

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

JÚLIA FERREIRA DE ALCÂNTARA

**PREDIÇÃO DO DESEMPENHO FISIOLÓGICO E DE CO₂ EM
SOJA: UMA ABORDAGEM UTILIZANDO SENSORIAMENTO
HIPERESPECTRAL E APRENDIZAGEM DE MÁQUINA.**

CHAPADÃO DO SUL – MS

2025

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CAMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

JÚLIA FERREIRA DE ALCÂNTARA

**PREDIÇÃO DO DESEMPENHO FISIOLÓGICO E DE CO₂ EM
SOJA: UMA ABORDAGEM UTILIZANDO SENSORIAMENTO
HIPERESPECTRAL E APRENDIZAGEM DE MÁQUINA.**

Orientador(a): Prof (a). Dr(a) Larissa Pereira Ribeiro Teodoro

Qualificação apresentada à Universidade Federal de Mato Grosso do Sul, como requisito para obtenção do título de Mestre em Agronomia, área de concentração: Produção Vegetal.

CHAPADÃO DO SUL – MS

2025



Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul



PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CERTIFICADO DE APROVAÇÃO

DISCENTE: Júlia Ferreira de Alcântara

ORIENTADORA: Dra. Larissa Pereira Ribeiro Teodoro

TÍTULO: Predição do desempenho fisiológico e de CO₂ em soja: uma abordagem utilizando sensoriamento hiperespectral e aprendizagem de máquina.

AVALIADORES:

Prof. Dra. Larissa Pereira Ribeiro Teodoro

Prof. Dr. Cid Naudi Silva Campos

Prof. Dr. Francisco Eduardo Torres

Chapadão do Sul, 10 de fevereiro de 2025.

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Larissa Pereira Ribeiro, Professora do Magistério Superior**, em 10/02/2025, às 10:37, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Cid Naudi Silva Campos, Professor do Magisterio Superior**, em 10/02/2025, às 19:47, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA
MÁXIMA
NO MEC

UFMS
É 10!!!



Documento assinado eletronicamente por **Francisco Eduardo Torres, Usuário Externo**, em 11/03/2025, às 15:28, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufms.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5420077** e o código CRC **41FC93BF**.

COLEGIADO DE PÓS-GRADUAÇÃO EM AGRONOMIA

Rod MS 306, Km 105, Caixa Postal 112

Fone: (67)3562-6351

CEP 79560-000 - Chapadão do Sul - MS

Referência: Processo nº 23455.000345/2024-04

SEI nº 5420077

DEDICATÓRIA

Ao excelso Pierre Curie

AGRADECIMENTOS

Para chegar até aqui precisei de 24 anos e muitas dádivas, dentre elas:

Meus avós: Valderes e Wolney, que foram meus tutores, zelaram pelo meu aprendizado desde muito pequena, proporcionaram-me qualidade de estudo e apoio à carreira. Minha mãe: Magali, que tornara os meus estudos uma diversão de cores e estímulo. Ela assistiu a todos os meus inícios, e fora corajosa ao me impulsionar em cada um deles. Meus tios e primos: Sergio, Eliene, Wolney Neto, Stephanie, Alice, Enzo e Helena, que participaram intimamente de minha formação pessoal também tiveram suas contribuições ao estudo, apoiando-me de diversos modos. Família, essa conquista é nossa! Aos dez anos prestei pela primeira vez a Olimpíada Brasileira de Matemática das Escolas Públicas, aqui agradeço veemente ao Instituto de Matemática Pura e Aplicada, pois isso mudou toda a minha história. A partir daí vieram diversas conquistas, não apenas em Matemática: agradeço à Sociedade Astronômica Brasileira e ao Instituto Ayrton Senna, que me proporcionaram olimpíadas, viagens, aprendizado e muita diversão. Agradeço ao meu pai: Nilson, pela iniciativa em investir na qualidade dos meus estudos, o que fez com que eu migrasse para um outro universo educacional e aumentasse exponencialmente o meu interesse pela Ciência. Agradeço às escolas EMEF Massami Emoto, E.E. Dr. Pio Antunes de Figueiredo e Extensão, e todos os seus funcionários pelo período de minha passagem. Agradeço ao Sistema de Seleção Unificada e à educação pública que me proporcionaram um ensino superior de excelência pela Fundação Universidade Federal de Mato Grosso do Sul, tal como ao Programa de Pós-Graduação em Agronomia da UFMS/CPCS. Agradeço a cada funcionário do CPCS, em especial àqueles que me tocaram a alma: a maioria dos Docentes. Muitíssimo obrigada, professora Larissa, pela sua orientação ao longo desses seis anos, obrigada professor Paulo por toda a confiança, esperança e apoio, sem vocês definitivamente eu não estaria concluindo este trabalho. Agradeço à CAPES, pelo financiamento de minha pesquisa. Agradeço aos amigos que fizeram da minha caminhada uma grande aventura repleta de boas lembranças: Pricila B., Cristina, Gustavo, Thayriq, Rubinaldo, Marcos E., Marcos P., Randerson, Juliano, Jorge, Priscila L., André, Matheus, Gabriela, Gabriel, Luana, Netto, Vitória, Mérite e Konai. Agradeço à espiritualidade amiga!

EPÍGRAFE

“Somos um só povo e temos uma única Terra”

Organização das Nações Unidas

LISTA DE FIGURAS

Figura 1. Dados diários de pluviosidade e temperatura média da safra 2023/24 na cidade de Chapadão do Sul/MS.

Figura 2. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável A (fotossíntese líquida).

Figura 3. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável gs (condutância estomática).

Figura 4. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável Ci (concentração interna de CO_2).

Figura 5. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável E (transpiração).

Figura 6. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável WUE (eficiência instantânea de uso de água).

Figura 7. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável ICE (eficiência instantânea de carboxilação).

Figura 8. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável emissão in situ de CO_2 do solo (FCO_2).

Figura 9. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável Fixação de CO pelo tecido foliar.

RESUMO

A fome e as mudanças climáticas assolam o planeta, e ações que visam suas mitigações têm sido o foco de entidades governamentais e centros de pesquisa. Técnicas que associam esses dois fatores são fundamentais para a vida e alimentação. As ciências ambientais associadas ao melhoramento da soja, à agricultura de precisão e à aprendizagem de máquinas (ML) são uma alternativa para esta agrura. Este trabalho objetiva prever variáveis fisiológicas, emissão in situ de dióxido de carbono do solo (FCO_2) e fixação de carbono no tecido foliar em genótipos de soja por meio de variáveis hiperespectrais e ML, tal como identificar os melhores algoritmos. Foi empregado delineamento de blocos casualizados com quatro repetições. As parcelas consistiram em cinco linhas de cinco metros com espaçamento de 0,45 m entre fileiras. As avaliações foram de: *fotossíntese líquida*; *condutância estomática*; *concentração interna de CO_2* ; *transpiração*; *eficiência instantânea de uso de água*; *eficiência instantânea de carboxilação*; análise espectral da planta; fluxo de dióxido de carbono, temperatura e umidade do solo in loco, e fixação de carbono no tecido foliar. Os modelos de ML utilizados foram: Redes neurais artificiais; Árvore de decisão REPTree; Árvore de decisão (M5P); Floresta aleatória (RF); Máquina de vetor suporte; e zero R. Os parâmetros utilizados foram: correlação de Pearson (r), erro médio absoluto e raiz quadrada do erro médio. É possível prever *fotossíntese líquida* com r acima de 0,75, o que é ótimo, e as demais variáveis fisiológicas permeando resultados médios. Já o fluxo e a fixação de carbono expressaram resultados insatisfatórios, o que requer mais estudos nessas previsões. As melhores técnicas de ML para essa abordagem são: RF e M5P.

PALAVRAS-CHAVE: Fisiologia vegetal; Machine learning; Fluxo de carbono.

ABSTRACT

Hunger and climate change plague the planet, and actions aimed at mitigating them have been the focus of government entities and research centers. Techniques that combine these two factors are fundamental for life and nutrition. Environmental sciences associated with soybean breeding, precision agriculture and machine learning (ML) are an alternative to this hardship. This work aims to predict physiological variables, in situ emission of soil carbon dioxide (FCO₂) and carbon fixation in leaf tissue in soybean genotypes through hyperspectral variables and ML, as well as identifying the best algorithms. A randomized block design with four replications was used. The plots consisted of five rows of five meters with a spacing of 0.45 m between rows. The assessments were: *net photosynthesis*; *stomatal conductance*; *internal CO₂ concentration*; *perspiration*; *instantaneous water use efficiency*; *instantaneous carboxylation efficiency*; plant spectral analysis; carbon dioxide flow, in situ soil temperature and humidity, and carbon fixation in leaf tissue. The ML models used were: Artificial neural networks; REPTree decision tree; Decision tree (M5P); Random Forest (RF); Support vector machine; and zero R. The parameters used were: Pearson correlation (r), mean absolute error and square root of the mean error. It is possible to predict *net photosynthesis* with r above 0.75, which is excellent, and the other physiological variables permeate average results. Carbon flow and carbon fixation showed unsatisfactory results, which requires further studies on these predictions. The best ML techniques for this approach are: RF and M5P.

KEYWORDS: Plant physiology; Machine learning; Carbon flow.

SUMÁRIO

AGRADECIMENTOS	6
EPÍGRAFE	7
LISTA DE FIGURAS	8
RESUMO	9
ABSTRACT	10
CAPÍTULO 1:	12
PREDIÇÃO DO DESEMPENHO FISIOLÓGICO EM SOJA.....	12
RESUMO	12
ABSTRACT	13
INTRODUÇÃO	14
MATERIAL E MÉTODOS	16
Condução do experimento a campo.....	16
Variáveis fisiológicas	17
Variáveis hiperespectrais	17
Aprendizagem de máquina	18
RESULTADOS E DISCUSSÃO	19
CONCLUSÕES.....	28
REFERÊNCIAS.....	29
CAPÍTULO 2:	34
PREDIÇÃO DA FIXAÇÃO E EMISSÃO DE CARBONO EM SOJA	34
RESUMO	34
ABSTRACT	35
INTRODUÇÃO	36
MATERIAL E MÉTODOS	38
Condução do experimento a campo.....	38
Fluxo de CO ₂	39
Aprendizagem de máquina	40
RESULTADOS E DISCUSSÃO	42
CONCLUSÕES.....	46

CAPÍTULO 1:

PREDIÇÃO DO DESEMPENHO FISIOLÓGICO EM SOJA

RESUMO

A soja é uma das culturas agrícolas mais populares por diversas culturas e tempos da humanidade. A preocupação no estudo de sua fisiologia remete-se ao desenvolvimento técnico e científico com a finalidade de superar a problemática da fome para além dos fatores de mercado. Entender claramente os mecanismos biológicos dos vegetais é necessário para se desenvolver manejos adequados e que obtenham maior produtividade, ainda mais neste cenário de mudanças climáticas e fenômenos naturais frequentes. Para isso, objetiva-se aqui prever variáveis fisiológicas por meio de variáveis hiperespectrais e utilizando algoritmo de aprendizagem de máquina (ML), tal como identificar os melhores algoritmos. Foi empregado delineamento de blocos casualizados com quatro repetições. As parcelas consistiram em cinco linhas de cinco metros com espaçamento de 0,45 m entre fileiras. As avaliações foram de: fotossíntese líquida; condutância estomática; concentração interna de CO₂; transpiração; eficiência instantânea de uso de água; eficiência instantânea de carboxilação; e análise espectral da planta. Os modelos de ML utilizados foram: Redes neurais artificiais; Árvore de decisão REPTree; Árvore de decisão (M5P); Floresta aleatória (RF); Máquina de vetor suporte; e zero R. Os parâmetros de acurácia utilizados foram: correlação de Pearson (r), erro médio absoluto e raiz quadrada do erro médio entre os valores observados e preditos. É possível prever fotossíntese líquida com r acima de 0,75, e as demais variáveis fisiológicas permeando resultados médios. As melhores técnicas de ML para essa abordagem são: RF e M5P.

PALAVRAS-CHAVE: Fisiologia vegetal; Machine learning; Sensoriamento remoto.

ABSTRACT

Soybeans are one of the most popular agricultural crops across different cultures and times of humanity. The concern in studying its physiology refers to technical and scientific development with the aim of overcoming the problem of hunger beyond market factors. Clearly understanding the biological mechanisms of plants is crucial to developing appropriate management that achieves greater productivity, even more so in this scenario of climate change and frequent natural phenomena. To this end, the objective here is to predict physiological variables through hyperspectral variables and using a machine learning (ML) algorithm, as well as identifying the best algorithms. A randomized block design with four replications was used. The plots consisted of five rows of five meters with a spacing of 0.45 m between rows. The assessments were: net photosynthesis; stomatal conductance; internal CO₂ concentration; perspiration; instantaneous water use efficiency; instantaneous carboxylation efficiency; and plant spectral analysis. The ML models used were: Artificial neural networks; REPTree decision tree; Decision tree (M5P); Random Forest (RF); Support vector machine; and zero R. The accuracy parameters used were: Pearson correlation (r), mean absolute error and square root of the mean error between the observed and predicted values. It is possible to predict net photosynthesis with r above 0.75, and the other physiological variables permeating average results. The best ML techniques for this approach are: RF and M5P.

KEYWORDS: Plant physiology; Machine learning; Remote sensing.

INTRODUÇÃO

O surgimento da agricultura auxiliou, de certo modo, no controle de um dos grandes problemas que assolam a humanidade, a fome. No entanto, esta agrura perdura até os dias atuais, e tem sido objeto de grande preocupação quando somada à estimativa do crescimento populacional para o século XXI (ONU, 2019). Dentre os 17 Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Organização das Nações Unidas (ONU) está a fome zero e agricultura sustentável. Neste viés, faz-se necessário o estudo de todos os fatores associados à produtividade a fim de se resolver ou amenizar significadamente o déficit nutricional.

No contexto da solução para a fome, e da consequente melhoria da produtividade das culturas, está a soja [*Glycine max* (L.) Merr.], considerada uma das principais culturas agrícolas do mundo, e a oleaginosa que obtém maior importância socioeconômica. O Brasil tem sido um dos maiores produtores do grão em todos os últimos anos e, na última safra, 2022/23, o país alcançou recorde histórico de área plantada, produtividade e produção. Contribuindo para isto, o estado de Mato Grosso do Sul produziu neste período a maior quantidade de soja em todo o seu histórico, sendo de 14 milhões de toneladas (CONAB, 2023).

A produtividade da cultura vem crescendo consideravelmente com os avanços tecnológicos e científicos aplicados nas lavouras. Todeschini et al. (2019) mostraram que nos últimos 50 anos, o Brasil aumentou a sua produtividade de 1.748 para 3.359 kg ha⁻¹. Ainda assim, faz-se necessário a abordagem de novas técnicas e métodos para suprir a demanda populacional associada aos fatores climáticos.

Neste sentido, o melhoramento da soja tem visado a obtenção de genótipos que expressem maior produtividade simultaneamente à resistência às adversidades climáticas. Para isso, é preciso realizar avaliações de desempenho de cultivares por meio de variáveis fisiológicas, tais como *fotossíntese líquida*, *transpiração*, *condutância estomática*, *concentração interna de CO₂* e *eficiência no uso da água (EUA)*.

Para auxiliar na mão-de-obra e reduzir os custos destas avaliações, bem como proporcionar acurácia dos resultados, pode-se fazer o uso do sensoriamento remoto, com a coleta e análise de variáveis hiperespectrais (LIU et al., 2016; RUTKOSKI et al., 2016). Além disso, o processamento de dados torna-se viável por meio de análises multivariadas devido à característica e extensão das variáveis.

As análises multivariadas são úteis em problemas que contenham variáveis aleatórias e independentes (ARTES e BARROSO, 2023), como as variáveis abordadas nesta pesquisa. Dentre as técnicas multivariadas aplicadas à agricultura, destaca-se o Random Forest (RF) pelo seu desempenho em correlacionar informações complexas e independentes dentro do sensoriamento remoto (RAMOS et al., 2020). Os autores Da Silva et al. (2021), Osco et al.

(2020) e Teodoro et. al. (2021) encontraram bons resultados para a predição de variáveis agronômicas por meio de radiação eletromagnética por meio de RF. Além disso, árvore de decisão e máquina de vetor suporte também são comumente aceitas para tais fins de predição.

Dessarte, este capítulo aborda a predição de variáveis fisiológicas na cultura da soja por meio de variáveis hiperespectrais, o que permitirá a elaboração de estratégias para aumentar a produção de alimentos pelo bom entendimento fisiológico das cultivares avaliadas. As hipóteses para esta pesquisa são que seja possível prever variáveis fisiológicas por meio de variáveis hiperespectrais e que técnicas multivariadas sejam capazes de expressar acurácia e precisão nessas correlações. Aqui objetiva-se estimar variáveis fisiológicas em grupos de soja, por meio de variáveis hiperespectrais em modelos de aprendizagem de máquina, bem como identificar os melhores modelos para este emprego.

MATERIAL E MÉTODOS

Condução do experimento a campo

O experimento ocorreu na safra 2023/24 na Universidade Federal de Mato Grosso do Sul, campus Chapadão do Sul (18°41'33''S, 52°40'45''W, com 810 m de altitude). O clima da região é Tropical de Savana (Aw); o solo classificado como Latossolo Vermelho Distrófico Argiloso, com as características na camada de 0 – 0,20 m: pH (H₂O) = 6,2; Al trocável (cmolc dm⁻³) = 0,0; Ca+Mg (cmolc dm⁻³) = 4,31; P (mg dm⁻³) = 41,3; K (cmolc dm⁻³) = 0,2; Matéria orgânica (g dm⁻³) = 19,74; V (%) = 45; m (%) = 0,0; Soma de bases (cmolc dm⁻³) = 2,3; CTC (cmolc dm⁻³) = 5,1.

Durante o período do ciclo da cultura, a pluviosidade concentrou-se na margem dos 10 mm/dia, alcançando esporadicamente picos de 30 a 50 mm/dia, enquanto a temperatura média oscilou entre os 20 e 30°C (Figura 1).

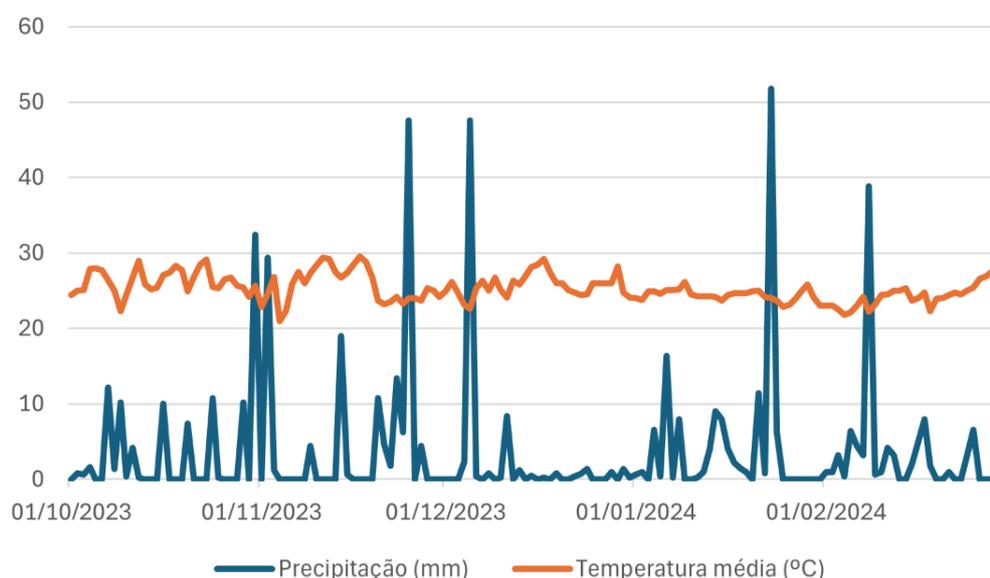


Figura 1. Dados diários de pluviosidade e temperatura média da safra 2023/24 na cidade de Chapadão do Sul/MS.

Os dados diários de pluviosidade e temperatura média durante os meses de outubro, novembro e dezembro de 2023, janeiro e fevereiro de 2024 foram coletados do banco de dados do Instituto Nacional de Meteorologia (INMET), disponibilizados na plataforma: <https://bdmep.inmet.gov.br/>.

O delineamento à campo ocorreu em blocos casualizados com parcelas contendo cinco linhas de um metro e meio de comprimento, espaçadas a 0,45 m e densidade de 15 plantas por metro. O experimento constou de quatro repetições e 25 populações de soja.

O solo foi preparado de modo convencional com aração e gradagem niveladora, seguido da semeadura, que ocorreu em outubro de 2023. Ocorreu adubação de base com 300kg/ha de

NPK (00-25-15), e adubação de KCl no quarto estágio vegetativo (v4) com a dosagem de 100 kg/ha. As sementes foram tratadas com fungicida (Piraclostrobina + Metil Tiofanato) e inseticida (Fipronil), na dose de 200 mL do produto a cada 100 Kg de sementes. Além disso, realizou-se a inoculação das sementes com bactérias do gênero *Bradyrhizobium*, na dose de 200 mL de inoculante líquido concentrado a cada 100 kg de sementes para a fixação biológica de nitrogênio. Os tratamentos culturais posteriores ocorreram de acordo com as necessidades da cultura.

Variáveis fisiológicas

As variáveis fisiológicas foram mensuradas aos 60 dias após a emergência (DAE) com o uso de um analisador de fotossíntese portátil (Infrared Gas Analyzer—IRGA) modelo Li-6400XT (LiCor Inc., Lincoln, Nebraska, USA). Também foi utilizado o fluxo de fótons fotossinteticamente ativos de $1000 \mu\text{mol m}^{-2} \text{s}^{-1}$ e concentrações de CO_2 ambiente ($372 \pm 10 \text{ mol m}^{-2} \text{s}^{-1}$).

As variáveis avaliadas foram: *fotossíntese líquida* (A, $\text{mmol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), *transpiração* (E, $\text{mmol H}_2\text{O m}^{-2} \text{ s}^{-1}$), *condutância estomática* (gs, $\text{mmol m}^{-2} \text{ s}^{-1}$), *concentração interna de CO_2* (C_i , $\text{mmol m}^{-2} \text{ s}^{-1}$), *eficiência instantânea de uso de água* (WUE) e *eficiência instantânea de carboxilação* (ICE). A WUE foi calculada pela razão entre A e E ($\mu\text{mol CO}_2 / \text{mmol H}_2\text{O}$), e a ICE pela razão entre A e C_i ($\text{CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ por $\mu\text{mol CO}_2 / \text{mol}^{-1}$).

No viés de maior eficácia na coleta de dados em detrimento da atmosfera, as medições ocorreram em dias sem nuvens, com temperaturas entre 26,0 e 26,5°C e umidade relativa e torno de 50 à 80%, entre 8:00 e 11:00 horas do horário local. Foram amostradas três plantas aleatórias em cada unidade experimental, tendo como padrão a terceira folha completamente desenvolvida a partir do ápice da planta.

Variáveis hiperespectrais

As folhas que passaram pela avaliação fisiológica também foram submetidas à análise espectral. Tal análise foi realizada com espectrorradiômetro (FieldSpec 4 Jr, Analytical Spectral Devices, Boulder, EUA). O equipamento tem a capacidade de medir o espectro na faixa de 350 a 2500 nm. Além disso, foi utilizado o equipamento ASD Plant Probe, que é projetado com a finalidade da medição espectral por meio do contato em materiais sólidos. A calibração do ASD Plant Probe se deu por meio de uma placa branca de sulfato de bário que reflete 100% da luz. O equipamento foi calibrado sempre após o término da leitura de um bloco, totalizando 4 calibrações.

O sensor foi conectado a um computador para os registros de cada leitura por meio de um software RS³, que é próprio do sensor e registra as leituras feitas pelo equipamento. Na sequência, os arquivos foram importados a outro software, ViewSpectroPro, para extração dos dados em extensão .txt. Uma das principais vantagens desse método é que ele não é destrutivo e não sofre interferência da luminosidade do ambiente, o que reduz erros ocasionado pela luz difusa.

Aprendizagem de máquina

O tratamento estatístico aplicado aos dados ocorreu por meio de análises de aprendizagem de máquina (ML). Os modelos de ML utilizados na predição foram: Redes neurais artificiais (MP, EGMONT-PETERSEN et al., 2002); Árvore de decisão REPTree (DT, SNOUSY et al., 2011); Árvore de decisão M5P (BLAIFI et al., 2018); Floresta aleatória (RF, BELGIU & DRĂGUȚ, 2016); Máquina de vetor suporte (SVM, NALEPA & KAWULOK, 2019) e Zero R, utilizado como modelo controle.

A variável predita (output) foi o fluxo de dióxido de carbono (FCO₂), enquanto os dados hiperspectrais foram utilizados como variáveis de entrada (input) nos modelos. A predição ocorreu de acordo com a validação cruzada estratificada com k-fold = 10 e dez repetições (100 execuções para cada modelo). Todos os parâmetros dos modelos foram estabelecidos pela configuração default do software Weka 3.8.5., com exceção para Redes Neurais, que obteve quatro neurônios por camada e duas camadas.

Os parâmetros de desempenho foram: métricas de coeficiente de correlação (r), erro absoluto médio (MAE) e raiz quadrada do erro médio (RMSE). Para verificar a significância dos algoritmos na predição foi realizada uma análise de variância. Nos casos de significância, foram gerados boxplots com as médias de r, MAE e RMSE, agrupados pelo teste de Scott-Knott (Scott & Knott, 1974) à de 5% de probabilidade. O agrupamento das médias e os boxplots foram gerados por meio dos pacotes ggplot2 e ExpDes.pt no software R.

RESULTADOS E DISCUSSÃO

Para a predição de fotossíntese líquida por meio das variáveis espectrais (Figura 2) todas as técnicas de aprendizagem de máquinas (ML) foram estatisticamente similares ao parâmetro correlação de Pearson (r), e expressaram resultados interessantes, com valores acima de 75%. Entretanto, analisando-se a variância de dados pelo espaçamento dos boxplots, pode-se dizer que Floresta aleatória (RF) destacou-se devido à menor dispersão dos valores.

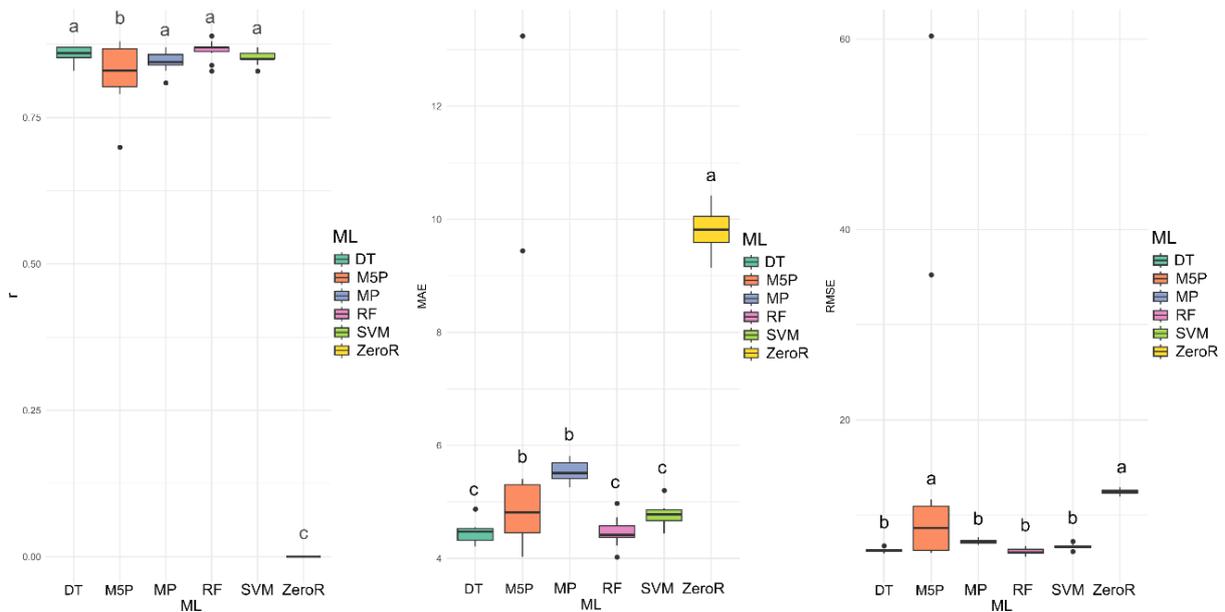


Figura 2. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável A (fotossíntese líquida).

Ao tratar-se dos erros MAE e RMSE, as melhores técnicas de ML com os menores valores, assemelhando-se entre si e diferindo-se das demais, foram Árvore de decisão REPTree (DT), RF e Máquina de Vetor Suporte (SVM). Embora Redes neurais artificiais (MP) tenha se assemelhado às técnicas citadas para raiz quadrada do erro médio (RMSE), divergiu-se em erro médio absoluto (MAE). A técnica menos favorável foi Zero R.

A *fotossíntese líquida* (A) remete-se à diferença entre a fixação total de CO_2 durante a fotossíntese e a respiração das plantas (POULTER et al., 2014). Portanto, a A se relaciona diretamente ao sequestro do carbono na atmosfera, e quantificá-la via sensoriamento remoto permite identificar cultivares com aspecto de sumidouro do gás. Para além disso, a predição desta variável por meio de informações hiperespectrais apresenta alta correlação devido à associação da fotossíntese com pigmentos presentes nas plantas. Os autores Silva et al., 2024

encontraram correlação entre fotossíntese e dados hiperespectrais por meio de espectrorradiometria e sensoriamento remoto via satélite.

Pigmentos fotossintetizantes interferem diretamente nos comprimentos de onda que são refletidos pelos tecidos vegetais (DOS SANTOS et al., 2019). As clorofilas *a* e *b* são os pigmentos predominantes do mesófilo na espécie da soja. Essas moléculas especializadas na captação da luz absorvem principalmente a energia nos comprimentos de onda do azul (cerca de 430 nm) e do vermelho (cerca de 660 nm). Deste modo, os picos de reflectância das clorofilas ocorrem na faixa do verde (cerca de 550 nm) para o espectro visível, e na faixa do infravermelho (TAIZ et al., 2017). Isso revela a possibilidade de se relacionar a fotossíntese com o espectro eletromagnético de reflectância dos tecidos foliares das plantas.

Neste sentido, vale ressaltar a importância da nutrição de plantas nesse processo, uma vez que as moléculas fotossintetizantes requerem os elementos disponíveis na solução do solo. O magnésio, por exemplo, é o átomo central do anel porfírico da clorofila, e segue-se rodeado de nitrogênio (BROWN et al., 2022). Os autores ressaltam também que o potássio e o cálcio são exemplos de ativadores enzimáticos que permitem os processos metabólicos de trocas gasosas. Portanto, aqui nota-se a relação dos macronutrientes com as variáveis fisiológicas e hiperespectrais das plantas.

Além disso, as avaliações espectrais ocorreram aos 60 dias após à emergência da soja, o que remete ao início do período reprodutivo (R1) (WALTER FEHR & CAVINESS, 1977). Neste sentido, torna-se necessário relacionar as características bioquímicas das folhas nesse estágio para melhor entendimento dos dados, visto que a reflectância espectral se expressa em detrimento das moléculas que são expostas ao eletromagnetismo.

O período entre o florescimento e o enchimento de grãos está entre àqueles de maior exigência hídrica pela soja (FARIAS et al., 2007), época que abrange R1. Isto justifica a presença considerável de água nos tecidos foliares submetidos à análise espectral. A água relaciona-se com todas as variáveis espectrais avaliadas, absorvendo, espalhando ou refletindo fótons e ondas eletromagnéticas (BARBOSA et al., 2019).

As moléculas de água inseridas nos tecidos foliares absorvem grande parte dessa energia, e influenciam inversamente em sua refletância espectral. Quando há diminuição no teor de água na planta, verifica-se um aumento da refletância em toda a faixa de comprimento de onda de 400 nm a 2.500 nm, principalmente na região do infravermelho médio (STRABELI et al., 2020). Por outro lado, as moléculas obtêm maiores reflectâncias no espectro visível, mais especificamente nos comprimentos de onda do azul e do verde (FERREIRA, A. B.; PEREIRA FILHO, W., 2009)

No processo de fixação de energia na fotossíntese a água apresenta-se ligada a um canal de prótons do citocromo f (NELSON; COX, 2014), além de atuar como fonte de hidrogênio para a formação de NADPH e ATP (SILVEIRA et al., 2015), e doar elétrons em seu processo de oxidação. Tudo isso exemplifica a relação da água com o sistema fotossintético e, consequentemente, a correlação da fotossíntese com a reflexão espectral.

Outrossim, a época da coleta de dados, que foi no verão de janeiro, também tem relevância na análise de dados, visto que o fotoperíodo associado à carga genética e o clima influenciam na resposta fisiológica dos vegetais (OLIVEIRA E GUALTIERI, 2017). Os autores afirmam que a intensidade luminosa proveniente do sol interfere na fotossíntese e, consequentemente, na fixação de carbono e no processo de carboxilação.

Com relação à predição de condutância estomática pelo parâmetro r (Figura 3), as técnicas de ML com destaque foram Árvore de decisão (M5P) e SVM. Todavia, o resultado de todas as técnicas aproximou-se entre si, com exceção de Zero R, oscilando entre 40 e 50%, aproximadamente. Ambos os erros MAE e RMSE sustentaram os mesmos resultados estatísticos, em que MP e Zero R expressam os maiores valores, diferenciando-se dos demais, que são similares entre si. Isso evidencia que MP e Zero R não são recomendados para essa predição.

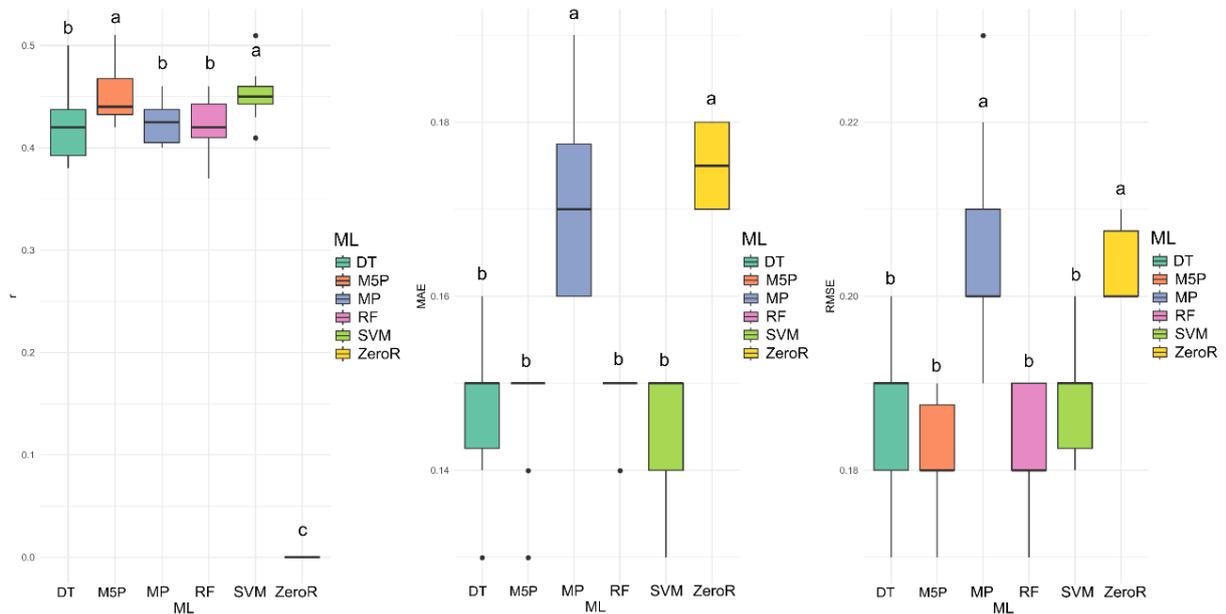


Figura 3. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável g_s (condutância estomática).

Na condutância estomática a água ocupa função de destaque pela concentração interna de vapor na folha (SILVA et al., 2024). Daí, nota-se mais uma vez a implicação da refletância

da água na relação espectral com a variável fisiológica. Além disso, sabe-se que o influxo de CO₂ ocorre simultaneamente ao efluxo de vapor d'água na transpiração, por meio da abertura estomática (SILVA et al., 2015). Os autores demonstram que a concentração interna de carbono em tecidos vegetais aumenta conforme a condutância estomática se eleva, tal como a transpiração em proporções similares.

Isso revela que as variáveis fisiológicas avaliadas são dependentes entre si e que, assim como as biomoléculas de pigmentos interferem em tais rotas metabólicas, a água também contribui para tal. Neste sentido, quando se avalia a reflexão espectral do dossel, analisa-se o comportamento espectral já conhecido dessas biomoléculas.

Em se tratando da predição da concentração interna de CO₂ (Figura 4), o parâmetro *r* aponta todas as técnicas como similares e eficazes, com exceção de Zero R, que obteve 0%. DT, M5P, MP, RF e SVM aproximaram-se dos 80% em eficácia na predição, que é ótimo para modelos algoritmos. Contudo, MP expressa maior dispersão de seus dados, que é refletido em seu afastamento às demais nos parâmetros de erros.

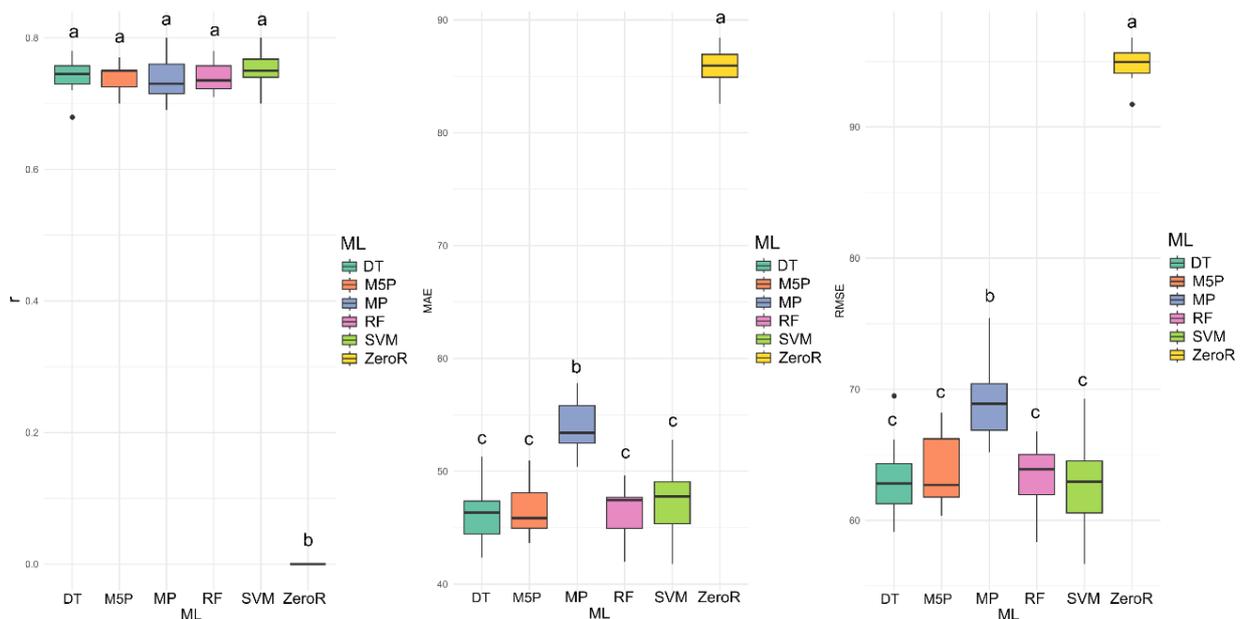


Figura 4. Boxplots para coeficiente de correlação de Pearson (*r*), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável *C_i* (concentração interna de CO₂).

MAE e RMSE obtiveram os mesmos resultados, em que Zero R se diferiu das demais com valores elevados, sendo uma técnica não recomendada. MP expressou valores maiores que as demais, que o fez se diferenciar como uma técnica intermediária. DT, M5P, RF e SVM destacaram-se com os menores erros.

Para a predição de transpiração pelo parâmetro r (Figura 5), a melhor técnica foi M5P, diferenciando-se das demais. DT, RF e SVM ilustraram valores intermediários entre 30 e 40% de eficácia. MP decaiu na porcentagem, diferenciando-se, e seguida pela Zero R, que foi a pior técnica utilizada. Os maiores erros (MAE e RMSE) foram apontados por MP, seguidos pelos erros de Zero R. A técnica de destaque, que abordou os menores erros, foi M5P, coincidindo com o parâmetro r para este output. As demais apresentaram erros intermediários.

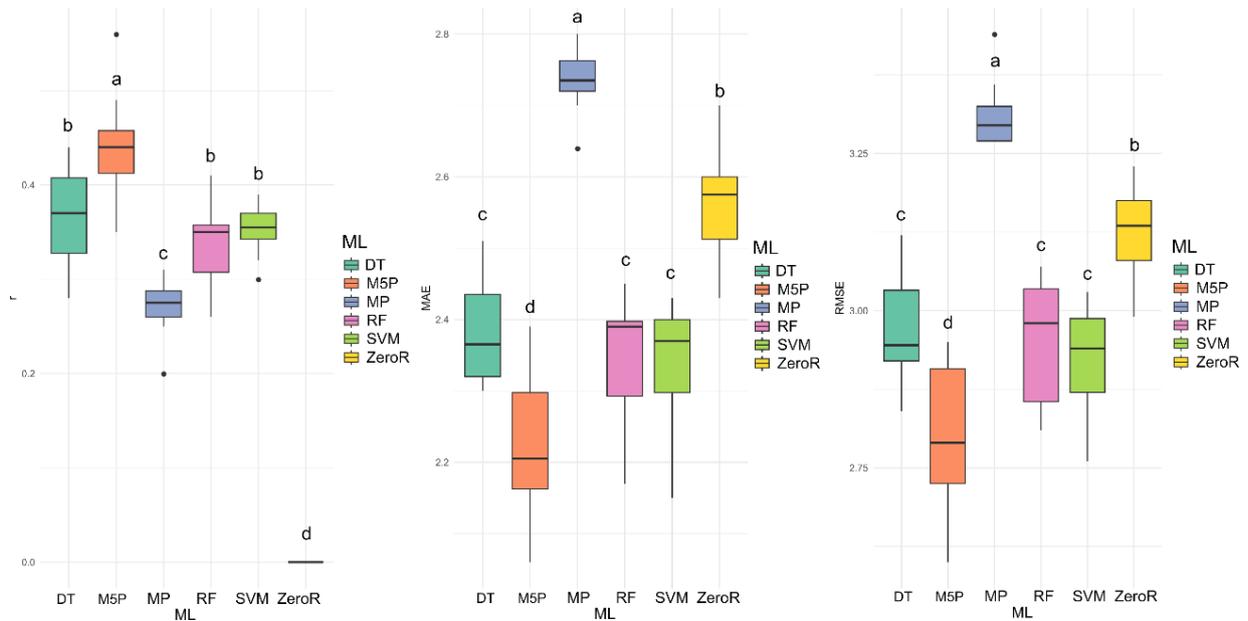


Figura 5. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável E (transpiração).

As melhores técnicas para prever a eficiência instantânea de uso de água por meio de variância espectral sob o parâmetro r (Figura 6) foram M5P e RF, com aproximação aos 60% de acurácia. Porém, as técnicas DT, MP e SVM, que foram similares entre si, oscilaram entre os valores acima de 50%, não apresentando grande diferença entre as melhores. Já para Zero R não foi possível obter dados satisfatórios. Em MAE, M5P assimilou-se com Zero R para os maiores erros, diferenciando-se das demais técnicas. Em RMSE, M5P expressou os maiores erros, enquanto as técnicas restantes não diferiram entre si.

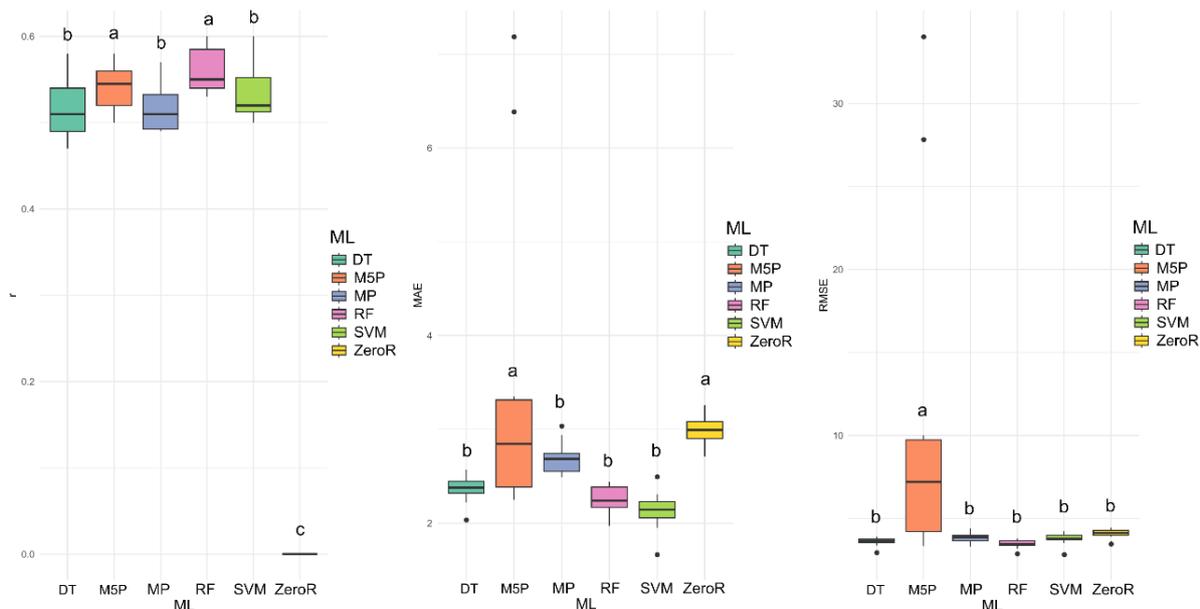


Figura 6. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável WUE (eficiência instantânea de uso de água).

As técnicas M5P, MP e SVM apresentaram os melhores resultados na predição de eficiência instantânea de carboxilação pelo parâmetro r (Figura 7), diferindo-se das demais. No entanto, essa predição não obteve bons resultados em nenhuma instancia, haja vista que os valores oscilaram entre 20 e 35% de acurácia, aproximadamente. Isto refletiu em erros elevados onde, para RMSE, nenhuma técnica diferiu entre si, e para MAE, Zero R expressou os maiores erros, e SVM, os menores.

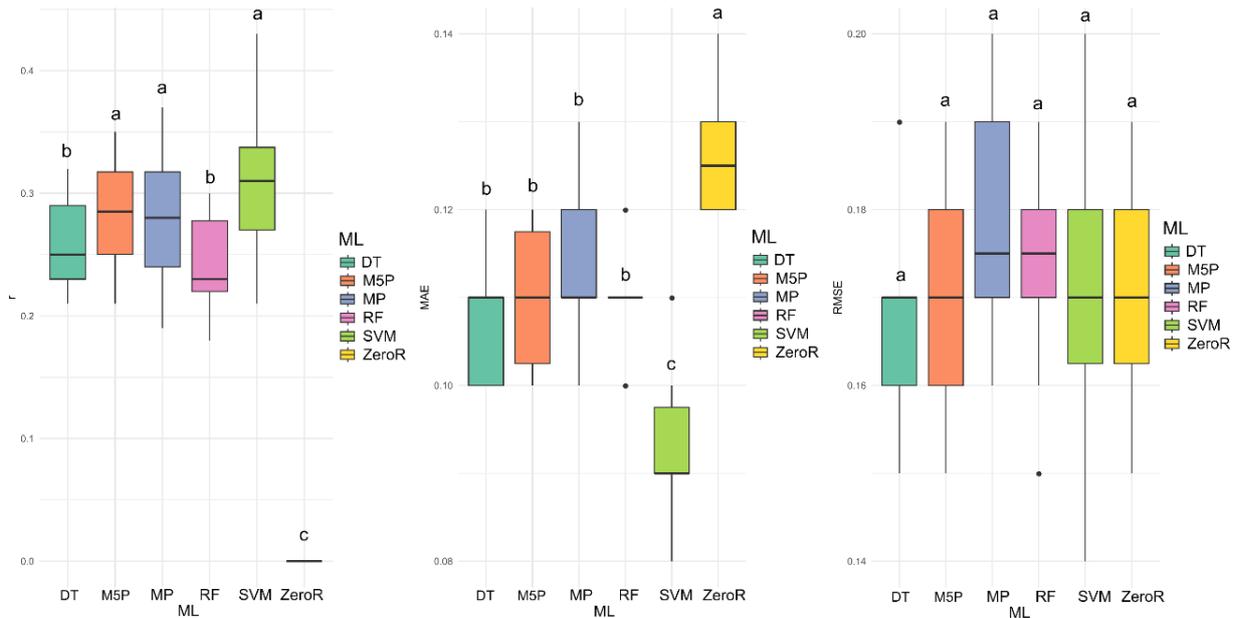


Figura 7. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável ICE (eficiência instantânea de carboxilação).

A variável eficiência instantânea de carboxilação, que expressou baixa porcentagem de correlação com os comprimentos de onda, pode ser analisada considerando-se que ela está inserida em um processo metabólico altamente correlacionado a outros fatores. O dióxido de carbono se reduz à açúcares, e o processo de carboxilação nada mais é do que a fixação de CO_2 no ciclo de redução fotossintética do carbono (NELSON; COX, 2014). Dessarte, afirma-se que essas variáveis fisiológicas são correlacionadas à fotossíntese e, por sua vez, ao espectro eletromagnético de refletância dos pigmentos fotossintetizantes.

O surgimento da computação associada ao sensoriamento remoto permite com maior amplitude a abordagem da sustentabilidade aplicada às práticas agrícolas (OGUNGBUYI et al., 2023). Técnicas de aprendizado de máquina na agricultura de precisão possibilitam estimar o crescimento e a produção de árvores em plantações florestais (DA SILVA et al., 2021), por exemplo, dentre outras aplicações agronômicas.

Entretanto, o sensoriamento hiperespectral não é possível ser trabalhado com processamentos estatísticos de modelos convencionais em detrimento da extensão da base de dados. Neste trabalho, por exemplo, as variáveis de entrada foram 2.150 comprimentos de onda. Isto se reflete no baixo aproveitamento do modelo Zero R, visto que é uma regressão linear. Desta maneira, faz-se necessária a abordagem de técnicas de aprendizagem de máquina (ML) com regressões multivariadas, como Redes neurais artificiais (MP), Árvore de decisão REPTree (DT), Árvore de decisão (M5P), Floresta aleatória (RF) e Máquina de vetor suporte (SVM).

Ao comparar-se as técnicas de aprendizagem de máquina entre si em relação aos parâmetros: coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE); as ML que expressaram melhores resultados, em uma abordagem geral das variáveis fisiológicas foram M5P e RF.

Árvore de Decisão M5P é um algoritmo que relaciona o input com output por meio de gráficos de ramificações que, sob condições e probabilidades, apresentam nós de convergência ou divergência entre os dados, tanto para regressão quanto para classificação (MONARD E BARANAUSKAS, 2003). Já a Floresta Aleatória (RF), que também é utilizada para a previsão de dados (SCHONLAU & ZOU, 2020), é o conjunto de diversas Árvores de Decisão que dispõe de menor aparição de overfitting e maior precisão e acurácia devido à repetição das árvores.

A técnica de RF tem ganhado espaço no âmbito das ciências agrárias, sobretudo em agricultura de precisão devido ao seu bom desempenho como algoritmo preciso (RAMOS et al., 2020). Por meio da Floresta Aleatória associada ao sensoriamento remoto pode-se prever a quantidade de ganho de biomassa em pastagem, como afirmam os autores (DE ROSA et al., 2021). Outrossim, o algoritmo foi estabelecido como a melhor técnica de Machine Learning para prever o crescimento e altura de espécies vegetais por meio de dados espectrais pelos autores (DA SILVA et al., 2021; OSCO et al., 2020; TEODORO et al., 2021). Há também a possibilidade de estimar a produtividade de culturas agrícolas em detrimento de índices de vegetação quando os dados são submetidos à RF (RAMOS et al., 2020).

O algoritmo SVM também expressou bons resultados nesta pesquisa, principalmente na predição de condutância estomática. Santana et al. (2023b) afirmam que essa técnica obteve êxito na correlação entre variáveis espectrais e teores de macronutrientes primários em folhas de soja, o que ressalta seu bom desempenho para estimativas fisiológicas.

Embora Redes Neurais Artificiais (MP) tenha apresentado resultados inferiores às demais técnicas, com exceção de Zero R, Santana et al. (2023) apontam o algoritmo, juntamente com Árvore de decisão REPTree (DT) e SVM, como sendo apropriados para análises em fenotipagem de alta precisão.

Associar variáveis fisiológicas às variáveis espectrais é algo recente no âmbito agrônomico, sobretudo com a aplicação de aprendizagem de máquina. Essa abordagem torna-se interessante por diversos aspectos: economia no financiamento de pesquisas e na prática em geral, visto que a aparelhagem empregada nas avaliações fisiológicas é de alto custo; redução de mão-de-obra e tempo de serviço pela praticidade do sensoriamento remoto; e maior acurácia nos resultados em decorrência do uso de sensores hiperespectrais e estatística avançada.

Portanto, com modelos de predição de variáveis fisiológicas por meio do espectro de reflexão das folhas, será possível renunciar a avaliações à campo e estimar tais variáveis apenas com sensoriamento remoto, sob o uso de sensores acoplados às aeronaves remotamente pilotadas, ou por meio de informações de satélites.

CONCLUSÕES

É possível estimar a fotossíntese líquida e a concentração interna de CO_2 por meio de variáveis espectrais com alta acurácia, em que o coeficiente de correlação de Pearson encontra-se em torno de 70%. No entanto, as demais variáveis fisiológicas que, embora tenham apresentado valores inferiores e requerem maior demanda de pesquisa, também permitem tal mensuração, sobretudo a eficiência instantânea de uso de água.

Os melhores modelos de aprendizagem de máquina neste emprego são Floresta aleatória e Árvore de decisão M5P.

REFERÊNCIAS

ARTES, R.; BARROSO, L. P. Métodos multivariados de análise estatística. 1ª ed. São Paulo: **Blucher**, 2023. 534 p.

BARBOSA, C. C. F. et al. **Introdução ao Sensoriamento Remoto de sistemas aquáticos: princípios e aplicações**. 1ª ed. São José dos Campos: INPE, 2019. 178 p.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BLAIFI, S. et al. MSP model tree based fast fuzzy maximum power point tracker. **Solar Energy**, v. 163, p. 405–424, 2018.

BROWN, P. H.; ZHAO, F. J.; DOBERMANN, A. What is a plant nutrient? Changing definitions to advance science and innovation in plant nutrition. **Plant and Soil**, v. 476, n. 1, p. 11-23, 2022.

CONAB. Acompanhamento da safra brasileira: grãos - safra 2022/23 – oitavo levantamento. Brasília: **Companhia Nacional de Abastecimento**, v. 10, n. 8, 2023.

DA SILVA, A. K. V. et al. Predicting eucalyptus diameter at breast height and total height with uav-based spectral indices and machine learning. **Forests**, v. 12, n. 5, 2021.

DE ROSA, D. et al. Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. **Computers and Electronics in Agriculture**, v. 180, 2021.

DOS SANTOS, F. C. et al. Relação entre transparência da água e a reflectância da imagem Landsat 5 tm do reservatório Passo Real. **Geoambiente on-line**, Jataí-GO, n. 33, 2019. Disponível em: <www.revistas.ufg.br/geoambiente>.

EGMONT-PETERSEN, M. et al. Image processing with neural networks - a review. **Pattern Recognition**, v. 35, n. 10, p. 2279–2301, 2002.

FARIAS, J. R. B. et al. Ecofisiologia da soja. **Circular Técnica**, Londrina: Embrapa, 2007.

FERREIRA, A. B.; PEREIRA FILHO, W. Avaliação da reflectância espectral de corpos d'água em Santa Maria-RS por meio de espectrorradiometria de campo. **Geoambiente On-line**, n. 13, p. 01-14, 2009.

LIU, C.; SUN, P. SEN; LIU, S. R. A review of plant spectral reflectance response to water physiological changes. **Chinese Journal of Plant Ecology**. Editorial Office of Chinese. 2016.

MARQUES RAMOS, A. P. 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, 2020.

MONARD M. C.; BARANAUSKAS J. A. Indução de Regras e Árvores de Decisão. In: MONARD M. C.; BARANAUSKAS J. A. **Sistemas inteligentes-Fundamentos e aplicações**: Conceitos sobre aprendizado de máquina. v. 1, n. 1, p. 57-74, 2003.

NALEPA, J.; KAWULOK, M. Selecting training sets for support vector machines: a review. **Artificial Intelligence Review**, v. 52, n. 2, p. 857–900, 2019.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. 6ª ed. Porto Alegre: Artmed, 2014. 1250 p.

OGUNGBUYI, M. G. et al. Enabling Regenerative Agriculture Using Remote Sensing and Machine Learning. **Land**, v. 12, n. 6, 2023.

OLIVEIRA, A. K. M.; GUALTIERI, S. C. J. Trocas gasosas e grau de tolerância ao estresse hídrico induzido em plantas jovens de *Tabebuia aurea* (paratudo) submetidas a alagamento. **Ciência Florestal**, Santa Maria, v. 27, n. 1, p. 181-191, 2017.

ONU - Organização das Nações Unidas. **World Population Prospects 2019: Highlights**. Department of Economic and Social Affairs, Population Division. ST/ESA/SER.A/423, p. 46, 2019.

OSCO, L. P. et al. Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. **Remote Sensing**, v. 12, n. 19, p. 1–17, 2020.

PARKINSON, K. J. An improved method for measuring soil respiration in the field. **Journal of applied ecology**, p. 221–228, 1981.

POULTER B. et al. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. **Nature** 509, 600–603, 2014. <https://doi.org/10.1038/nature13376>

RAMOS, A. P. M. 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.

ROSSI, F. S. et al. 19-year remotely sensed data in the forecast of spectral models of the environment. **International Journal of Digital Earth**, v. 14, n. 8, p. 1040–1066, 2021.

RUTKOSKI, J. et al. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. **G3: Genes, Genomes, Genetics**, v. 6, n. 9, p. 2799–2808, 2016.

SANTANA, D. C. et al. Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. **Remote Sensing Applications: Society and Environment**, v. 29, 2023a.

SANTANA, D. C. et al. Machine Learning in the Classification of Soybean Genotypes for Primary Macronutrients' Content Using UAV–Multispectral Sensor. **Remote Sensing**, v. 15, n. 5, 2023b.

SCHONLAU, M.; ZOU, R. Y. The random forest algorithm for statistical learning. **Stata Journal**, v. 20, n. 1, 2020.

SCOTT, A. A. J.; KNOTT, M. A Cluster Analysis Method for Grouping Means in the Analysis of Variance Published by: International Biometric Society Stable. **Biometrics**, v. 30, n. 3, p. 507–512, 1974.

SILVA, F. G. et al. Gas exchange and chlorophyll fluorescence of eggplant grown under different irrigation depths. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 10, p. 946–952, 1 set. 2015.

SILVA, J. N. B. et al. Análise da confiabilidade da produção primária bruta e da fotossíntese líquida determinados com espectrorradiometria em campo e dados de satélite. **Geoambiente On-line**, n. 50, 2024.

SILVEIRA, M. T. et al. **Fisiologia Vegetal**. Florianópolis: Universidade Federal de Santa Catarina, 2015. 182 p.

SNOUSY, M. B. AL et al. Suite of decision tree-based classification algorithms on cancer gene expression data. **Egyptian Informatics Journal**, v. 12, n. 2, p. 73–82, 2011.

STRABELI, T. F. et al. Influence of relative water content variation on the spectral behavior of Eucalyptus spp. **Scientia Forestalis/Forest Sciences**, v. 48, n. 128, 2020.

TAIZ, L. et al. **Fisiologia e Desenvolvimento Vegetal**. 6ª ed. Porto Alegre: Artmed, 2017. 888 p.

TEODORO, P. E. et al. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. **Remote Sensing**, v. 13, n. 22, 2021.

TODESCHINI, M. H. et al. Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. **Euphytica**, v. 215, n. 7, 2019.

WALTER FEHR, BY R.; CAVINESS, C. E. Stages of Soybean Development. **Special Report** 80. v. 80, p. 1–12, 1977.

CAPÍTULO 2:

PREDIÇÃO DA FIXAÇÃO E EMISSÃO DE CARBONO EM SOJA

RESUMO

A soja é uma das culturas agrícolas mais populares por diversas culturas e tempos da humanidade, e neste cenário de mudanças climáticas e fenômenos naturais frequentes é fulcral associar as necessidades de mercado com as exigências ambientais. Deste modo, as culturas agrícolas podem se comportar como sumidouro de dióxido de carbono (CO_2), um dos mais agravantes gases do efeito estufa. Para isso, objetiva-se aqui prever a emissão in situ de dióxido de carbono do solo (FCO_2) e a fixação de carbono no tecido foliar da soja por meio de variáveis hiperespectrais e utilizando algoritmo de aprendizagem de máquina (ML), tal como identificar os melhores algoritmos. Foi empregado delineamento de blocos casualizados com quatro repetições. As parcelas consistiram em cinco linhas de cinco metros com espaçamento de 0,45 m entre fileiras. As avaliações foram de: fluxo de dióxido de carbono, temperatura e umidade do solo in loco; fixação de carbono no tecido foliar; e análise espectral da planta. Os modelos de ML utilizados foram: Redes neurais artificiais; Árvore de decisão REPTree; Árvore de decisão (M5P); Floresta aleatória (RF); Máquina de vetor suporte; e zero R. Os parâmetros de acurácia utilizados foram: correlação de Pearson (r), erro médio absoluto e raiz quadrada do erro médio entre os valores observados e preditos. As previsões de FCO_2 e fixação de carbono por meio de variáveis hiperespectrais expressaram resultados insatisfatórios, o que requer mais estudos nesse ramo. As melhores técnicas de ML para essa abordagem são: RF e M5P.

PALAVRAS-CHAVE: Fluxo de carbono; Machine learning; Sensoriamento remoto.

ABSTRACT

Soy is one of the most popular agricultural crops across different cultures and times of humanity, and in this scenario of climate change and frequent natural phenomena, it is crucial to associate market needs with environmental requirements. In this way, agricultural crops can act as a sink for carbon dioxide (CO₂), one of the most aggravating greenhouse gases. To this end, the objective here is to predict the in situ emission of soil carbon dioxide (FCO₂) and carbon fixation in soybean leaf tissue through hyperspectral variables and using a machine learning (ML) algorithm, such as identifying the best algorithms. A randomized block design with four replications was used. The plots consisted of five rows of five meters with a spacing of 0.45 m between rows. The assessments were: carbon dioxide flow, temperature and soil humidity in situ; carbon fixation in leaf tissue; and plant spectral analysis. The ML models used were: Artificial neural networks; REPTree decision tree; Decision tree (M5P); Random Forest (RF); Support vector machine; and zero R. The accuracy parameters used were: Pearson correlation (r), mean absolute error and square root of the mean error between the observed and predicted values. Predictions of FCO₂ and carbon fixation using hyperspectral variables expressed unsatisfactory results, which requires further studies in this field. The best ML techniques for this approach are: RF and M5P.

KEYWORDS: Carbon flow; Machine learning; Remote sensing.

INTRODUÇÃO

Para que a fome seja solucionada, ou ao menos significativamente minimizada, é preciso que se faça destacar o termo “sustentável” do objetivo: fome zero e agricultura sustentável, presente nos 17 Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pela Organização das Nações Unidas (ONU). Não há sentido em solucionar um problema às custas de outro, além de que o bom desempenho das lavouras está intimamente relacionado aos fatores climáticos. Ou seja, se a humanidade não se atentar às pautas ambientais, em pouco tempo será impossível cultivar, sobretudo em escala mundial (ONU, 2024).

Um dos fatores de extrema importância para o meio ambiente é a emissão de dióxido de carbono (CO₂) na atmosfera. O aumento da concentração desse gás tem provocado mudanças climáticas que, além de serem maléficas ao equilíbrio ambiental, prejudicam diretamente o rendimento das lavouras.

Outra associação interessante entre a agricultura e a emissão de CO₂ é que as lavouras podem atuar tanto como emissoras, quanto como sumidouro do gás. O revolvimento do solo em práticas agrícolas convencionais permite que o carbono agregado em suas partículas seja disperso pela atmosfera (BARANDU, 2021). Por outro lado, a incorporação do CO₂ pelas plantas na fotossíntese é um excelente meio para o sequestro do carbono atmosférico, o que permite que lavouras com manejo adequado se comportem como sumidouro (ROSSI et al., 2021).

A soja [*Glycine max* (L.) Merr.], considerada uma das principais culturas agrícolas do mundo, e a oleaginosa que obtém maior importância socioeconômica, pode auxiliar na temática da alimentação sustentável. O Brasil tem sido um dos maiores produtores do grão em todos os últimos anos e, na última safra, 2022/23, o país alcançou recorde histórico de área plantada, produtividade e produção. Contribuindo para isto, o estado de Mato Grosso do Sul produziu neste período a maior quantidade de soja em todo o seu histórico, sendo de 14 milhões de toneladas (CONAB, 2023). Neste sentido, é de grande relevância a abordagem do estudo sobre fluxo e fixação de carbono nos genótipos de soja, que podem auxiliar, para além da problemática da fome, no controle do carbono no meio ambiente.

Para auxiliar na mão-de-obra e reduzir os custos destas avaliações, bem como proporcionar acurácia dos resultados, pode-se fazer o uso do sensoriamento remoto, com a coleta e análise de variáveis hiperespectrais (LIU et al., 2016; RUTKOSKI et al., 2016). Além disso, o processamento de dados torna-se viável por meio de análises multivariadas devido à característica e extensão das variáveis.

Portanto, neste capítulo aborda-se a predição de fluxo de CO₂ no solo e fixação de carbono no tecido foliar da cultura da soja por meio de variáveis hiperespectrais, o que

possibilitará a elaboração de estratégias para aumentar a produção de alimentos com o pressuposto de mitigar os efeitos do CO₂ na atmosfera. As hipóteses para esta pesquisa são que seja possível prever variáveis de carbono por meio de variáveis hiperespectrais e que técnicas multivariadas sejam capazes de expressar acurácia e precisão nessas correlações. O objetivo deste capítulo é estimar CO₂ no solo e fixação de carbono no tecido foliar em grupos de genótipos de soja por meio de variáveis hiperespectrais em parâmetros de aprendizagem de máquina, bem como identificar os melhores modelos para este emprego.

MATERIAL E MÉTODOS

Condução do experimento a campo

O experimento ocorreu na safra 2023/24 na Universidade Federal de Mato Grosso do Sul, campus Chapadão do Sul (18°41'33''S, 52°40'45''W, com 810 m de altitude). O clima da região é Tropical de Savana (Aw); o solo classificado como Latossolo Vermelho Distrófico Argiloso, com as características na camada de 0 – 0,20 m: pH (H₂O) = 6,2; Al trocável (cmolc dm⁻³) = 0,0; Ca+Mg (cmolc dm⁻³) = 4,31; P (mg dm⁻³) = 41,3; K (cmolc dm⁻³) = 0,2; Matéria orgânica (g dm⁻³) = 19,74; V (%) = 45; m (%) = 0,0; Soma de bases (cmolc dm⁻³) = 2,3; CTC (cmolc dm⁻³) = 5,1.

Durante o período do ciclo da cultura, a pluviosidade concentrou-se na margem dos 10 mm/dia, alcançando esporadicamente picos de 30 a 50 mm/dia, enquanto a temperatura média oscilou entre os 20 e 30°C (Figura 1).

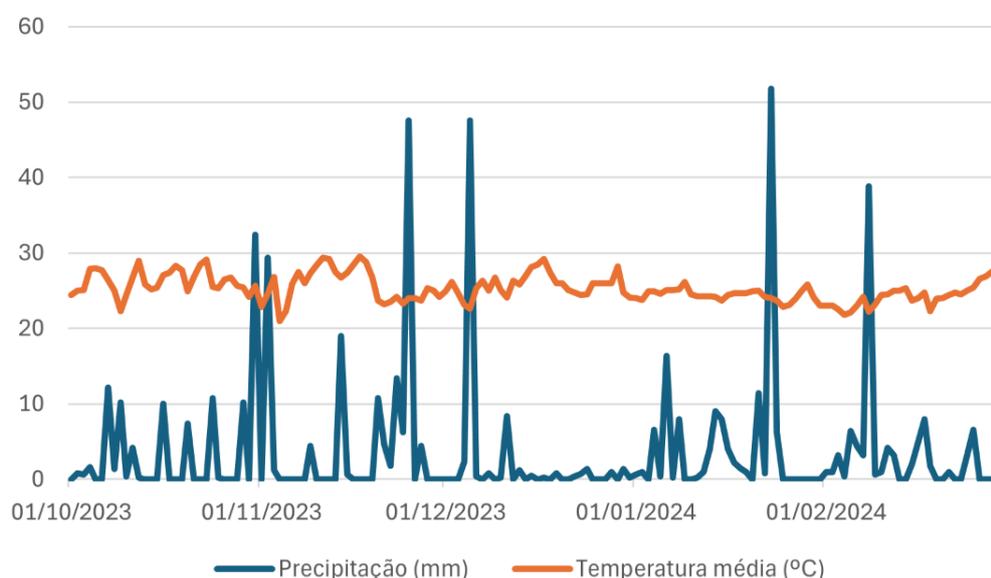


Figura 1. Dados diários de pluviosidade e temperatura média da safra 2023/24 na cidade de Chapadão do Sul/MS.

Os dados diários de pluviosidade e temperatura média durante os meses de outubro, novembro e dezembro de 2023, janeiro e fevereiro de 2024 foram coletados do banco de dados do Instituto Nacional de Meteorologia (INMET), disponibilizados na plataforma: <https://bdmep.inmet.gov.br/>.

O delineamento à campo ocorreu em blocos casualizados com parcelas contendo cinco linhas de um metro e meio de comprimento, espaçadas a 0,45 m e densidade de 15 plantas por metro. O experimento constou de quatro repetições e 25 populações de soja.

O solo foi preparado de modo convencional com aração e gradagem niveladora, seguido da semeadura, que ocorreu em outubro de 2023. Ocorreu adubação de base com 300kg/ha de NPK (00-25-15), e adubação de KCl no quarto estágio vegetativo (v4) com a dosagem de 100

kg/ha. As sementes foram tratadas com fungicida (Piraclostrobina + Metil Tiofanato) e inseticida (Fipronil), na dose de 200 mL do produto a cada 100 Kg de sementes. Além disso, realizou-se a inoculação das sementes com bactérias do gênero *Bradyrhizobium*, na dose de 200 mL de inoculante líquido concentrado a cada 100 kg de sementes para a fixação biológica de nitrogênio. Os tratos culturais posteriores ocorreram de acordo com as necessidades da cultura.

Fluxo de CO₂

As mensurações de fluxo de CO₂ foram realizadas em colares de PVC fixados no solo 24 horas antes da primeira leitura, que permaneceram na área até a última mensuração (trigésima leitura). Cada parcela obteve um ponto para a coleta de dados. O equipamento utilizado na leitura foi o sistema portátil EGM-5 (PP-Systems, Amesbury, EUA), modelo AGA560. Os procedimentos aplicados na mensuração da emissão in situ de CO₂ do solo (FCO₂) deram-se de acordo com Teodoro et al. (2024).

O sistema portátil EGM-5 monitora as variações da concentração de CO₂ no interior da câmara de PVC do solo por meio de espectroscopia de absorção óptica na região espectral do infravermelho (ACS041). Essa câmara é um sistema fechado com volume interno de 854,2 cm³ e área de contato circular de 83.7 cm².

A mensuração de FCO₂ foi submetida a um ajuste da concentração de CO₂ do ar dentro da câmara em função de uma regressão quadrática no tempo após o seu fechamento e a compensação de água. A aquisição dos dados da emissão de CO₂ do solo contou com o registo de 30 segundo para cada ponto amostral, enquanto a pressão barométrica dentro da câmara foi determinada a cada 10 segundos. O FCO₂ (μmol m⁻² s⁻¹) foi calculado conforme Equação 1, proposta por (Parkinson, 1981):

$$FCO_2 = \frac{dC}{dT} \times \frac{P}{1013} \times \frac{273}{273 + T} \times \frac{V}{A} \times 44.615 \quad ($$

Em que: dC é mudança na concentração de CO₂ em ppm; dT é o tempo de execução do processo (1 segundo); P é a pressão barométrica mensurada em mbar; T é a temperatura do ar em °C; V é o volume da câmara do EGM5 em m³; A é a área dos anéis de PVC instalados no solo em m².

Fixação de carbono

Foram dissolvidos 139 g de Solução de sulfato ferroso (FeSO₄.7H₂O 0,5 mol L⁻¹) em 15 mL de H₂SO₄ concentrado, e diluído em até o volume de 1 litro. Foi titulada a solução contra 10 mL de K₂Cr₂O₇ 1 mol L⁻¹. Foram dissolvidos 61,30 g de K₂Cr₂O₇ 0,25 mol L⁻¹ (secado

a 105°C por duas horas) em 1 litro de água destilada. Foram dissolvidos 1 g de difenilamina em 100 mL de H₂SO₄ (95%).

Em um frasco de erlenmeyer de 250 mL foi adicionado 0,05 g do material vegetal de cada parcela e 10 mL de K₂Cr₂O₇ 0,25 mol L⁻¹. Agitou-se levemente o frasco e adicionou-se 10 mL de H₂SO₄ concentrado. Depois de agitar novamente e aguardar 30 minutos, a solução foi diluída para 100 mL.

Em outro erlenmeyer adicionou-se 100 mL de H₂O destilada, 5 mL da solução (material vegetal, 10 mL de K₂Cr₂O₇ 0,25 mol L⁻¹, 10 mL de H₂SO₄ concentrado), 5 mL de ácido fosfórico e 3 gotas de difenilamina. Titulou-se até a cor mudar de violeta para verde.

O método apresentado é Walkley-Black com calor externo com pequenas modificações a fim de aumentar a quantidade de material e facilitar a execução das análises.

Variáveis hiperspectrais

As folhas que passaram pela avaliação fisiológica também foram submetidas à análise espectral. Tal análise foi realizada com espectrorradiômetro (FieldSpec 4 Jr, Analytical Spectral Devices, Boulder, EUA). O equipamento tem a capacidade de medir o espectro na faixa de 350 a 2500 nm. Além disso, foi utilizado o equipamento ASD Plant Probe, que é projetado com a finalidade da medição espectral por meio do contato em materiais sólidos. A calibração do ASD Plant Probe se deu por meio de uma placa branca de sulfato de bário que reflete 100% da luz. O equipamento foi calibrado sempre após o término da leitura de um bloco, totalizando 4 calibrações.

O sensor foi conectado a um computador para os registros de cada leitura por meio de um software RS³, que é próprio do sensor e registra as leituras feitas pelo equipamento. Na sequência, os arquivos foram importados a outro software, ViewSpectroPro, para extração dos dados em extensão .txt. Uma das principais vantagens desse método é que ele não é destrutivo e não sofre interferência da luminosidade do ambiente, o que reduz erros ocasionado pela luz difusa.

Aprendizagem de máquina

O tratamento estatístico aplicado aos dados ocorreu por meio de análises de aprendizagem de máquina (ML). Os modelos de ML utilizados na predição foram: Redes neurais artificiais (MP, EGMONT-PETERSEN et al., 2002); Árvore de decisão REPTree (DT, SNOUSY et al., 2011); Árvore de decisão M5P (BLAIFI et al., 2018); Floresta aleatória (RF, BELGIU & DRĂGUȚ, 2016); Máquina de vetor suporte (SVM, NALEPA & KAWULOK, 2019) e Zero R, utilizado como modelo controle.

A variável predita (output) foi o fluxo de dióxido de carbono (FCO_2), enquanto os dados hiperspectrais foram utilizados como variáveis de entrada (input) nos modelos. A predição ocorreu de acordo com a validação cruzada estratificada com $k\text{-fold} = 10$ e dez repetições (100 execuções para cada modelo). Todos os parâmetros dos modelos foram estabelecidos pela configuração default do software Weka 3.8.5., com exceção para Redes Neurais, que obteve quatro neurônios por camada e duas camadas.

Os parâmetros de desempenho foram: métricas de coeficiente de correlação (r), erro absoluto médio (MAE) e raiz quadrada do erro médio (RMSE). Para verificar a significância dos algoritmos na predição foi realizada uma análise de variância. Nos casos de significância, foram gerados boxplots com as médias de r , MAE e RMSE, agrupados pelo teste de Scott-Knott (Scott & Knott, 1974) à de 5% de probabilidade. O agrupamento das médias e os boxplots foram gerados por meio dos pacotes ggplot2 e ExpDes.pt no software R.

RESULTADOS E DISCUSSÃO

A predição da emissão in situ de CO₂ do solo (FCO₂) por meio de variáveis espectrais das folhas torna-se mais complexa e de difícil mensuração que as anteriores, visto que os dados input e output são coletados em regiões dispersas do sistema de produção. Ainda assim, o modelo RF destacou-se com a maior porcentagem para o parâmetro r e menores erros MAE e RMSE. Os modelos menos adequados à predição foram Zero R e MP. Este expressou grande dispersão dos dados para o parâmetro r, e os maiores erros MAE e RMSE. Enquanto Zero R não possibilitou a análise. Os demais modelos encontraram-se com resultados intermediários, menos prováveis à recomendação (Figura 8).

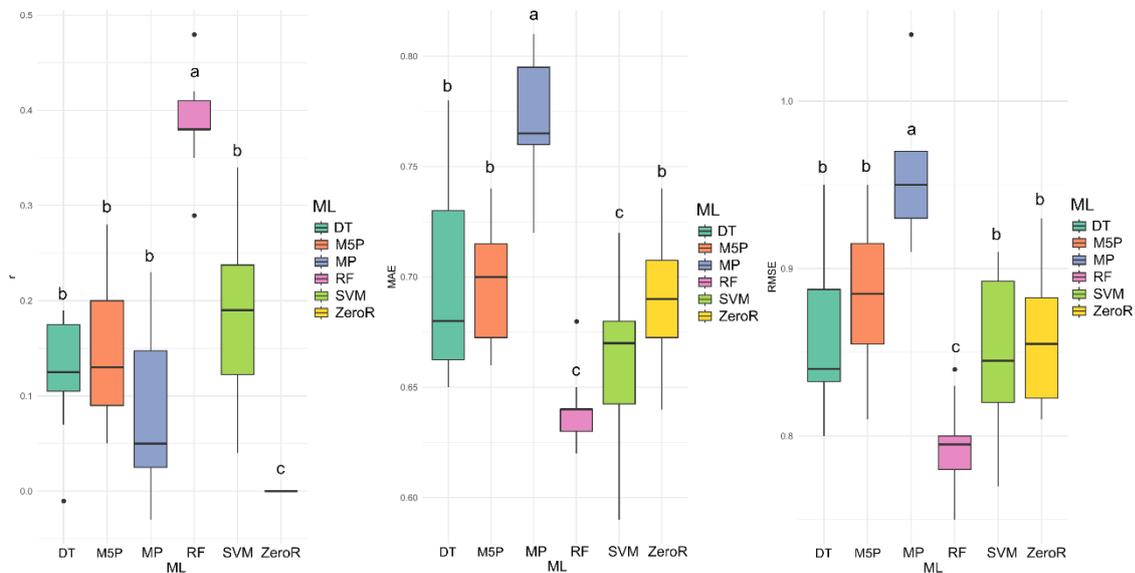


Figura 8. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável emissão in situ de CO₂ do solo (FCO₂).

Em se tratando de FCO₂, além de retrain carbono em matéria orgânica, o solo em sua porção gasosa comporta-se como um excelente armazenador de gás carbônico, o que pode ser constatado em estudos sob áreas nativas do Cerrado (COSTA et al., 2018). Entretanto, os autores afirmam que o substrato pode atuar tanto como sumidouro do gás na atmosfera, quanto como emissor, e isso varia de acordo com uma série de fatores. Distúrbios naturais e antrópicos compõem esses fatores (PAIVA A. O. et al., 2011), tais como temperatura, umidade, preparo e manejo do solo (BRÖRING, 2013).

Diante dessa série de fatores, verifica-se uma das justificativas da baixa porcentagem de correlação ao predizer o carbono no solo. A complexidade por trás dessas informações numéricas é abrangente, desde a questão da microbiota do substrato até fenômenos naturais mais abrangentes.

Mas, alinhado ao objetivo desta pesquisa, outro fator torna-se relevante para o estudo do carbono no solo, que é a vegetação, podendo potencializar o processo de sequestro e/ou emissão do gás (OLIVEIRA et al., 2023). As plantas, além de fixar carbono fotossintetizado à matéria orgânica em caule e raízes (COSTA et al., 2018), contribui para o sistema microbiológico que atua na reação do nutriente a gás.

Isso revela que a escolha da cultura e das cultivares deve ser estrategicamente minuciosa para que se tenha um manejo alinhado à sustentabilidade. Tudo isso requer a análise do comportamento da planta com o meio, com outros tipos de vida, e com o meio físico do solo.

Para a predição de fixação de carbono foliar por meio de variáveis hiperespectrais os resultados foram os inferiores dentre toda esta pesquisa. Acredita-se que isso ocorreu devido a divergência de fatores que atuam entre o input e o output. É interessante analisar esses dados, pois a fotossíntese se correlacionou com os dados hiperespectrais, e o carbono, que está inteiramente relacionado à fotossíntese, não obteve tais resultados. Talvez uma investigação mais assídua sobre as rotas metabólicas no processo de construção do tecido foliar auxiliasse na interpretação desses dados. Mesmo assim, pode-se selecionar os algoritmos que se sobressaíram.

De modo geral, M5P e RF foram as melhores técnicas para o parâmetro r e erros MAE e RMSE. Embora elas não tenham se diferenciado das demais pelo teste Scott-Knot, foram as que expressaram melhores valores e menor dispersão de dados para todos os parâmetros (Figura 9).

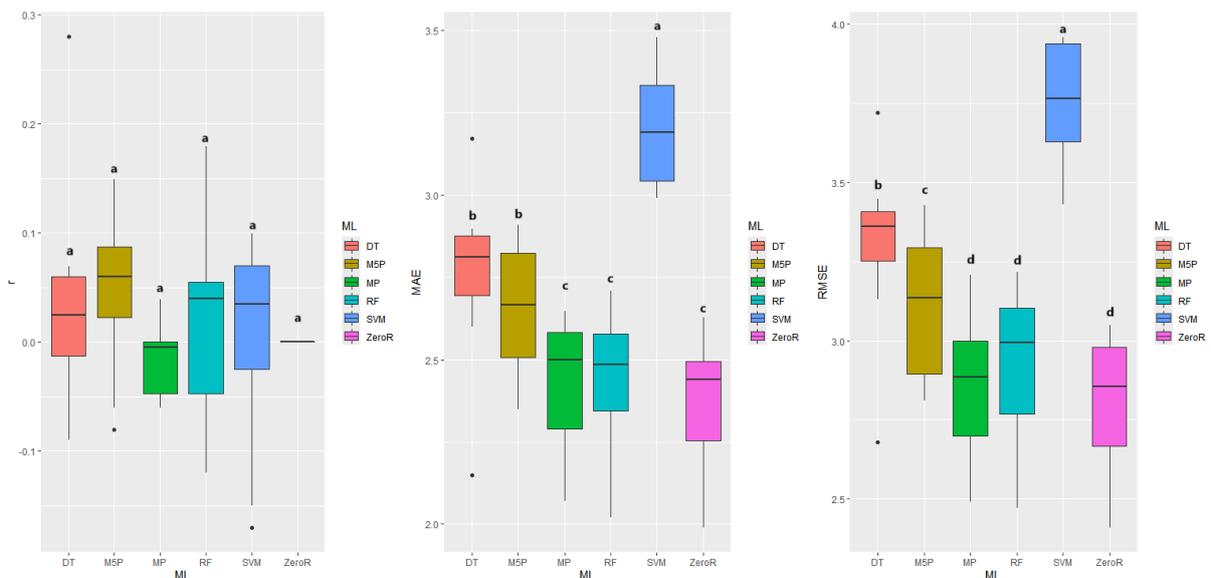


Figura 9. Boxplots para coeficiente de correlação de Pearson (r), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE) para diferentes modelos de aprendizado de máquina para a variável Fixação de CO₂ pelo tecido foliar.

As variáveis fisiológicas abordadas no Capítulo 1 (fotossíntese líquida, condutância estomática, transpiração, concentração interna de CO₂, eficiência instantânea de uso de água e eficiência instantânea de carboxilação) são dependentes entre si (SILVA et al., 2015) e associam-se diretamente à fixação de carbono foliar. O processo de carboxilação é a fixação de CO₂ no ciclo de redução fotossintética do carbono (NELSON; COX, 2014). Isso permite inferir em uma predição indireta, visto que a correlação entre dados hiperespectrais e fixação de carbono não tenha sido satisfatória.

Em relação à machine learning, a técnica de Floresta aleatória conseguiu estimar em torno de 40% de assertividade o fluxo de carbono no solo por meio das variáveis fisiológicas. Este resultado expressa a possibilidade de novas abordagens neste sentido, embora a literatura ainda careça de estudos aprofundados sobre a correlação entre fisiologia vegetal e comportamento físico do substrato, como a retenção de um gás. Além de que o surgimento da inteligência computacional associada ao sensoriamento remoto permite abordar a sustentabilidade às práticas agrícolas (OGUNGBUYI et al., 2023).

Entretanto, o sensoriamento hiperespectral não é possível ser trabalhado com processamentos estatísticos de modelos convencionais em detrimento da extensão da base de dados. Neste trabalho, por exemplo, as variáveis de entrada foram 2.150 comprimentos de onda. Isto reflete no baixo aproveitamento do modelo Zero R, visto que é uma regressão linear. Desta maneira, faz-se necessária a abordagem de técnicas de aprendizagem de máquina (ML) com regressões multivariadas, como Redes neurais artificiais (MP), Árvore de decisão REPTree (DT), Árvore de decisão (M5P), Floresta aleatória (RF) e Máquina de vetor suporte (SVM).

Para estas análises M5P destacou-se. Ela trata-se de um algoritmo que trabalha com gráficos de ramificações e pontos convergentes e divergentes entre os dados, tanto para regressão quanto para classificação (MONARD E BARANAUSKAS, 2003). Floresta aleatória (RF) que não ficou de fora dos melhores resultados consta do conjunto de Árvores de decisão. Essa junção de diversas árvores diminui a ocorrência de overfitting e aumenta a precisão das análises.

RF tem se destacado nas ciências agrárias, sobretudo em agricultura de precisão devido ao seu bom desempenho como algoritmo preciso (RAMOS et al., 2020). O algoritmo foi estabelecido como a melhor técnica de Machine Learning para predizer o crescimento e altura de espécies vegetais por meio de dados espectrais pelos autores (DA SILVA et al., 2021; OSCO et al., 2020; TEODORO et al., 2021), variáveis fisiológicas essas que se associam à fixação de carbono nos tecidos vegetais.

Associar FCO₂ e fixação de carbono nos tecidos foliares às variáveis espectrais apresenta grande complexidade e é uma abordagem inovadora no âmbito científico. As

vantagens dessa relação são rapidez e acurácia no processamento de dados a fim de se potencializar práticas sustentáveis. Neste viés, a mensuração do gás no solo e a incorporação de carbono atmosférico pelas plantas são importantes para a tomada de decisão à cerca das culturas, cultivares, manejo e sistemas de produção que possam servir como sumidouro de carbono.

Portanto, modelos de predição de FCO_2 e fixação de carbono nos tecidos foliares por meio do espectro de reflexão das folhas auxiliarão na velocidade e qualidade de avaliações à campo. Tudo isso devendo-se ao sensoriamento remoto sob o uso de sensores acoplados às aeronaves remotamente pilotadas, ou por meio de informações de satélites, acoplado à inteligência computacional.

CONCLUSÕES

Embora as abordagens sobre FCO₂ e fixação de carbono no tecido foliar da soja não tenham tido resultados satisfatórios, elas comporta-se como uma sugestão para futuras pesquisas e revelam que os melhores modelos de aprendizagem de máquina neste emprego são Árvore de decisão M5P e Floresta aleatória.

REFERÊNCIAS

BANDARU, V. Climate data induced uncertainties in simulated carbon fluxes under corn and soybean systems. **Agricultural Systems**, v. 196, p. 103341, 2021. Doi:

<https://doi.org/10.1016/j.agry.2021.103341>

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BLAIFI, S. et al. M5P model tree based fast fuzzy maximum power point tracker. **Solar Energy**, v. 163, p. 405–424, 2018.

BRÖRING, J. M. **Dinâmica do carbono em sistemas de uso do solo no oeste e planalto de Santa Catarina**. Lages: Universidade do estado de Santa Catarina – UDESC, 2013.

CONAB. Acompanhamento da safra brasileira: grãos - safra 2022/23 – oitavo levantamento. Brasília: **Companhia Nacional de Abastecimento**, v. 10, n. 8, 2023.

COSTA, T. G. A. et al. Dinâmica de carbono do solo em unidade de conservação do cerrado brasileiro sob diferentes fitofisionomias. **Revista Gestão & Sustentabilidade Ambiental**, v. 7, n. 4, p. 306, 12 dez. 2018.

DA SILVA, A. K. V. et al. Predicting eucalyptus diameter at breast height and total height with uav-based spectral indices and machine learning. **Forests**, v. 12, n. 5, 2021.

EGMONT-PETERSEN, M. et al. Image processing with neural networks - a review. **Pattern Recognition**, v. 35, n. 10, p. 2279–2301, 2002.

INMET. <<https://bdmep.inmet.gov.br/>>.

LIU, C.; SUN, P. SEN; LIU, S. R. A review of plant spectral reflectance response to water physiological changes. **Chinese Journal of Plant Ecology**. Editorial Office of Chinese. 2016.

MONARD M. C.; BARANAUSKAS J. A. Indução de Regras e Árvores de Decisão. In: MONARD M. C.; BARANAUSKAS J. A. **Sistemas inteligentes-Fundamentos e aplicações**: Conceitos sobre aprendizado de máquina. v. 1, n. 1, p. 57-74, 2003.

NALEPA, J.; KAWULOK, M. Selecting training sets for support vector machines: a review. **Artificial Intelligence Review**, v. 52, n. 2, p. 857–900, 2019.

OLIVEIRA, C. L. et al. A Caatinga Emite ou Sequestra Carbono? **Revista Brasileira de Geografia Física**, v. 16, n 02, p. 791-804, 2023.

OSCO, L. P. et al. Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. **Remote Sensing**, v. 12, n. 19, p. 1–17, 2020.

PAIVA A. O. et al. Estoque de carbono em cerrado sensu stricto do Distrito Federal. **Revista Árvore**, Viçosa-MG, v. 35, n. 3, p. 527-538, 2011.

PARKINSON, K. J. An improved method for measuring soil respiration in the field. **Journal of applied ecology**, p. 221–228, 1981.

ROSSI, F. S. et al. 19-year remotely sensed data in the forecast of spectral models of the environment. **International Journal of Digital Earth**, v. 14, n. 8, p. 1040–1066, 2021.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. 6ª ed. Porto Alegre: Artmed, 2014. 1250 p.

OGUNGBUYI, M. G. et al. Enabling Regenerative Agriculture Using Remote Sensing and Machine Learning. **Land**, v. 12, n. 6, 2023.

ONU. Relatório anual de 2024. **Organização das Nações Unidas**. 2024. <
<https://brasil.un.org/pt-br/279199-relat%C3%B3rio-anual-das-na%C3%A7%C3%B5es-unidas-2024>>

RAMOS, A. P. M. 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.

RUTKOSKI, J. et al. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. **G3: Genes, Genomes, Genetics**, v. 6, n. 9, p. 2799–2808, 2016.

SCOTT, A. A. J.; KNOTT, M. A Cluster Analysis Method for Grouping Means in the Analysis of Variance Published by: International Biometric Society Stable. **Biometrics**, v. 30, n. 3, p. 507–512, 1974.

SILVA, F. G. et al. Gas exchange and chlorophyll fluorescence of eggplant grown under different irrigation depths. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 10, p. 946–952, 1 set. 2015.

SNOUSY, M. B. AL et al. Suite of decision tree-based classification algorithms on cancer gene expression data. **Egyptian Informatics Journal**, v. 12, n. 2, p. 73–82, 2011.

TEODORO, P. E. et al. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. **Remote Sensing**, v. 13, n. 22, 2021.

TEODORO, P. E. et al. Soil CO₂ emissions under different land-use managements in Mato Grosso do Sul, Brazil. **Journal of Cleaner Production**, v. 434, p. 139983, 2024.