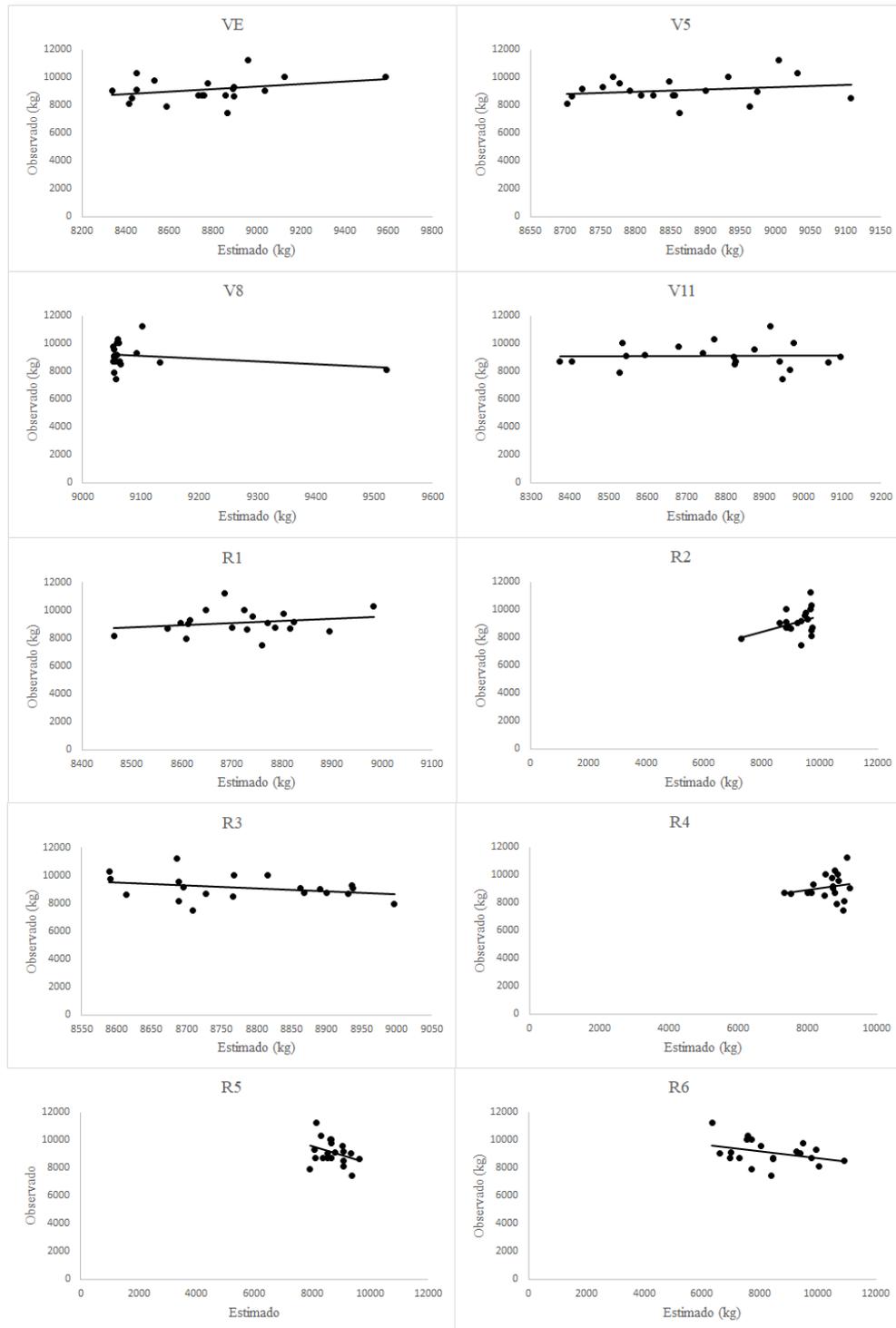


FIGURA 9. Valores observados x estimados para os estádios fenológicos da área de estudo

Os gráficos expostos na Figura 9 mostram os pontos associados aos valores estimados, que estão presentes no eixo das ordenadas, e os valores observados, que estão contidos no eixo das abscissas. Para os estádios fenológicos VE, V5, V8 e V11, os dados apresentam uma maior dispersão ao longo do eixo das abscissas, seguindo uma tendência linear.

3.3.3 Espacialização da variável estimada

As Figuras 10 e 11 apresenta o resultado dos mapas gerados em função da estimativa de produtividade para as Áreas 1 e 2 respectivamente.

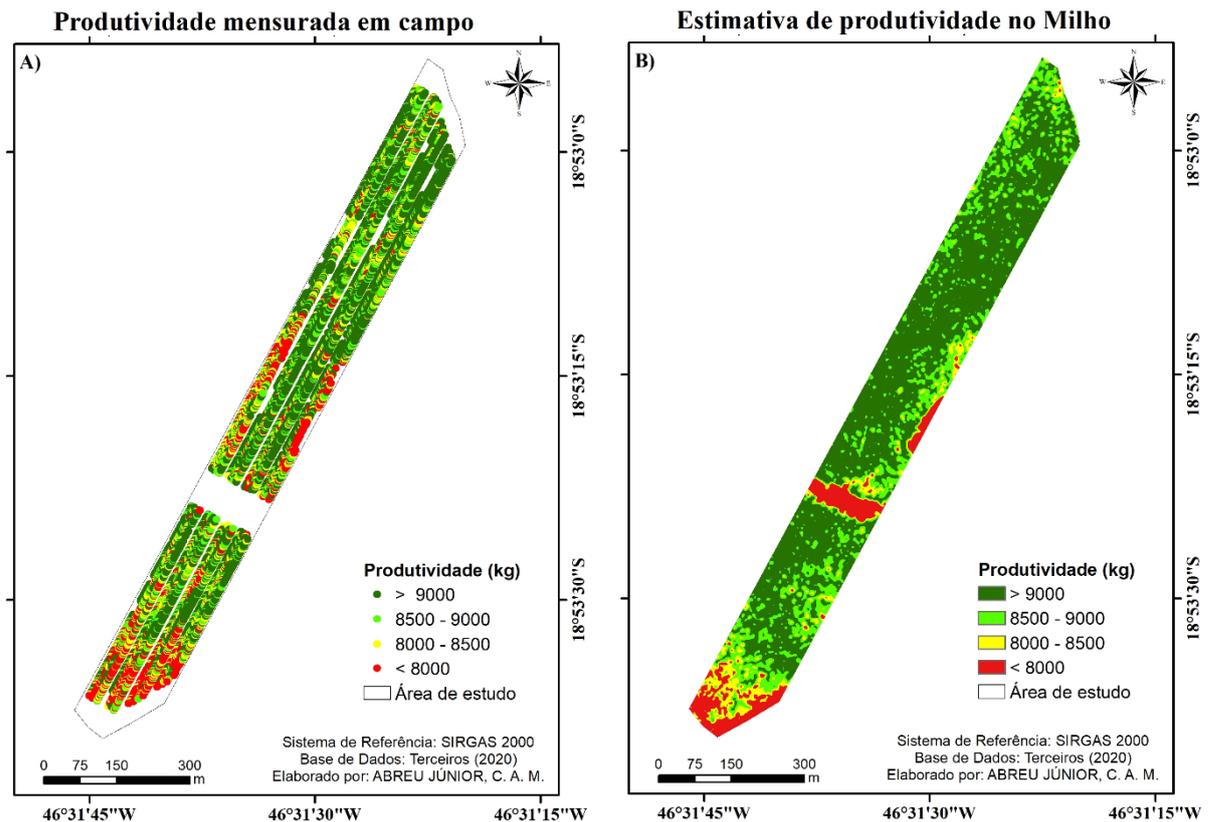


FIGURA 10. Distribuição espacial da produtividade na área de estudo, proveniente do algoritmo SVM aplicado sobre as bandas/índices advindos do satélite Planet; produtividade mensurada em campo

Na Figura 10, a imagem A) apresenta os pontos mensurados em campo, enquanto a imagem B), o resultado da espacialização da variável produtividade, estimada pelo modelo SVM. Analisando a imagem, é possível identificar diferenças entre o cenário mensurado e o estimado. A principal delas é a quantidade de valores compreendidos no intervalo inferior a 3000 kg, representado pela cor vermelha. O segundo mapa apresenta uma redução significativa na quantidade de área compreendida nesse intervalo.

A Figura 11 apresenta a diferença normalizada entre os valores estimados e os valores observados para cada área de estudo.

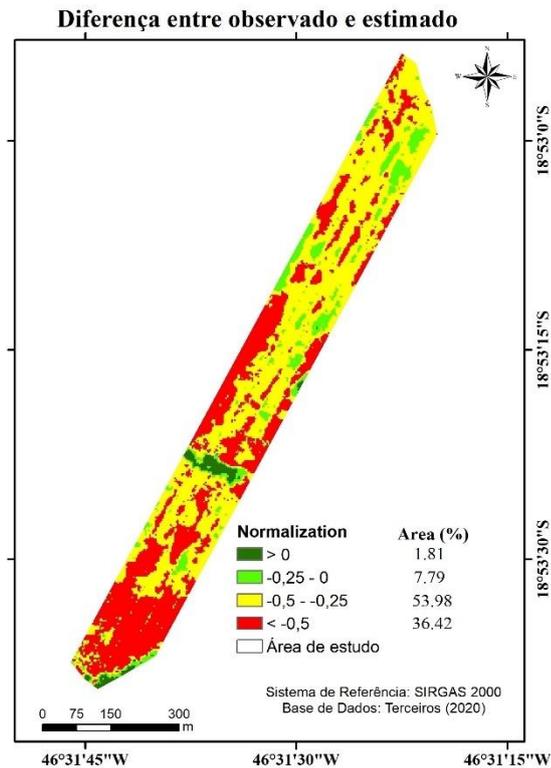


FIGURA 11. Diferença entre valores observados e estimados

TABELA 6. Estatísticas das superfícies observadas e estimadas, onde Root Mean Square Error Percentagem (RMSE%) e Mean Absolute Percentagem Error (MAPE)

Métricas analisadas	Área de estudo	
	Observado	Estimado
Média (kg/ha)	9137,66	8448,98
Desvio Padrão (kg/ha)	924,62	917,76
RMSE%	12,31	
MAPE (%)	8,38	

A área de estudo apresenta maiores diferenças na porção inferior do talhão e em algumas faixas nos extremos esquerdo e direito. As maiores proporções de valores estão localizadas entre o intervalo -0,5 até -0,25 com 53,98% das áreas centrada nessa classe. Esses resultados corroboram os encontrados na Tabela 5, embora sejam numericamente diferentes. Nesse contexto, os resultados são mais quantitativos, sendo estes mais próximos dos valores observados em campo.

3.4 CONSIDERAÇÕES FINAIS

Os modelos espectrais de estimativa da produtividade mostram que:

- Os modelos espectrais são mais acurados e precisos quando construídos com imagens espectrais referentes aos estádios R2, que está próximo do período de alto vigor vegetativo da cultura, definindo-o como ideal;
- Devido à influência da resposta espectral do solo, a partir de imagens tomadas no estágio VE, é possível estimar produtividade com acurácia e precisão satisfatória;
- A composição das variáveis preditoras e a acurácia e precisão dos modelos estão associados a estágio fenológico de desenvolvimento, ou seja, a arquitetura do modelo é variável para diferentes áreas e estágio de senescência da planta.

4 CAPÍTULO 2: Estimativa de produtividade na cultura do milho a partir de modelos multiespectrais mais robustos baseados em algoritmos de aprendizado de máquina

RESUMO

O uso de algoritmos de aprendizado de máquina vem ganhando cada vez mais espaço na agricultura. Por se tratar de modelos robustos, as aplicações voltadas à agricultura permeiam diversas áreas de interesse dessa ciência, tais como detecção de estresse hídrico, detecção de doenças, análises sobre doses de nitrogênio, estimativa de produtividade, dentre outros. Entretanto, normalmente os modelos são aplicados na generalização da variável na sua própria área de estudo. Embora não estejam erradas, as aplicações de modelos que ficam restritas a uma mesma área não tem o seu potencial de estimativa testado em áreas de condições similares, o que seria possível com a construção de um modelo mais robusto. Sendo assim, esta pesquisa se propõe a responder se é possível estimar a produtividade do milho considerando mais de uma área experimental, utilizando um modelo mais robusto, por meio de imagens multiespectrais e técnicas de aprendizado de máquina. Para tanto, a partir de produtividade derivada de três áreas experimentais, foram gerados modelos multiespectrais com base em imagens do satélite Sentinel 2 e os algoritmos Support Vector Machine, Neural Net, Radom Forest e Random Tree. Para a validação dos modelos, foram utilizadas as métricas avaliativas RMSE% e MAPE. Os resultados indicam que os melhores modelos foram aqueles originados dos algoritmos SVM e RF. Para aplicações individuais, ou seja, considerando uma única área para treinamento e validação dos modelos, foi possível estimar a produtividade com um para RMSE% e MAPE de 17,75% e 15,27%, respectivamente. Na construção de um modelo mais robusto, ou seja, um modelo único para todas as áreas, os valores para RMSE% e MAPE foram inferiores a 30%.

Palavras-chave: Cultura do milho; Produtividade; aprendizado de máquina; Modelos robusto.

4 CHAPTER 2: Corn yield estimation from more robust multispectral models based on machine learning algorithms

ABSTRACT

Machine learning algorithms have increasingly gained ground in agriculture. Their robustness makes them suitable for different areas of interest in this field, such as detecting water stress and diseases, analyzing nitrogen doses and estimating yield, among other applications. However, these models are typically used for general yield estimates in a specific study area. Although this application is not incorrect, restricting these models to a single area precludes testing their estimation potential in areas with similar conditions, which would be possible with a more robust model. As such, this study aimed to determine whether corn yield can be estimated considering more than one experimental area and using a more robust model with multispectral images and machine learning techniques. To that end, the yield of three experimental areas was used to generate multispectral models based on Sentinel 2 satellite images and support vector machine, neural network, random forest and random tree algorithms. RMSE% and MAPE evaluation metrics were used to validate the models. The results indicated that the best models were derived from the SVM and RF algorithms. For individual applications, that is, considering a single area to train and validate the models, it was possible to estimate yield with an RMSE% and MAPE of 17.75% and 15.27%, respectively. In order to generate a more robust model, that is, a single model for all the areas, RMSE% and MAPE values were lower than 30%.

Keywords: Corn crops; yield; machine learning; robust models.

4.1 INTRODUÇÃO

O uso de algoritmos de aprendizado de máquina vem ganhando cada vez mais espaço na agricultura. Essas técnicas baseiam-se na identificação de padrões estruturais (CARLEO et al., 2019) e na capacidade de processar uma alta demanda de dados de diversas naturezas, tais como ambientais, desenvolvimento sustentável, agricultura, entre outros. Por se tratar de modelos robustos, as aplicações voltadas à agricultura permeiam diversas áreas de interesse dessa ciência, tais como detecção de estresse hídrico, detecção de doenças, análises sob doses de nitrogênio, estimativa de produtividade, entre outros.

Além da diversidade de áreas, os algoritmos de aprendizado de máquina também se pluralizam na aplicação de diversas culturas. Vários estudos baseiam-se na estimativa de produtividade de diversas culturas agrícolas, tais como o milho (ZHANG et al., 2020), café (KOUADIO et al., 2018), soja (MAIMAITIJIANG, et al. 2020) e cana (SUMESH et al., 2021). A confecção desses modelos varia desde escalas locais ou individuais até escalas maiores (PANG, SHEN, YAO, 2021), sejam elas estaduais ou nacionais.

Entretanto, normalmente, os modelos são aplicados na espacialização da variável na sua própria área de estudo. Embora não estejam erradas, as aplicações de modelos que ficam restritas a uma mesma área podem provocar uma certa tendência na estimativa dos valores de produtividade. Isto é, quando o conjunto de dados utilizados para a geração do modelo pertencem à mesma área daquele que se deseja estimar não são testados em outras áreas com condições similares àquelas do modelo. Uma alternativa para a atenuação de uma tendência presente nesses modelos é a construção de modelos em escala local que se baseiam em múltiplas áreas, sendo elas distintas entre as etapas de treinamento e aplicação. Diante desse contexto, para que a produtividade continue a ser incrementada, necessita-se de uma metodologia coesa e prática para a estimativa da produtividade, que pondere situações que possam limitar o processo de estimativa de dados do modelo. Sendo assim, esta pesquisa busca responder se é possível estimar a produtividade de uma área diferente por meio de um modelo mais robusto, utilizando imagens multiespectrais e técnicas de aprendizado de máquinas.

Diante do exposto, o objetivo desta pesquisa foi estimar a produtividade de áreas da cultura do milho utilizando imagens multiespectrais e técnicas de aprendizado de máquinas. Inicialmente, foi avaliado o desempenho dos algoritmos para a estimativa individual das áreas de estudos e, sequencialmente, a geração de um modelo em escala local mediante a junção dos seus dados de produtividade. O intuito é que seja criado um modelo sem tendência, isto é, que os dados para a geração do modelo e aplicação sejam feitos em áreas distintas. Para tanto, foram

utilizadas imagens do satélite Sentinel 2 juntamente com os algoritmos de Support Vector Machine e Random Forest, Neural Net e Random Tree.

4.2 MATERIAL E MÉTODO

4.2.1 Área de estudo

Para o desenvolvimento desta pesquisa, foi utilizado um experimento que está localizada no município de Patos de Minas. A área está inserida na mesorregião do Triângulo Mineiro e Alto Paranaíba, Minas Gerais, Brasil (Figura 12). As áreas de estudo possuem latossolo vermelho, altitude média de 938 m, relevo plano e ondulado e precipitação anual média de 950 mm.

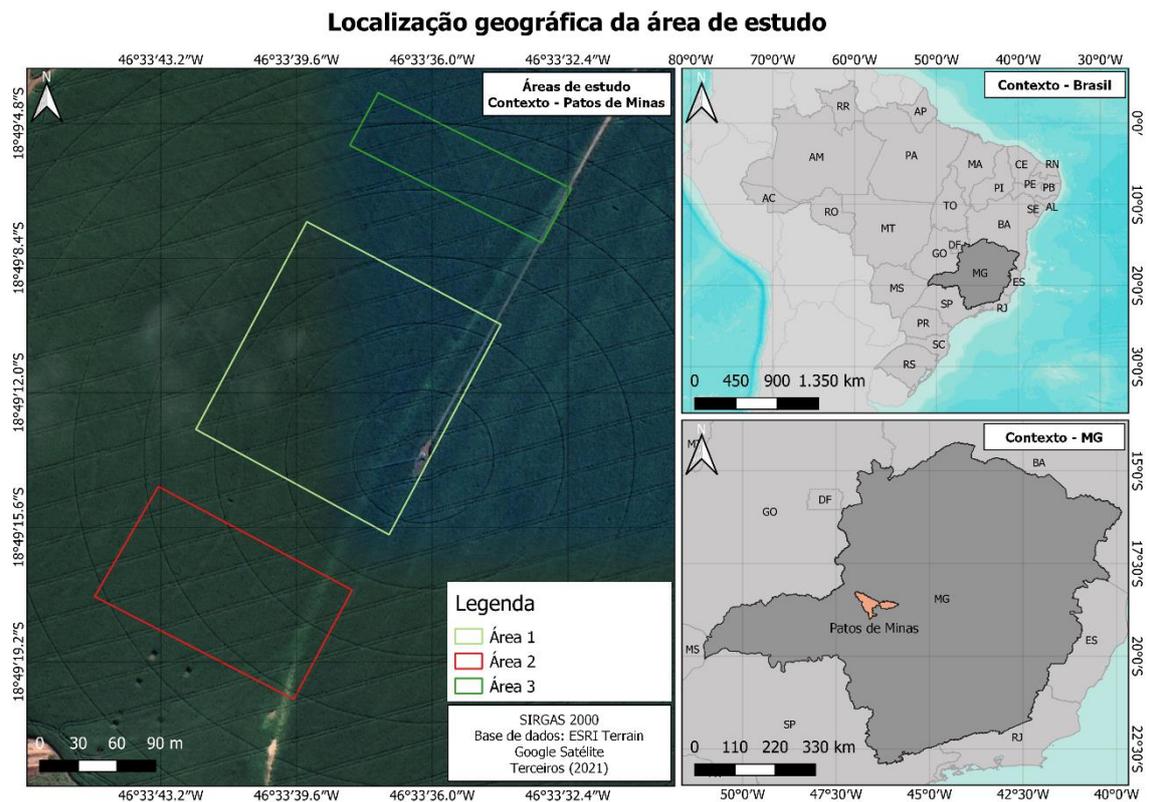


FIGURA 12. Imagem da área usada na pesquisa: contexto do município de Patos de Minas, estado de Minas Gerais, Brasil

4.2.2 Metodologia utilizada para o processamento dos dados

Inicialmente, baseamo-nos em Ji et al. (2021) e Sun et al. (2020) para criar os modelos de estimativa e as respectivas variáveis preditoras, a saber, as bandas das imagens e os índices multiespectrais, os algoritmos de regressão (aprendizado de máquina) e a variável estimada (produtividade). Em linhas gerais, a metodologia deste trabalho (Figura 13) foi desenvolvida permeando etapas de aquisição de dados de produtividade e multiespectrais, definição de algoritmos para a estimativa de produtividade, a validação do método e o mapeamento da distribuição espacial da produtividade e comparação entre os produtos gerados por cada modelo. Assim, genericamente, as etapas da pesquisa estão definidas da seguinte forma: (I) Aquisição de dados de produtividade e multiespectrais; (II) Pré-processamento dos dados multiespectrais; (III) Amostragem e filtragem dos dados de produtividade; (IV) Geração e validação dos modelos de predição; (V) Comparação entre os dados gerados pelo modelo em escala local e individual.

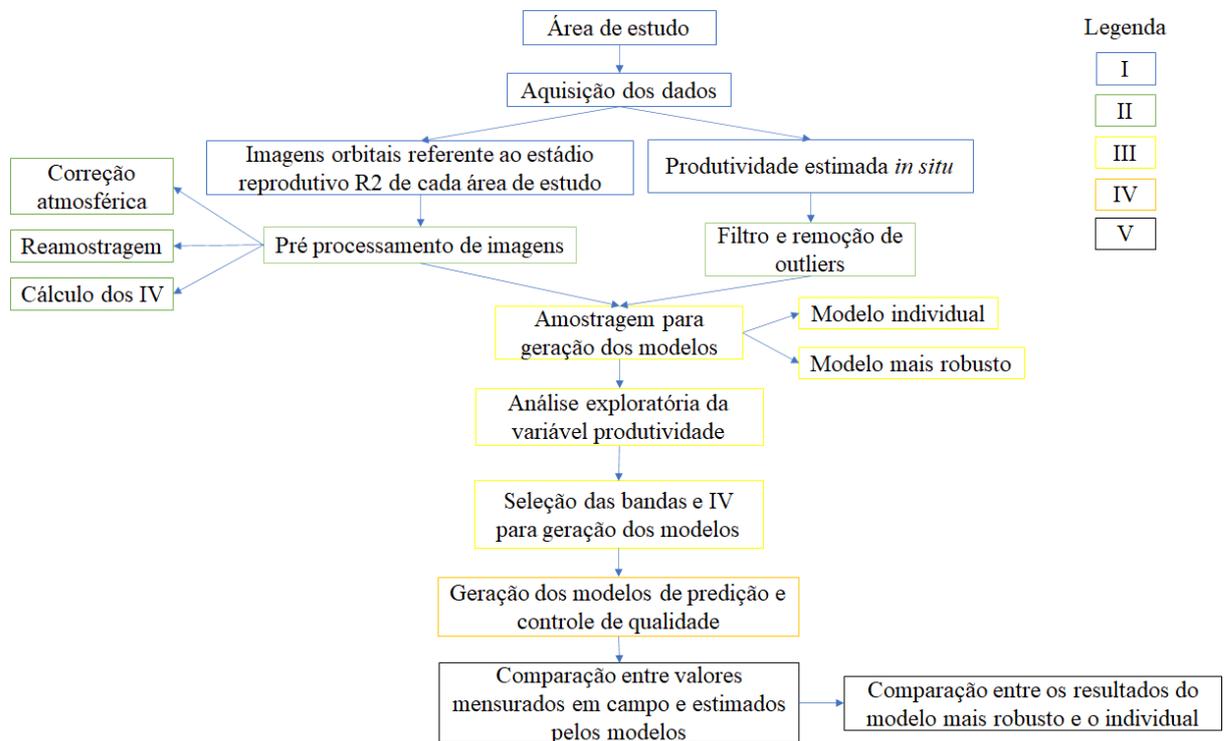


FIGURA 13. Fluxograma metodológico

4.2.3 Delimitação do experimento

As áreas de estudo utilizadas neste trabalho na análise da produtividade foram padronizadas no município de Patos de Minas. A colheita foi realizada no dia 18 de agosto de 2021 para as áreas 1, 2 e 3, sendo essa data referente à segunda safra do mercado nacional brasileiro. O experimento foi montado utilizando espaçamento entre linhas de 0,5 m e com distribuição de 3 plantas a cada metro. Dessa forma, o espaçamento vertical foi de aproximadamente 0,33 m. As áreas de estudo possuem áreas iguais a 3,32 ha, 1,77 ha e 0,83 ha para as áreas 1, 2 e 3, respectivamente. Para a estimativa de produtividade, foi considerado o índice da colhedora, sendo esta o modelo CASE IH Axial Flow 6150. Para dados dessa natureza, a produtividade é apresentada em toneladas por hectare em um espaçamento um m². Os arquivos são disponíveis em formato vetorial georreferenciado (*shapefile*).

4.2.4 Filtragem e remoção de outliers

A metodologia utilizada foi a mesma aplicada na seção “Filtragem dos dados de produtividade da cultura do milho” do Capítulo 1.

4.2.5 Aquisição de dados multiespectrais

Para a construção de modelos multiespectrais de estimativa da produtividade, foram utilizadas imagens do satélite Sentinel 2, capturadas pelo sensor instrumento multiespectral (MultiSpectral Instrument - MSI), que possui resolução radiométrica de 12 bits, com dois planos focais baseados em um detector monolítico Complementary Metal Oxide Semiconductor (CMOS), que registra a radiação proveniente do canal do visível e do infravermelho próximo, e um detector Mercury-cadmium-telluride (MCT) hibridado em uma CMOS, para a captura da radiação proveniente do infravermelho de ondas curtas. Além disso o sensor apresenta resolução espectral de 13 bandas e resolução espacial de 10 metros para as bandas (comprimento de onda): azul (ρ_{490}), verde (ρ_{560}), vermelho (ρ_{665}), infravermelho próximo (ρ_{842}) e resolução espacial de 20 metros para as bandas vermelho limítrofe 1 (ρ_{705}), vermelho limítrofe 2 (ρ_{740}), vermelho limítrofe 3 (ρ_{783}), vermelho limítrofe 4 (ρ_{865}), infravermelho de ondas curtas 1 (ρ_{1610}) e infravermelho de ondas curtas 2 (ρ_{2190}).

A imagem utilizada foi a disponível com a data mais próxima do estádio R2, referente a data 14 de maio de 2021.

4.2.6 Processamento dos dados multiespectrais

A atmosfera tem interferência nas imagens orbitais de duas maneiras completamente distintas. Ela pode atuar como um refletor, incrementando o sinal de radiância registrado pelo sensor, ou então ela pode absorver o sinal, atenuando a intensidade de energia que é transmitida ao alvo na superfície terrestre. Ressalta-se que em qualquer uma das situações o efeito é propagado ao longo de toda a imagem (KHANAL et al., 2017).

Além desses erros, as imagens do satélite Sentinel 2 são adquiridas a um nível 1C e, dessa forma, precisam ser corrigidas. Além do mais, a correção atmosférica faz-se necessária para que seja obtida a reflectância de superfície. A correção foi executada por meio do software SNAP, pelo plugin sen2cor. O plugin realiza as correções atmosférica, terrestre e cirros do topo da atmosfera, transformando as imagens para level 2A (ESA, 2018).

Após a correção atmosférica, foi feita a reamostragem das imagens com resolução espacial de 20 m (ρ_{705} , ρ_{740} , ρ_{783} , ρ_{865} , ρ_{1610} e ρ_{2190}), para que todas fossem padronizadas em uma resolução espacial de 10 m. Esse procedimento foi feito no software SNAP. Para cálculo dos novos valores dos pixels reamostrados, foram utilizados os algoritmos pelo método de interpolação do vizinho mais próximo. Os métodos de *downsampling* e *flag aggregation* utilizados foram o de *first encounter*, no qual o valor de cada pixel na imagem de saída foi definido de acordo com o valor do pixel mais próximo encontrado.

4.2.7 Análise descritiva e exploratória da produtividade

A metodologia utilizada foi a mesma aplicada na seção “Análise descritiva e exploratória da produtividade” do Capítulo 1.

4.2.8 Geração de modelos de predição e controle de qualidade

Nessa etapa, foi realizada a confecção de dois tipos de modelos, sendo eles em área individual e em escala local. A priori, foram gerados os modelos referentes a cada área de estudo, separadamente. Isto é, modelos nos quais os dados de treinamento e validação eram pertencentes a mesma área de estudo. Nesse contexto, foram construídos, ao todo, três modelos, para as áreas 1, 2 e 3, respectivamente. Num segundo momento, foi elaborado um modelo em

escala local. Para este, foram considerados os dados de produtividade de duas das três áreas de estudo para treinamento e uma terceira para validação do algoritmo.

Para a geração de modelos em área individual, seguiu-se a mesma metodologia de separação dos pontos descrita na seção “Geração dos modelos de predição e controle de qualidade” do Capítulo 1.

Para a geração do modelo em escala local, foram utilizados 200 pontos – 100 pontos referentes à área 2 e 100 pontos referentes à área 3 –, os quais haviam sido selecionadas aleatoriamente para compor a etapa de treinamento do algoritmo. O procedimento de seleção dos pontos foi o mesmo da construção dos modelos em área individual. Entretanto, dessa vez foram utilizados 160 pontos para treinamento e 40 para validação do modelo.

Para a extração dos valores de reflectância de superfície, adotou-se a mesma metodologia descrita na seção “Geração dos modelos de predição e controle de qualidade” do Capítulo 1.

Após a extração dos valores de reflectância de cada ponto, foi feita a seleção de atributos para selecionar quais bandas e índices de vegetação iriam compor o modelo. O critério adotado para tal foi o cálculo da importância de cada variável para o modelo, realizado por meio da ferramenta de seleção do software Weka, considerando a correlação como métrica de importância. Para calcular o score de cada variável, o software usa como padrão um algoritmo (escolhido pelo usuário) que se baseia no cálculo do RMSE% para o modelo, no qual é ponderado se há aumento ou diminuição do mesmo quando aquela variável é removida da composição. Desse modo, é possível ranquear os índices e bandas, de forma que, quanto maior seu score, maior é sua contribuição para a geração do modelo.

Posteriormente, foi feita a seleção de atributos dos índices e bandas. Essa estratégia foi escolhida para reduzir o sobreajuste do modelo em situações em que há um excesso de variáveis preditoras. Nesse contexto, foram realizados diversos testes para estabelecer qual o número ideal de parâmetros para compor o modelo, priorizando o menor número, sem alterar significativamente a variabilidade dos resultados. Os testes resultaram na construção de modelos, de forma a remover o índice/banda que houvesse menor importância na elaboração do próximo. Dessa forma, a priori, o modelo possuía 32 índices/bandas, posteriormente, 31 índices/bandas, e assim sucessivamente. Durante a criação de cada modelo, foi analisado o RMSE% (Fórmula 4) e os valores de produtividade estimados, para analisar o desempenho do modelo. Por fim, após os testes, os resultados apontaram para redução dos 32 índices/bandas para modelos compostos apenas por 5 índices/bandas, capazes de estimar a produtividade sem

comprometer a variabilidade dos valores estimados. Para essa etapa, foi utilizado o algoritmo de RF, tendo a correlação como a métrica de seleção dos atributos mais importantes.

Após a seleção das bandas, foram feitos testes para encontrar os melhores modelos entre os algoritmos de Support Vector Machine, Neural Net, Random Forest e Random Tree. Os parâmetros de entrada para a geração dos modelos foram os *default* do software Weka. Para cada área de estudo, foi gerado um modelo em cada algoritmo. Após isso, foram calculadas as métricas avaliativas RMSE% e MAPE, para analisar o desempenho de cada modelo, para que fosse selecionado o de maior precisão.

4.2.9 Comparação entre os valores de produtividade mensurados em campo e os estimados *in situ*

A metodologia utilizada foi a mesma aplicada na seção “Mapa de discrepância entre valor observado em campo e estimado pelo modelo” do Capítulo 1.

4.3 RESULTADOS E DISCUSSÕES

4.3.1 Resultados da Análise exploratória da variável produtividade

A Tabela 7 apresenta os valores descritivos para a variável produtividade, considerando cada uma das áreas de estudo. A Figura 14 apresenta os histogramas sobrepostos pela curva de distribuição normal de cada área.

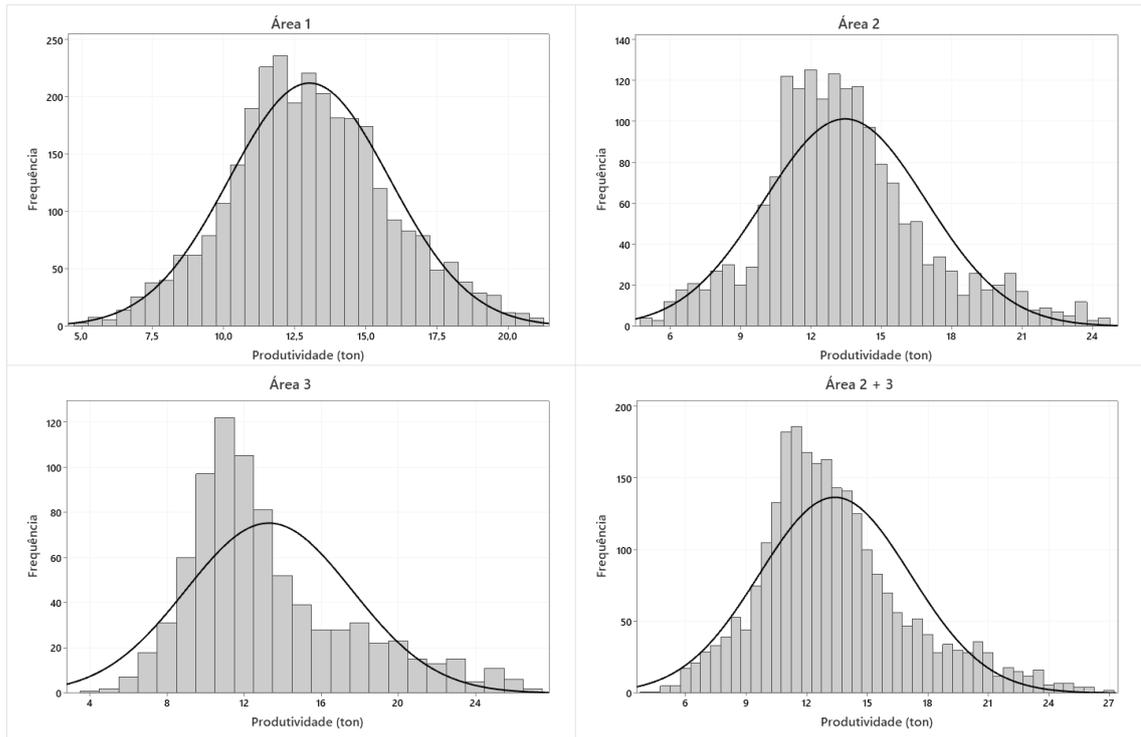


FIGURA 14. Histograma e curva de distribuição normal da área de estudo na lavoura do milho

TABELA 7. Análise exploratória da variável produtividade

Variável	Média (t/ha)	D. P. (t/ha)	C. V. (%)	Mínimo (t/ha)	Q1 (t/ha)	Mediana (t/ha)	Q3 (t/ha)	Máximo (t/ha)
Área1	13,02	2,82	21,66	4,86	11,19	12,91	14,82	21,10
Área2	13,45	3,45	25,66	4,91	11,25	13,13	15,22	24,66
Área3	13,29	4,31	32,46	4,11	10,40	12,12	15,35	26,90
Área2 + 3	13,40	3,74	27,96	4,11	10,99	12,82	15,23	26,90

Ao analisar a Tabela 7, é possível observar que as métricas utilizadas apresentam valores relativamente próximos em todas as áreas. A área que teve maior média de produtividade foi a área 2, com um total de 13,45 t/ha, seguida pela Área 3 (13,29 t/ha) e, por último, pela área 1 (13,02 t/ha). Segundo Conab (2021), a produtividade média do estado de Minas Gerais para a safra de 2020/2021 foi de 4,62 t/ha, o que garante que todas as áreas de estudo utilizadas tenham produzido mais do que a média do estado. Em uma análise conjunta, entre os valores de média e os gráficos dispostos na Figura 14, nota-se que a média apresenta valor inferior à mediana, o que é característico de curvas assimétricas positivas. Nesse contexto, a Área 3 é a que apresenta maior tendência para a esquerda no seu conjunto de dados, seguida pela Área 2 e 1.

A proximidade entre os valores de produtividade presente nas áreas de estudo fez com que a junção de duas delas (Área 2 + 3) ainda apresentasse um histograma com características similares aos valores individuais de cada área. O menor valor para coeficiente de variação pertence à Área 1 (21,66%), seguido pela Área 2 (25,66%), Área 2 + 3 (27,96%) e Área 3 (32,46%). Os valores de C.V. sugerem que as áreas utilizadas possuem baixa variabilidade e alta homogeneidade. Esse padrão pode ser constatado de forma mais clara quando analisados os histogramas supracitados juntamente com os valores de Q1 e Q3. Utilizando os quartis 1 e 3 como limiares inferiores e superiores, respectivamente, observa-se que 50% dos dados de produtividade estão contidos em intervalos pequenos, sendo estes: 3,63 t/ha para Área 1; 3,97 t/ha para Área 2; 4,95 t/ha para Área 3; e 4,24 t/ha para Área 2 + 3. A alta concentração de valores num intervalo pequeno pode fazer com que o modelo seja menos sensível aos valores de máximo e mínimo, uma vez que estes tendem a ter menor incidência dentro das áreas de estudo. Por sua vez, quando separadas as amostras para treinamento do modelo, esses valores apresentam menor probabilidade de serem selecionados.

4.3.2 Índices/bandas escolhidos para geração dos modelos e análise dos modelos individuais

A Figura 15 apresenta gráficos para cada área de estudo, considerando os valores de score para cada banda/índice.

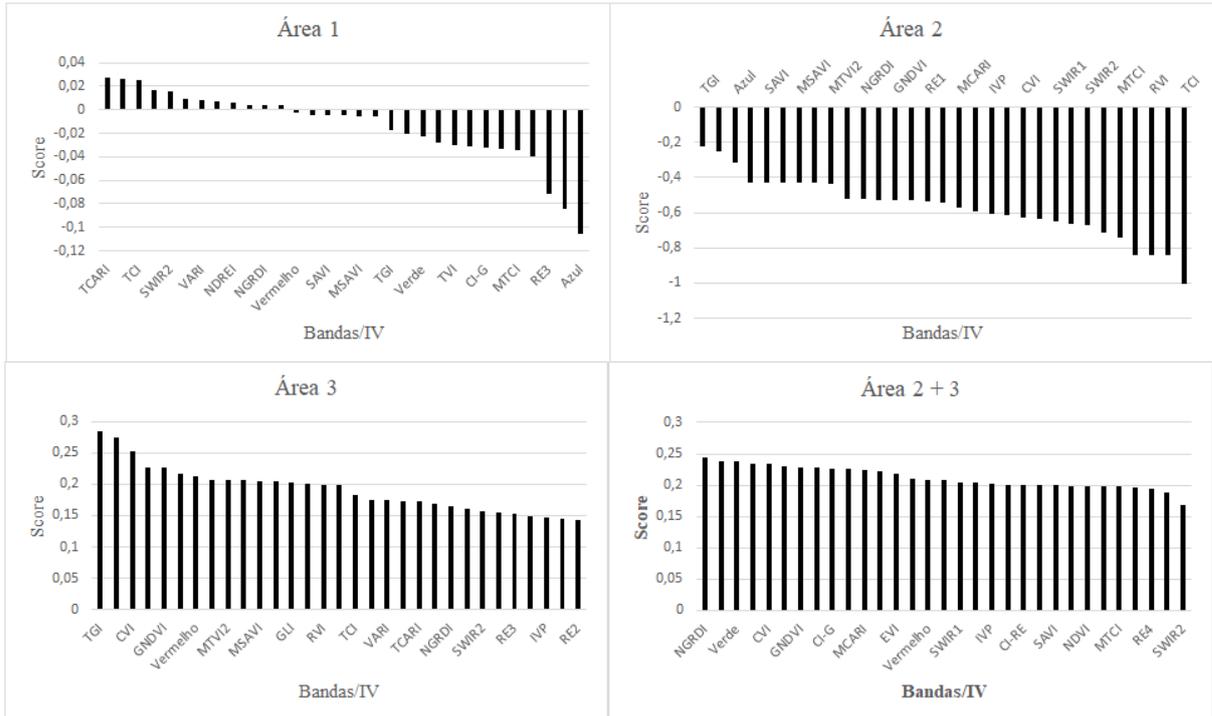


FIGURA 15. Importância da variável para geração do modelo em sua respectiva área de estudo

Os valores obtidos demonstram a importância de cada variável para a confecção do modelo utilizando um algoritmo específico. Ao todo, foram selecionados os cinco maiores valores para a construção do modelo. Dessa forma, tem-se: TCARI, TCI, SWIR2, VARI e NDREI para Área 1; TGI, Azul, SAVI, MSAVI e MTVI2 para Área 2; TGI, CVI, GNDVI, Vermelho e MTVI2 para Área 3; e NGRDI, Verde, CVI, GNDVI e CI-G para Área 2 + 3. Embora alguns índices se repitam entre as áreas de estudo, nota-se uma variedade na seleção deles para a construção de cada modelo, o que é um indicativo de que a sensibilidade dos índices depende do contexto específico de cada talhão, algo atrelado às variáveis macro e microbióticas que cercam aquela área.

A Tabela 8 apresenta os resultados de RMSE% e MAPE para os modelos gerados em cada área de estudo. Ao analisar os valores obtidos, nota-se que todos os valores são inferiores a 30% e 22%, respectivamente.

TABELA 8. Métricas avaliativas dos modelos gerados e suas respectivas bandas selecionadas

Área	Algoritmo	RMSE (%)	MAPE	IV/Bandas
Área 1	SVM	17,75	15,27	TCARI, TCI, SWIR2, VARI e NDREI
Área 2	RF	25,79	18,7	Azul, TGI, SAVI, MSAVI, MTVI2
Área 3	SVM	29,07	21,57	Vermelho, TGI, CVI, GNDVI, MTVI2
Área 2+3	SVM	27,86	20,01	Verde, CVI, GNDVI, CI-G, NGRDI

Diante dos valores mostrados pela Tabela 8, o melhor desempenho está associado à Área 1, com RMSE% e MAPE de 17,75% e 15,27%, respectivamente. O segundo menor valor de RMSE% e MAPE pertence ao modelo da Área 2, com os valores de 25,79% e 21,57%, respectivamente. Esses resultados corroboram o C.V. apresentado anteriormente na Tabela 7, o qual mostra que os menores valores estavam associados as áreas 1 e 2, respectivamente. Além disso, estas também são as áreas que apresentam o histograma com conjunto de dados mais simétrico. Para a Área 3, os valores foram de 29,07% para RMSE% e de 21,57% para MAPE, sendo estes os valores mais altos encontrados entre os modelos gerados. Para área 2 + 3, os resultados de 27,86% para RMSE% e de 20,01% MAPE mostram que o modelo gerado por meio dos dados das áreas 2 e 3 obteve um desempenho médio quando comparado às duas áreas de origem. Os resultados aos obtidos na Área 2 foram inferiores a área 1, porém superiores aos encontrados na Área 3.

Em linhas gerais, os valores de RMSE% e MAPE mostram o potencial das imagens advindas do satélite Sentinel 2 para a estimativa de produtividade na cultura do milho. Na tentativa de estimar produtividade mediante uma aeronave remotamente pilotada de baixo custo, Zhang et al. (2020) conseguem gerar modelos com MAPE entre 6,2% e 15,1% para a cultura do milho. Embora exista uma diferença entre os valores encontrados pelos autores e os dessa pesquisa.

Diante dos modelos gerados, o algoritmo com maior repetição foi o SVM. Esse algoritmo é utilizado na agricultura de precisão em diversos casos, entre eles, na estimativa de produtividade. Joshi et al. (2021) conseguem gerar modelos para a estimativa de produtividade de milho e soja, com RMSE% de 4,9% e 5,5%, respectivamente. Embora tenham obtido o melhor resultado em apenas um dos casos, o algoritmo de RF tem sido utilizado amplamente na estimativa de produtividade de culturas agrícolas (SEGARRA et al., 2022; RAMOS et al. 2020; SAKAMOTO, 2020).

Os cinco melhores IV e bandas selecionados em cada modelo variam de acordo com cada área de estudo. Dentre os selecionados, os índices TGI, MTVI2, GNDVI, CVI são os com maior incidência, todos aparecendo em pelo menos dois modelos diferentes. Os índices selecionados são comumente relacionados à estimativa de produtividade em diversas culturas. Para aprimorar o desempenho do algoritmo RF na estimativa de produtividade na cultura do milho, Ramos et al. (2020) aplicam o método de seleção de atributos (*feature selection*) para escolher os três melhores índices de vegetação entre 33 calculados. No trabalho, os autores apontam GNDVI como um dos três melhores contribuintes. García-Martinez et al. (2020) utilizam o índice TGI, entre outros, para a geração de modelos de estimativa de produtividade na cultura do milho e conseguem, assim, encontrar correlação de 0,77. Segarra et al. (2022) utilizam sete índices para estimar o rendimento de grãos de trigo, entre eles, GNDVI, CVI e TGI. No trabalho, os autores conseguem RMSE 0,74 t/ha, com um modelo baseado no algoritmo de RF.

A Figura 16 mostra a distribuição de pontos formados pelos valores observados em campo e os estimados pelos modelos.

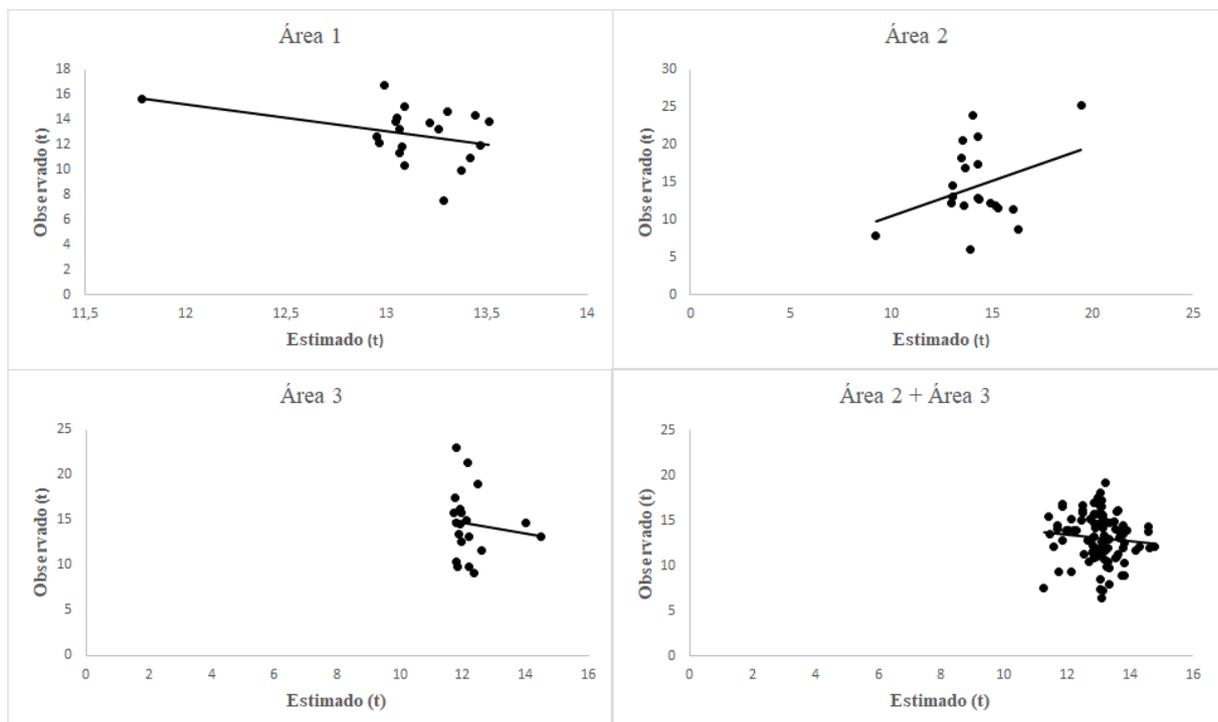


FIGURA 16. Gráficos entre valores observados e estimados

Os valores correspondentes ao eixo das abscissas são referentes aos valores estimados, enquanto o eixo das ordenadas possui os valores observados em campo. As Áreas 1, 3 e 2 + 3

apresentam uma linha de tendência decrescente, um coeficiente angular baixo e dados restritos a limiares de baixa amplitude. Em contrapartida, a Área 2 apresenta comportamento crescente e um conjunto de dados dispersos num intervalo de maior amplitude. Dessa forma, os modelos se mostraram adequados.

4.3.3 Análise da diferença entre os valores mensurados e estimados pelos modelos

A Figura 17 mostra a superfície de diferença entre os valores mensurados e os valores observados em campo.

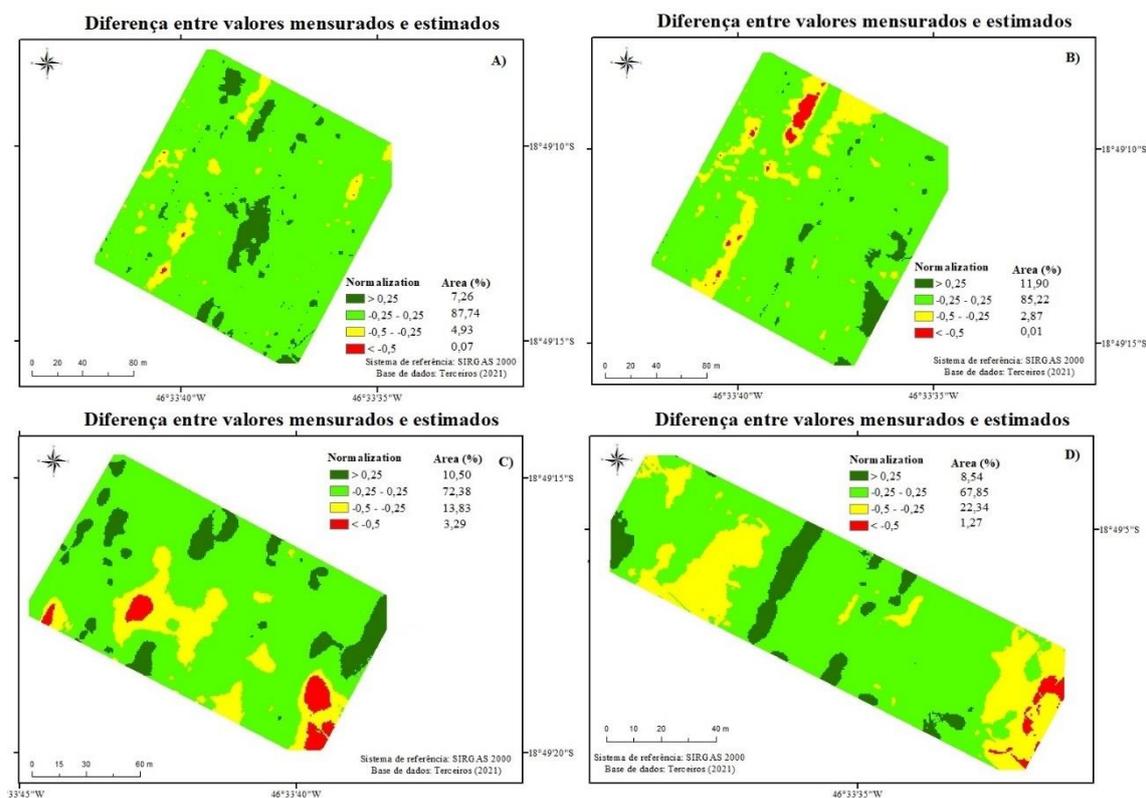


FIGURA 17. Diferença entre valores mensurados e estimados em campo A) Modelo Área 1; B) Modelo Área 1 local; C) Modelo Área 2; e D) Modelo Área 3

Os mapas de diferença entre os valores observados e estimado foram separados em 4 classes, sendo elas referentes aos seguintes intervalos: vermelho, diferença superior a -0,5; amarelo, diferença entre -0,5 e -0,25; verde, intervalo entre -0,25 e 0,25; e verde escuro, maior que 0,25. O comportamento das discrepâncias é proporcional em todas as áreas de estudo. Em

todos os casos, as diferenças entre os valores observados e estimados estão majoritariamente presentes entre os intervalos de -0,25 a 0,25, o que mostra que os modelos conseguem ter um desempenho consistente ao longo da maior parte da área ao qual é aplicado. Um fator que podem contribuir para isso é a sensibilidade da imagem às variações que ocorrem na cultura.

A fim de comparar as diferenças entre a superfície gerada pelo modelo local e o modelo específico para Área 1, foi gerada a Figura 18, que mostra o resultado da diferença entre as duas superfícies geradas.

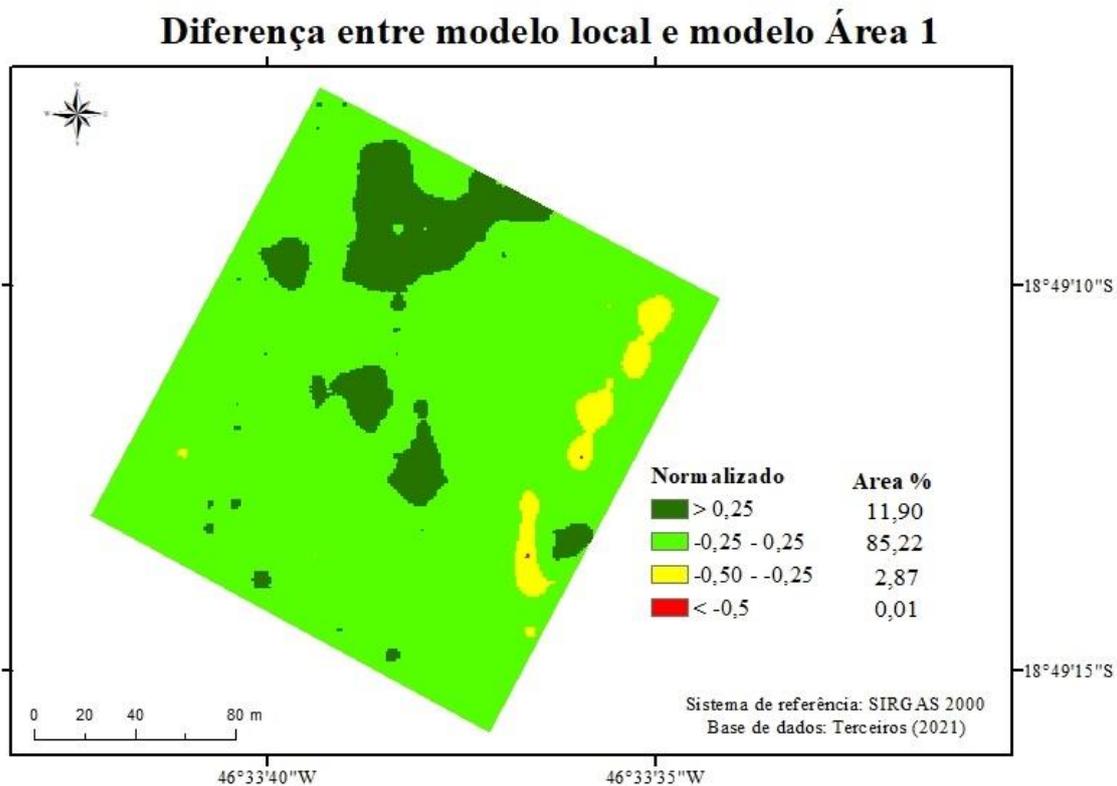


FIGURA 18. Mapa de diferença entre o modelo local e específico da Área 1

TABELA 9. Estatísticas das superfícies local e específica

Métricas analisadas	Área de estudo	
	Local	Específica
Média	13,03	13,07
Desvio Padrão	0,28	0,28
RMSE%	20,97	
MAPE	19,19	

As classes foram distribuídas conforme o padrão supracitado anteriormente. Ao comparar as superfícies geradas pelos modelos local e específico da Área 1, nota-se que 85,22%

das diferenças pertencem à classe com limiares $-0,25$ e $0,25$. Em linhas gerais, isso indica que o modelo local teve a capacidade de estimar os valores de produtividade para a Área 1 de forma similar ao calculado pelo modelo específico da própria área. A capacidade de estimar valores de produtividade de áreas com condições similares mediante um modelo mais robusto, utilizando imagens orbitais, apresenta grande potencial, uma vez que o monitoramento pode ser feito de forma precoce, já que o modelo independe dos valores de produtividade daquela área. O RMSE% e MAPE de $20,97\%$ e de $19,19\%$ respectivamente, mostram um desempenho superior ao calculado pelo modelo na etapa de treinamento do algoritmo.

4.4 CONSIDERAÇÕES FINAIS

Nossos modelos espectrais de estimativa da produtividade mostram que:

- O modelo de estimativa de produtividade local consegue resultados similares ao modelo específico da Área 1;
- Os modelos de estimativa local podem ser uma alternativa para a predição de produtividade, uma vez que o modelo é subsidiado por valores de produtividade de outras áreas daquele contexto local.

REFERÊNCIAS

- ADDAN, D. **Support Vector Machine. Unibrasil**, 2019. Disponível em: <<https://www.inf.ufpr.br/dagoncalves/IA07.pdf>>. Acesso em: 21 de junho de 2021.
- AHMAD, I. et al. Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. **Journal of the Indian society of remote sensing**, v. 46, n. 10, p. 1701-1711, 2018. <https://doi.org/10.1007/s12524-018-0825-8>
- AHMAD, I. et al. Remote sensing-based framework to predict and assess the interannual variability of maize yields in Pakistan using Landsat imagery. **Computers and Electronics in Agriculture**, v. 178, p. 105732, 2020. DOI: <https://doi.org/10.1016/j.compag.2020.105732>.
- AGHIGHI, H.; M. AZADBAKHT, D. ASHOURLOO, H. S. SHAHRABI, S. RADIOM. Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 OLI. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 12, p. 4563-4577, 2018. <https://doi.org/10.1109/JSTARS.2018.2823361>
- ARAVIND, K. R. et al. Disease classification in maize crop using bag of features and multiclass support vector machine. In: **2018 2nd International Conference on Inventive Systems and Control (ICISC)**, p. 1191-1196, 2018. <https://doi.org/10.1109/ICISC.2018.8398993>
- ALEHEGN, E. Ethiopian maize diseases recognition and classification using support vector machine. **International Journal of Computational Vision and Robotics**, v. 9, p. 90-109, 2019. <https://dx.doi.org/10.1504/IJCVR.2019.098012>
- ALMEIDA, L. F., Spers, E. E. Coffee Consumption and Industry Strategies in Brazil: A Volume in the Consumer Science and Strategic Marketing Series. **Woodhead Publishing**, 2019.
- AMARAL, F. C. S. Sistema Brasileiro de Classificação de Terras para Irrigação: enfoque na região semiárida. **Embrapa Solos-Livro técnico (INFOTECA-E)**, 2011.
- ARAB, S. T., NOGUCHI, R., MATSUSHITA, S., AHAMED, T. Prediction of grape yields from time-series vegetation indices using satellite remote sensing and a machine-learning approach. **Remote Sensing Applications: Society and Environment**, v. 22, p. 100485, 2021. <http://dx.doi.org/10.1016/j.rsase.2021.100485>.
- BAUM, M. E.; ARCHONTOULIS, S. V.; LICHT, M. A. Planting date, hybrid maturity, and weather effects on maize yield and crop stage. **Agronomy Journal**, v. 111, n. 1, p. 303-313, 2019. <https://doi.org/10.2134/agronj2018.04.0297>

BABU, S., R. SINGH, R. K. AVASTHE, G. S. YADAV, A. DAS, V. K. SINGH, K. P. MOHAPATRA, S. S. RATHORE, P. CHANDRA, A. KUMAR. Impact of land configuration and organic nutrient management on productivity, quality and soil properties under baby corn in Eastern Himalayas. **Scientific Reports**, v. 10, n. 1, p. 1-14, 2020.

<https://doi.org/10.1038/s41598-020-73072-6>

BERTOLIN, N. O.; FILGUEIRAS, R.; VENANCIO, L. P.; MANTOVANI, E. C. Predição da produtividade de milho irrigado com auxílio de satélite. **Rev. Bras. Agri. Irri.** v. 11, n. 4, Fortaleza, p. 1627-1638, Jul-Ago, 2017. <http://dx.doi.org/10.7127/rbai.v11n400567>

BROGE, N.H., LEBLANC, E. Comparing predictive power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. **Remote Sensing of Environment**, v. 76, p. 156–172, 2000.

[https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8)

BORGHI, E. et al. Sistemas de cultivo de soja e milho na região do Alto Paranaíba-MG e resultados de avaliações na safra 2014/15. **Embrapa Milho e Sorgo-Documentos** (INFOTECA-E), 2016.

BRUNNSCHWEILLER, Seasonal Changes in Agricultural Pattern - a Study in Comparative A.P.I. **P. Eng.** p. 131, 1957.

CARLEO, G. et al. Machine learning and the physical sciences. **Reviews of Modern Physics**, v. 91, 2019. <https://doi.org/10.48550/arXiv.1903.10563>

CIAMPITTI, I. A.; ELMORE, R. W.; LAUER, J. **FASES DO DESENVOLVIMENTO DA CULTURA DO MILHO**. Kansas State University Agricultural Experiment Station and Cooperative Extension Service, 2016. Disponível em:

< [https://www.npct.com.br/npctweb/npct.nsf/article/BRS-3137/\\$File/MF3305BP-CornGrowth-portuguese_FINAL.pdf](https://www.npct.com.br/npctweb/npct.nsf/article/BRS-3137/$File/MF3305BP-CornGrowth-portuguese_FINAL.pdf) >, acesso em: 02/2/2022.

CHEN, Y., Z. ZHANG, F. TAO. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. **European journal of agronomy**, v. 101, p. 163-173, 2018. <https://doi.org/10.1016/j.eja.2018.09.006>

CONAB. **Acompanhamento da safra brasileira de grãos: décimo segundo levantamento, setembro 2020, safra 2019/2020**, 2020. Acesso: 11, ago, 2021, Disponível:

<<https://www.conab.gov.br/infoagro/safras/grãos>>.

CONAB. **Acompanhamento da safra brasileira de grãos: décimo segundo levantamento, setembro 2021, safra 2020/2021**, 2021. Acesso: 19, set, 2021, Disponível:

<<https://www.conab.gov.br/infoagro/safras/grãos>>.

CHLINGARYAN, A., SUKKARIEH, S., WHELAN, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. **Computers and electronics in agriculture**, v. 151, p. 61-69, 2018.

<https://doi.org/10.1016/j.compag.2018.05.012>

DASH, J., CURRAN, P. J. The MERIS terrestrial chlorophyll index. **International Journal of Remote Sensing**, v. 25, p. 5403–5413, 2004.

<https://doi.org/10.1080/0143116042000274015>

DAUGHTRY, C.S.T.; WALTHALL, C.L.; KIM, M.S.; BROWN DE COLSTOUN, E., MCMURTREY, J.E. Estimating corn leaf chlorophyll concentration for leaf and canopy reflectance. **Remote Sensing of Environment** v. 74, p. 229–239, 2000.

[https://doi.org/10.1016/S0034-4257\(00\)00113-9](https://doi.org/10.1016/S0034-4257(00)00113-9)

DE MELO, R. W., FONTANA, D. C., BERLATO, M. A., DUCATI, J. R. An agrometeorological–spectral model to estimate soybean yield, applied to southern Brazil. **International Journal of Remote Sensing**, v. 29, n. 14, p. 4013-4028, 2008.

<https://doi.org/10.1080/01431160701881905>

DILL, H. W. The Use of the Comparison Method in Agricultural A.P.I. **P. Eng.**, p. 49, 1959.

DOKIC, K.; BLASKOVIC, L.; MANDUSIC, D. D. From machine learning to deep learning in agriculture—the quantitative review of trends. In **IOP Conference Series: Earth and Environmental Science**, v. 614, n. 1, p. 012138, 2020. <http://dx.doi.org/10.1088/1755-1315/614/1/012138>

<http://dx.doi.org/10.1088/1755-1315/614/1/012138>

ESA (European Space Agency). **Science Toolbox Exploitation Platform**. 2018. Disponível em: < <https://step.esa.int/main/snap-supported-plugins/sen2cor/> >. Acesso em: 12/10/2021

ESTES, J. Some Applications of Aerial IR Imagery, **Annals AA.G**, p. 637, 1966.

<https://doi.org/10.1111/j.1467-8306.1966.tb00584.x>

EUGENIO, F. C.; GROHS, M.; VENANCIO, L. P.; SCHUH, M.; BOTTEGA, E. L.; RUOSO, R.; SCHONS, C.; MALLMANN, C. L.; BADIN, T. L.; FERNANDES, P. Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. **Remote Sensing Applications: Society and Environment**, v. 20, p. 100397, 2020.

<https://doi.org/10.1016/j.rsase.2020.100397>

ERENSTEIN, O., CHAMBERLIN, J.; SONDER, K. Estimating the global number and distribution of maize and wheat farms. **Global Food Security**, V. 30, p.100558, 2021.

<https://doi.org/10.1016/j.gfs.2021.100558>

FAN, J., ZHENG, J., WU, L., ZHANG, F. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. **Agricultural Water Management**, v. 245, p. 106547, 2021.

<https://doi.org/10.1016/j.agwat.2020.106547>

FATHIPOOR, H., H. AREFI, R. SHAH-HOSSEINI, H. Moghadam. Corn forage yield prediction using unmanned aerial vehicle images at mid-season growth stage. **Journal of Applied Remote Sensing**, v. 13, n. 3, p. 034503, 2019.

<http://dx.doi.org/10.1117/1.JRS.13.034503>

FOSTER, F. H. Field Use of Aerial Photographs by Geographers, **P. Eng.**, p. 771, 1951.

GARCÍA-MARTÍNEZ, H. et al. Corn Grain Yield Estimation from Vegetation Indices, Canopy Cover, Plant Density, and a Neural Network Using Multispectral and RGB Images Acquired with Unmanned Aerial Vehicles. **Agriculture**, v. 10, n. 7, p. 277, 2020. <https://doi.org/10.3390/agriculture10070277>

GATES, P. Characteristics of Soil and Vegetated Surfaces to Reflected and Emmitted Radiation, **Proc. 3rd in R.S. of Environment**, Ann Arbor, Mich.: IR Physics Lab, 1965.

GUIMARÃES, T. T. et al. Evaluation of regression analysis and neural networks to predict total suspended solids in water bodies from unmanned aerial vehicle images. **Sustainability**, v. 11, n. 9, p. 2580, 2019. <https://doi.org/10.3390/su11092580>

GUO, Y., H. WANG, Z. WU, S. WANG, H. SUN, J. SENTHILNATH, J. WANG, C. R. BRYANT, Y. FU. Modified red blue vegetation index for chlorophyll estimation and yield prediction of maize from visible images captured by UAV. **Sensors**, v. 20, n. 18, p. 5055, 2020. <https://doi.org/10.3390/s20185055>

GITELSON, A.A., KAUFMAN, Y.J., MERZLYAK, M.N. USE of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment**, v. 58, p. 289–298, 1996. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)

GITELSON, A. A., GRITZ, Y., MERZLYAK, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for nondestructive chlorophyll assessment in higher plant leaves. **Journal of Plant Physiology**, v. 160, n. 3, p. 271–282, 2003. <https://doi.org/10.1078/0176-1617-00887>

GOODMAN, N. S. Technique for the Identification of Farm Crops on Aerial Photographs. **P. Eng.**, p. 131, 1959.

HUETE, A., DIDAN, K., MIURA, T., RODRIGUEZ, E.P., GAO, X., FERREIRA, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, p. 195–213, 2002. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)

HUETE, A.R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, p. 295–309, 1988. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

HABOUDANE, D., MILLER, J.R., PATTEY, E., ZARCO-TEJADA, P.J., STRACHAN, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. **Remote Sensing of Environment**, v. 90, p. 337–352, 2004. <https://doi.org/10.1016/j.rse.2003.12.013>

HABOUDANE, D., MILLER, J.R., TREMBLAY, N., ZARCO-TEJADA, P.J., DEXTRAZE, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. **Remote Sensing of Environment**, v. 81, p. 416–426, 2002. <https://doi.org/10.1016/j.rse.2003.12.013>

HUNT, E.R., DAUGHTRY, C.S.T., EITEL, J.U.H., LONG, D.S. Remote sensing leaf chlorophyll content using a visible band index. **Agronomy Journal**, v. 103 n. 4, p. 1090–1099, 2011. <https://doi.org/10.2134/agronj2010.0395>

ISLAM, M. M., MATSUSHITA, S., NOGUCHI, R., AHAMED, T. Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. **Remote Sensing Applications: Society and Environment**, v. 22, p. 100494, 2021. <https://doi.org/10.1016/j.rsase.2021.100494>

JEŠOVNIK, A.; I. BLAŽEVIĆ, D.; LEMIC, I. PAJAČ ŽIVKOVIĆ. Ant fauna of annual and perennial crops. **Appl. Ecol. Environ. Res**, v. 17, p. 12709-12722, 2019. http://dx.doi.org/10.15666/aer/1706_1270912722

JENSEN, J.R. **Sensoriamento remoto do ambiente: Uma perspectiva em recursos terrestres**. Parêntese, São José dos Campos, 598 p, 2009.

Jl, Z., PAN, Y.; ZHU, X., WANG. J.; LI, Q. Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. **Sensors**, v. 21, n. 4, p. 1406, 2021. <https://doi.org/10.3390/s21041406>.

JOSHI, V. R.; KAZULA, M. J.; COULTER, J. A.; NAEVE, S. L.; GARCIA, A. G. IN-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. **International journal of biometeorology**, v. 65, n. 4, p. 489-502, 2021. <https://doi.org/10.1007/s00484-020-02039-z>

JORDAN, C.F. Derivation of leaf area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663–666, 1969. <https://doi.org/10.2307/1936256>

KHANAL, S. ET al. Assessing the impact of agricultural field traffic on corn grain yield using remote sensing and machine learning. **Soil And Tillage Research**, v. 208, p. 104880, 2021. <https://doi.org/10.1016/j.still.2020.104880>

KHANAL, S.; FULTON, J.; SHEARER, S. An overview of current and potential applications of thermal remote sensing in precision agriculture. **Computers and Electronics in Agriculture**, v. 139, p. 22-32, 2017. <https://doi.org/10.1016/j.compag.2017.05.001>

KHAKI, S., H. PHAM, L. WANG. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. **Scientific Reports**, v. 11, n. 1, p. 1-14, 2021. <https://doi.org/10.1038/s41598-021-89779-z>

KOUADIO, L.; DEO, R. C.; BYRAREDDY, V.; ADAMOWSKI, J. F.; MUSHTAQ, S. 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. <https://doi.org/10.1016/j.compag.2018.10.014>

KOVÁCS, Z. **Redes Neurais Artificiais**. 4 ed. Editora: Livradia da Física. São Paulo, 2006.

- LIESENBERG, V., GALVÃO, L. S.; PONZONI, F. J. Variations in reflectance with seasonality and viewing geometry: implications for classification of Brazilian savanna physiognomies with MISR/Terra data. **Remote Sensing of Environment**, v. 107, n. 1-2, p. 276-286, 2007. <https://doi.org/10.1016/j.rse.2006.03.018>
- LOUHAICHI, M., BORMAN, M.M., JOHNSON, D.E. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. **Geocarto International** v. 16, p. 65–70, 2001. <https://doi.org/10.1080/10106040108542184>
- MELO, R. W., FONTANA, D. C., BERLATO, M. A., DUCATI, J. R. An agrometeorological–spectral model to estimate soybean yield, applied to southern Brazil. **International Journal of Remote Sensing**, v. 29, n. 14, p. 4013-4028, 2008. <https://doi.org/10.1080/01431160701881905>
- MYNENI, R., HALL, R., SELLERS, P., MARSHAK, A. The interpretation of spectral vegetation indexes. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, p. 481-486, 1995. <https://doi.org/10.1109/TGRS.1995.8746029>
- MEKONNEN, Y., NAMUDURI, S., BURTON, L., SARWAT, A., BHANSALI, S. Machine learning techniques in wireless sensor network based precision agriculture. **Journal of the Electrochemical Society**, v. 167, n. 3, p. 037522, 2019. <http://dx.doi.org/10.1149/2.0222003JES>
- MAGALHÃES, P. C., DURÃES, F., O. M. Fisiologia da produção de milho (Physiology of corn production). **Circular Técnica** (INFOTECA-E), 2006.
- LEROUX, L. et al. Using remote sensing to assess the effect of trees on millet yield in complex parklands of Central Senegal. **Agricultural Systems**, v. 184, p. 102918, 2020. <https://doi.org/10.1016/j.agsy.2020.102918>
- LOBELL, D.B.; ORTIZ-MONASTERIO, J.I.; ASNER, G.P.; NAYLOR, R.L.; FALCON, W.P. Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. **Agronomy Journal**, v. 97, n. 1, p. 241-249, 2005. <https://doi.org/10.2134/agronj2005.0241a>
- MAIMAITIJIANG, M.; SAGAN, V.; SIDIKE, P.; HARTLING, S.; ESPOSITO, F.; FRITSCHI, F. B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote sensing of environment**, v. 237, p. 111599, 2020. <https://doi.org/10.1016/j.rse.2019.111599>
- MA, Y., ZHANG, Z., KANG, Y., ÖZDOĞAN, M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, 259, 112408, 2021. <https://doi.org/10.1016/j.rse.2021.112408>
- OLIVEIRA, A. S.; MELLO, C. M.; MARQUES, R. F. P. V. Temporal trends of climate indices associated with precipitation and air temperature in Minas Gerais, Brazil. **Revista Brasileira de Climatologia**, v. 26, 2020. <http://dx.doi.org/10.5380/abclima.v26i0.71770>

PHILPOTTS, V. Wallen, IR for Crop Disease Identification, **P. Eng.**, p. 1116, 1969.

PÁDUA, L., MARQUES, P., HRUŠKA, J., ADÃO, T., PERES, E., MORAIS, R. AND SOUSA, J.J. Multi-temporal vineyard monitoring through UAV-based RGB imagery. **Remote Sensing**, v. 10, n. 12, p.1907, 2018. <https://doi.org/10.3390/rs10121907>

PANG, L.; SHEN, S.; YAO, M. A Spline-High Dimensional Model Representation for SRAM Yield Estimation in High Sigma and High Dimensional Scenarios. **IEEE Access**, v. 9, p. 47320-47329, 2021. <https://doi.org/10.1109/ACCESS.2021.3067510>

PAUL, G. C., SAHA, S. HEMBRAM, T. K. Application of phenology-based algorithm and linear regression model for estimating rice cultivated areas and yield using remote sensing data in Bansloi River Basin, Eastern India. **Remote Sensing Applications: Society and Environment**, v. 19, p. 100367, 2020. <https://doi.org/10.1016/j.rsase.2020.100367>

PANDA, S. S.; AMES, D. P.; PANIGRAHI, S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. **Remote Sensing**, v. 2, n. 3, p. 673-696, 2010. <http://dx.doi.org/10.3390/rs2030673>

PANTAZI, X.E.; MOSHOU, D.; ALEXANDRIDIS, T.; WHETTON, R.L.; MOUAZEN, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. **Computers and Electronics in Agriculture**, v. 121, p. 57–65, 2016. <https://doi.org/10.1016/j.compag.2015.11.018>

PLANET. **Planet Imagery Product Specifications**. Planet Labs, 2021.

PLANET. **Planet Labs Specifications: Spacecraft Operations & Ground Systems**. Planet Labs, 2015.

PEARSON, R.L., MILLER, L.D. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. In: **Proceedings of the Eighth International**, p. 2-6, 1972.

QI, J., CHEHBOUNI, A., HUETE, A.R., KERR, Y.H., SOROOSHIAN, S. A modified soil adjusted vegetation index. **Remote Sensing Environment**, v. 48, p. 119–126, 1994. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1)

RAMOS, A. P. M., L. P. OSCO, D. E. G. FURUYA, W. N. GONÇALVES, D. C. SANTANA, L. P. R. TEODORO, C. A. DA SILVA JUNIOR et al. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. **Computers and Electronics in Agriculture**, v. 178, p. 105791, 2020. <https://doi.org/10.1016/j.compag.2020.105791>

RAUBER, T. **Redes Neurais Artificiais**. Universidade Federal do Espírito Santo, v. 29, 2005.

RAY, K., H. BANERJEE, S. DUTTA, A. K. HAZRA, K. MAJUMDAR. Macronutrients influence yield and oil quality of hybrid maize (*Zea mays* L.). **PloS one**, v. 14, n. 5, p. 0216939, 2019. <https://doi.org/10.1371/journal.pone.0216939>

RYERSON, R. A. **A Model to Provide a Measure of Agricultural Productivity Using Remote Sensing Techniques**. PhD thesis. 1970.

RIQUETI, G.; RIBEIRO, C.; ZÁRATE, L. Classificando perfis de longevidade de bases de dados longitudinais usando Floresta Aleatória. **Conference Paper**, 2018.

ROUSE, J. W., HAAS, R. H., SCHELL, J. A., DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. **NASA Special Publication**, v. 351, p. 309-317, 1974.

RONDEAUX, G., STEVEN, M., BARET, F., Optimization of soil-adjusted vegetation indices. **Remote Sensing of Environment**, v. 55, p. 95–107, 1996.
[https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7)

SEGARRA, J.; ARAUS, J. L.; KEFAUVER, S. C. Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. **International Journal of Applied Earth Observation and Geoinformation**, v. 107, p. 102697, 2022.
<https://doi.org/10.1016/j.jag.2022.102697>

SAKAMOTO, T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 160, p. 208-228, 2020. <https://doi.org/10.1016/j.isprsjprs.2019.12.012>

SHANAHAN, J. Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. **Agronomy Journal**, v. 93, n. 3, p. 583-589, 2001. <https://doi.org/10.2134/agronj2001.933583x>

SELVI, K. C.; GÜLSER, C.; BEYHAN, M. A. Short Term Effects of Different Tillage Methods on Nitrate Content in Soil and Corn Yield. **Malaysian Journal of Soil Science**, v. 23, p. 55-68, 2019.

SOLOGUREN, L. Milho: Brasil amplia cultivo para atender demanda crescente. **Visão Agrícola**, 2015.

SUMESH, K.C., NINSAWAT, S.; SOM-ARD, J. Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. **Computers and Electronics in Agriculture**, v. 180, p.105903, 2021. <https://doi.org/10.1016/j.compag.2020.105903>

SUN, J.; LAI, Z.; DI, L.; SUN, Z.; TAO, J.; SHEN, Y. Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt, in **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 5048-5060, 2020, <https://doi.org/10.1109/JSTARS.2020.3019046>

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

TREBOUX, J., GENOUD, D. Improved machine learning methodology for high precision agriculture. **2018 Global Internet of Things Summit (GIoTS)**, p. 1-6, 2018.

<https://doi.org/10.1109/GIOTS.2018.8534558>

TAGARAKIS, A., KETTERINGS, Q. In-season estimation of corn yield potential using proximal sensing. **Agronomy Journal**, v. 109, n. 4, p. 1323-1330, 2017.

<https://doi.org/10.2134/agronj2016.12.0732>

VINCINI, M., FRAZZI, E., D'Alessio, P. A broad-band leaf chlorophyll index at the canopy scale. **Precision Agriculture**, v. 9, p. 303–319, 2008. <http://dx.doi.org/10.1007/s11119-008-9075-z>

WANG, L., WANG, P., LIANG, S., ZHU, Y., KHAN, J., FANG, S. Monitoring maize growth on the North China Plain using a hybrid genetic algorithm-based back-propagation neural network model. **Computers and Electronics in Agriculture**, v. 170, p. 105238, 2020.

<https://doi.org/10.1016/j.compag.2020.105238>

WAHAB, I., HALL, O. JIRSTRÖM, M. Remote sensing of yields: Application of uav imagery-derived ndvi for estimating maize vigor and yields in complex farming systems in sub-saharan africa. **Drones**, v. 2, n. 3, p. 28, 2018. <https://doi.org/10.3390/drones2030028>

ZHANG, M.; ZHOU, J.; SUDDUTH, K. A.; KITCHEN, N. R. Estimation of maize yield and effects of variable-rate nitrogen application using UAV-based RGB imagery. **biosystems engineering**, v. 189, p. 24-35, 2020. <https://doi.org/10.1016/j.biosystemseng.2019.11.001>

ZHENG, J., FAN, J., ZHANG, F., WU, L., ZOU, Y., ZHUANG, Q. Estimation of rainfed maize transpiration under various mulching methods using modified Jarvis-Stewart model and hybrid support vector machine model with whale optimization algorithm. **Agricultural Water Management**, v. 249, p. 106799, 2021. <https://doi.org/10.1016/j.agwat.2021.106799>