

ROBSON ARGOLO DOS SANTOS

**MODELAGEM DA EVAPOTRANSPIRAÇÃO DA CANA-DE-AÇÚCAR
UTILIZANDO ABORDAGENS DE SENSORIAMENTO REMOTO E *MACHINE
LEARNING***

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

Orientador: Everardo Chartuni Mantovani

Coorientadores: Vinícius Bof Bufon
Christopher Michael Usher Neale

**VIÇOSA - MINAS GERAIS
2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

Santos, Robson Argolo dos, 1990-
S237m Modelagem da evapotranspiração da cana-de-açúcar
2023 utilizando abordagens de sensoriamento remoto e *machine
learning* / Robson Argolo dos Santos. – Viçosa, MG, 2023.
1 tese eletrônica (121 f.): il. (algumas color.).

Orientador: Everardo Chartuni Mantovani.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2023.
Referências bibliográficas: f. 110-121.
DOI: <https://doi.org/10.47328/ufvbbt.2023.576>
Modo de acesso: World Wide Web.

1. Evapotranspiração - Medição. 2. Cana-de-açúcar.
3. Aprendizado do computador. 4. Sensoriamento remoto.
I. Mantovani, Everardo Chartuni, 1958-. II. Universidade Federal
de Viçosa. Departamento de Engenharia Agrícola. Programa de
Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 551.572


ROBSON ARGOLO DOS SANTOS

**MODELAGEM DA EVAPOTRANSPIRAÇÃO DA CANA-DE-AÇÚCAR
UTILIZANDO ABORDAGENS DE SENSORIAMENTO REMOTO E *MACHINE
LEARNING***


Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

APROVADA: 28 de julho de 2023

Assentimento:

Documento assinado digitalmente
 **ROBSON ARGOLO DOS SANTOS**
Data: 20/09/2023 16:02:42-0300
Verifique em <https://validar.iti.gov.br>

Robson Argolo dos Santos
Autor

Documento assinado digitalmente
 **EVERARDO CHARTUNI MANTOVANI**
Data: 20/09/2023 15:59:39-0300
Verifique em <https://validar.iti.gov.br>

Everardo Chartuni Mantovani
Orientador

AGRADECIMENTOS

Agradeço a Deus por permitir esta grandiosa conquista!

Agradeço aos meus pais, Romilton Jesus dos Santos e Nilzete Argolo dos Santos, pela confiança, esforço, dedicação e ensinamentos.

Também quero expressar minha gratidão aos meus irmãos, Romário Argolo dos Santos, Ronald Argolo dos Santos, à minha namorada, Myriam Marta Soares, e a todos os familiares que, direta ou indiretamente, proporcionaram o caminho até aqui.

Lembro com carinho do meu avô paterno, Antônio, avô materno, Artur (*in memoriam*), minha avó materna, Idalina (*in memoriam*), minha avó paterna, Luísa (*in memoriam*), e minha tia, Berenisse (*in memoriam*), que desempenharam papéis importantes em minha formação.

À Universidade Estadual de Feira de Santana, que me capacitou e concedeu o título de Engenheiro Agrônomo, meu sincero agradecimento.

Gostaria de expressar minha profunda gratidão à minha orientadora de iniciação científica e TCC, Rosângela Leal.

Aos amigos do curso de Agronomia da UEFS, especialmente da primeira turma (veteranos de todos, calouros de ninguém), agradeço por compartilharmos essa jornada.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Engenharia Agrícola, meu agradecimento pela oportunidade de capacitação.

Ao meu orientador, Professor Everardo Chartuni Mantovani, agradeço pelos ensinamentos, amizade e apoio inestimáveis durante o mestrado e o doutorado.

Aos meus coorientadores, Vinicius Bof Bufon e Professor Elpídio Inácio Fernandes Filho, agradeço o valioso auxílio durante a execução da pesquisa.

Quero expressar minha gratidão à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudos, meu sincero agradecimento.

Aos meus amigos da pós-graduação e a todos os colegas e professores do GESAI e do CRRH, agradeço a amizade e companheirismo.

Não posso esquecer de mencionar os funcionários da Universidade Federal de Viçosa que, direta e indiretamente, auxiliaram em todas as etapas durante o doutorado.

Enfim, a todos que me ajudaram de alguma forma na execução desse projeto, fica minha total gratidão.

BIOGRAFIA

ROBSON ARGOLO DOS SANTOS nasceu em Taperoá, Bahia, no dia 22 de fevereiro de 1990. Filho de Romilton Jesus dos Santos e Nilzete Argolo dos Santos e tem dois irmãos: Romário Argolo dos Santos e Ronald Argolo dos Santos. Em março de 2012, ingressou no curso de Engenharia Agrônômica na Universidade Estadual de Feira de Santana. Concluiu o curso em julho de 2017, após a defesa do Trabalho de Conclusão de Curso (TCC) intitulado “Avaliação e ajuste de métodos alternativos de estimativa da evapotranspiração de referência a partir da equação de Penman-Monteith-FAO56 em Feira de Santana, Bahia”, sob a orientação da Professora Doutora Rosângela Leal. Em agosto de 2017, deu início ao curso de Mestrado em Engenharia Agrícola, com ênfase na área de Recursos Hídricos e na linha de pesquisa de Manejo de Água/Planta na Agricultura Irrigada. Defendeu a Dissertação em julho de 2019, com o título “Demanda hídrica e modelagem dos parâmetros biofísicos na cultura do milho a partir do sensor RGNIR a bordo de um VANT”, sendo orientado pelo Professor Doutor Everardo Chartuni Mantovani e coorientado pelo Professor Doutor Elpídio Inácio Fernandes-Filho. Em agosto de 2019, iniciou o curso de Doutorado em Engenharia Agrícola, com ênfase na área de Recursos Hídricos e na linha de pesquisa de Manejo de Água/Planta na Agricultura Irrigada. Em julho de 2023 submeteu-se à defesa da Tese de doutorado com título “Modelagem da evapotranspiração da cana-de-açúcar utilizando abordagens de sensoriamento remoto e machine learning” alcançando a aprovação dos cinco membros da banca.

“No agro vencemos etapas, temos programas, geramos soluções e mantemos viva a economia brasileira.”

(Alysson Paolinelli - Ex-ministro da Agricultura)

RESUMO

SANTOS, Robson Argolo dos, D.Sc., Universidade Federal de Viçosa, julho de 2023. **Modelagem da evapotranspiração da cana-de-açúcar utilizando abordagens de sensoriamento remoto e *machine learning***. Orientador: Everardo Chartuni Mantovani. Coorientadores: Vinícius Bof Bufon e Christopher Michael Usher Neale.

O uso de sensoriamento remoto combinando informações de estações meteorológicas tem sido estudado como uma alternativa a estimativa da demanda hídrica, porém algumas limitações ainda persistem, requerendo novas metodologias que possa superá-las. Desta forma, essa tese teve como objetivo geral estimar a ET_a (evapotranspiração atual) da cana-de-açúcar utilizando *machine learning* com base em dados obtidos por sensoriamento remoto ativo e passivo, bem como múltiplas fontes de informação. Os objetivos específicos foram divididos em três capítulos: 1) Estimar a evapotranspiração utilizando *machine learning* e sensoriamento remoto passivo; 2) Estimar a evapotranspiração utilizando *machine learning* e sensoriamento remoto passivo sem a necessidade de estação meteorológica em campo; e 3) Estimar a evapotranspiração utilizando *machine learning* e sensoriamento remoto ativo na presença de nuvens. Para tal, foram adquiridos dados dos sensores OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*) embarcados nos satélites Landsat 8 e 9, a fim de determinar a ET_{rF} (fração evapotranspirativa) e posterior ET_a utilizando o modelo METRIC (*Mapping Evapotranspiration at High Resolution with Internalized Calibration*). Essas variáveis foram utilizadas como resposta, enquanto as variáveis explicativas foram obtidas dos sensores MSI (*Multispectral Instrument*), embarcado no Sentinel-2, e SAR (*Synthetic Aperture Radar*), no Sentinel-1. Os dados meteorológicos, climatológicos, de solos, elevação do terreno e astronômicas, foram utilizados para garantir um treinamento e desempenho adequados dos algoritmos de *machine learning*. No primeiro capítulo, foi possível observar a capacidade dos modelos de *machine learning* em estimar a ET_{rF} utilizando dados do sensor MSI, que não possui banda termal. O modelo que obteve melhor desempenho nessa abordagem foi o XgbLinear, com $R^2 = 0.91$ e $RMSE = 0.10$, apresentando resultados satisfatórios. Os valores estimados de ET_{rF} , neste capítulo, quando multiplicados pela ET_r (Evapotranspiração de referência alfa), forneceram a ET_a , que é utilizada para determinar a reposição adequada da irrigação. Os resultados do segundo capítulo mostraram que os modelos de *machine learning*, utilizando dados de sensoriamento remoto passivo sem banda termal em conjunto com outras variáveis explicativas, permitiu quantificar diretamente a ET_a , sem a necessidade do valor de ET_r determinado via estação meteorológica. Dos 13 modelos utilizados neste capítulo, o brnn

(*Bayesian Regularized Neural Networks*) se destacou, com R^2 de 0.73 e RMSE de 1.10, respectivamente. Por fim, os resultados do terceiro capítulo foram promissores, pois foi possível modelar e quantificar a ET_a na cana-de-açúcar mesmo na presença de nuvens. O brnn novamente se destacou, apresentando R^2 de 0.45 e RMSE de 1.5. Apesar das métricas inferiores, devido às limitações espectrais do sensor SAR, esse resultado representa um grande avanço, uma vez que o sensor SAR pode ser utilizado quando não é possível o uso de sensores remotos passivos. Com base em todos esses resultados, é possível inferir que os algoritmos de *machine learning* XgbLinear e brnn, devidamente treinados com dados obtidos por sensoriamento remoto passivo e ativo, juntamente com outras fontes de informação, podem ser eficazes na estimativa da evapotranspiração atual da cana-de-açúcar.

Palavras-chave: Satélite. *Machine learning*. Agricultura irrigada.

ABSTRACT

SANTOS, Robson Argolo dos, D.Sc., Universidade Federal de Viçosa, July 2023. **Modeling sugarcane evapotranspiration using remote sensing and *machine learning* approaches.** Adviser: Everardo Chartuni Mantovani. Co- advisers: Vinícius Bof Bufon and Christopher Michael Usher Neale.

The use of remote sensing, combined with information from weather stations, has been studied as an alternative method to estimate water demand. However, some limitations persist, necessitating the development of new methodologies to overcome them. Therefore, the main objective of this thesis was to estimate the actual evapotranspiration (ET_a) of sugarcane using *machine learning* based on data obtained from both active and passive remote sensing, as well as multiple sources of information. The specific objectives were divided into three chapters: 1) Estimating evapotranspiration using *machine learning* and passive remote sensing; 2) Estimating evapotranspiration using *machine learning* and passive remote sensing without the need for a field weather station; e 3) Estimating evapotranspiration using *machine learning* with active remote sensing even in the presence of clouds. Data from the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) sensors on Landsat 8 and 9 satellites were acquired to determine the evapotranspiration fraction (ET_rF) and subsequent ET_a using the Mapping Evapotranspiration at High Resolution with Internalized Calibration (METRIC) model. These variables were used as responses, while explanatory variables were obtained from the Multispectral Instrument (MSI) sensor on Sentinel-2 and Synthetic Aperture Radar (SAR) on Sentinel-1. Meteorological, climatological, soil, terrain elevation, and astronomical data were utilized to ensure proper training and performance of *machine learning* algorithms. In the first chapter, we observed the capability of *machine learning* models to estimate ET_rF using MSI sensor data, which lacks thermal bands. The model that performed best in this approach was XgbLinear, achieving an R^2 of 0.91 and RMSE of 0.10, yielding satisfactory results. The estimated values of ET_rF in this chapter, when multiplied by ET_r (reference evapotranspiration alfalfa), provided ET_a , which is crucial for determining appropriate irrigation replenishment. The results of the second chapter demonstrated that *machine learning* models, using passive remote sensing data without thermal bands in conjunction with other explanatory variables, allowed for the direct quantification of ET_a without the need for ET_r values determined via weather stations. Among the 13 models used in this chapter, Bayesian Regularized Neural Networks (brnn) stood out with an R^2 of 0.73 and RMSE of 1.10, respectively. Finally, the results of the third chapter were promising, as it was possible to model and quantify ET_a in

sugarcane even in the presence of clouds. The brnn model once again stood out, with an R^2 of 0.45 and RMSE of 1.5. Despite the lower METRICs due to the spectral limitations of the SAR sensor, this result represents a significant advancement since SAR can be employed when passive remote sensors cannot. Based on all these results, it can be inferred that *machine learning* algorithms, specifically XgbLinear and brnn, when properly trained with data obtained from passive and active remote sensing, along with other sources of information, can be effective in estimating the actual evapotranspiration of sugarcane.

Keywords: Satellite. Machine learning. Irrigated agriculture.

SUMÁRIO

1	Introdução geral	11
2	Artigos científicos	15
2.1	Artigo 1	15
1	Introdução	16
2	Materiais e Métodos	18
3	Resultados e Discussão	29
4	Conclusão	42
2.2	Artigo 2	43
1	Introdução	44
2	Materiais e Métodos	46
3	Resultados e Discussão	59
4	Conclusão	75
2.3	Artigo 3	76
1	Introdução	77
2	Materiais e Métodos	80
3	Resultados e Discussão	93
4	Conclusão	106
3	Conclusões gerais	108
	Referências	110

1 Introdução geral

A cana-de-açúcar possui participação fundamental no agronegócio brasileiro, ocupando vastas áreas em todo o país, principalmente nos biomas da Mata Atlântica, Cerrado e Caatinga. De acordo com os dados da Companhia Nacional de Abastecimento (CONAB, 2023), na safra 2022/2023, a área plantada com cana-de-açúcar alcançou cerca de 8.29 milhões de hectares, e para a safra 2023/2024, estima-se uma expansão para 8.41 milhões de hectares, representando um aumento de 1.5%, desempenhando um papel essencial na produção de açúcar, utilizado na indústria alimentícia, etanol e derivados, destinados ao setor energético. Além disso, possibilita a geração de diversos produtos secundários, como biomassa, biofertilizantes, biometano, bioquerosene, bioeletricidade e hidrogênio verde. Apesar da diversidade de produtos e subprodutos, o açúcar é que se destaca nas exportações brasileiras. Segundo os dados apresentados pelo Instituto de Pesquisa Econômica Aplicada (IPEA, 2023), em 2022, foram exportadas 1.35 milhões de toneladas, gerando uma receita de 520 milhões de dólares. Para o ano de 2023, os dados foram ainda mais favoráveis, com um aumento de 50% nas exportações (2.03 milhões de toneladas) e um aumento de 68% na receita (872 milhões de dólares).

Para alcançar essa expressividade nacional e relevância econômica, foram necessários investimentos significativos em estudos e pesquisas, que têm proporcionado inovações tecnológicas refletindo em melhores resultados. Dentre as tecnologias desenvolvidas e aprimoradas no setor sucroenergético brasileiro, a introdução da irrigação na matriz de produção tem permitido, em curto e médio prazo, vantagens em termos de aumento de eficiência agrônômica, financeira e ambiental, além de reduzir a pressão por expansão de novas áreas (BUFON *et al.*, 2021).

Conforme dados da Agência Nacional de Águas e Saneamento Básico (ANA, 2021), o território brasileiro dedicado à irrigação e fertirrigação é cerca de 8.2 milhões de hectares. Desse total, aproximadamente 5.3 milhões de hectares são direcionados para irrigação, enquanto os restantes 2.9 milhões de hectares são utilizados predominantemente para fertirrigação, em grande parte na cultura da cana-de-açúcar para aplicação de vinhaça. Contudo, é previsto que até 2040 a extensão total dedicada a essas práticas alcance os 12.4 milhões de hectares, com um crescimento concentrado em irrigação. A projeção aponta um aumento significativo, elevando a área irrigada de 5.3 para 9.5 milhões de hectares, o que equivale a um impressionante acréscimo de 79% no total. Em relação à cultura da cana-de-açúcar, a área total destinada é de aproximadamente 3.65 milhões de hectares, sendo os 2.9 milhões da fertirrigação e o restante ocupado por sistemas de irrigação para salvamento (15.5%), caracterizados por baixas lâminas

de irrigação em épocas críticas para a rebrota; suplementação (4.2%), com lâminas de irrigação mais substanciais, porém que não atendem a 100% da demanda hídrica; e irrigação plena (0.8%), com lâminas de irrigação devidamente controladas para suprir toda a necessidade hídrica da planta. Os 0.75 milhão de hectares representam 14% da área total de irrigação no Brasil, e estima-se que esse número alcance 1.13 milhão de hectares até 2040 (ANA, 2021), o que corresponde a um incremento de 50.67%. Este notável crescimento reflete os avanços tecnológicos na tecnologia da irrigação na produção de cana-de-açúcar. No entanto, é importante destacar que esse progresso também acarretará uma demanda crescente de recursos hídricos em um contexto em que as discussões ambientais sobre o uso e a qualidade da água estão ganhando cada vez mais destaque.

Em função das discussões e conflitos sobre o uso da água, é necessário aprimorar a gestão hídrica para garantir uma maior eficiência no seu uso. Nos sistemas irrigados, há constantemente necessidade de estimar a demanda de água perdida por unidade de área, a fim de repor adequadamente, evitando o uso excessivo desse recurso. A necessidade hídrica de uma área irrigada ou de cultivo de sequeiro é determinada por meio de métodos meteorológicos (via clima), instrumentação de solo (via solo) ou instrumentação de plantas (via plantas), ou ainda alguma combinação desses três métodos (BERNADO *et al.*, 2019). O método meteorológico requer a presença de estações meteorológicas no campo, que utilizam sensores para captar as condições atmosféricas e calcular a demanda evapotranspirativa da atmosfera por meio da equação da ET_0 (evapotranspiração de referência) (ALLEN *et al.*, 1998). Considerando o coeficiente de cultura único para o coeficiente de transpiração e evaporação (K_c), multiplicando-se a ET_0 por ele e pelo K_s (coeficiente de estresse), obtém-se o conceito de ET_a (evapotranspiração atual), que representa a quantidade de água perdida pelo processo evapotranspirativo na área cultivada. No entanto, esse método apresenta duas limitações. A primeira é o custo elevado para implementação e manutenção das estações meteorológicas em campo. A segunda limitação está relacionada ao fato de que o K_c é um coeficiente tabelado, logo estático, que, geralmente, é utilizado em três fases do ciclo da cultura, o que pode não representar adequadamente a evapotranspiração atual da cultura. Já os métodos que utilizam instrumentação de solo e planta são mais onerosos e, na maioria dos casos, são empregados em estudos e pesquisas, tornando-se menos viáveis para a aplicação em larga escala. Portanto, os métodos baseados no clima são amplamente utilizados por conta de sua praticidade e menor custo (BERNADO *et al.*, 2019).

Devido às desvantagens mencionadas anteriormente, têm sido realizados estudos com o objetivo de utilizar o sensoriamento remoto como uma alternativa aos métodos convencionais

na estimativa da evapotranspiração atual. O sensoriamento remoto possibilita o desenvolvimento de modelos híbridos, que combinam informações provenientes de estações meteorológicas em campo, permitindo assim a quantificação da ET_a . Diversos métodos já foram desenvolvidos, como o TSEB - *Two-Source Energy Balance* (NORMAN; KUSTAS; HUMES, 1995), SEBAL - *Surface Energy Balance Algorithm for Land* (BASTIAANSEN, 1999), SSEBI - *Simplified Surface Energy Balance Index* (ROERINK; SU; MENENTI, 2000), METRIC - *Mapping EvapoTranspiration at high Resolution with Internalized Calibration* (ALLEN; TASUMI; TREZZA, 2007) e SAFER - *Simple Algorithm for Evapotranspiration Retrieving* (TEIXEIRA, 2010). Contudo, esses métodos apresentam três desvantagens. A primeira é a dependência de um sensor termal, o qual não está presente em muitos satélites ou outras plataformas. A segunda é que eles são limitados quando há presença de nuvens, uma vez que as nuvens bloqueiam a radiação refletida e/ou emitida pela superfície. Por fim, a terceira é a necessidade de uma estação meteorológica em campo.

Em relação à necessidade de um sensor termal, Dos Santos *et al.* (2022) e Costa *et al.* (2023) utilizaram técnicas de *machine learning* e dados do sensor MSI (*Multispectral Instrument*), embarcado no Sentinel-2, que não possui banda termal, para estimar a ET_{rF} (Fração evapotranspirativa), superando a limitação dos dados térmicos. No entanto, ainda enfrentam limitações quanto à necessidade de uma estação meteorológica e à presença de nuvens. Desta forma, torna-se viável estimar a evapotranspiração atual mesmo em condições de nebulosidade utilizando informações obtidas por meio do sensor SAR (*Radar de Abertura Sintética*), um dispositivo ativo capaz de emitir micro-ondas. Isso se deve à característica dessas micro-ondas, que conseguem penetrar as camadas de nuvens e interagir com os elementos na superfície terrestre (CHINTALA *et al.*, 2022; DOS SANTOS *et al.*, 2022a; FILGUEIRAS *et al.*, 2019a; SHORACHI; KUMAR; STEELE-DUNNE, 2022). Quanto à solução para evitar o uso de uma estação meteorológica, é possível utilizar fontes de dados adicionais, além do sensoriamento remoto, como dados meteorológicos, do produto Nasa Power, e climatológicos, dos produtos do WordClim, além de dados que contenham informações sobre as características físicas e hídricas do solo, estabelecendo assim uma relação solo-planta-atmosfera. Essa integração de múltiplas fontes de dados para modelar uma variável de interesse requer modelos complexos com essa finalidade e potencialidade, sendo os algoritmos de *machine learning* ferramentas adequadas para essa tarefa.

O *machine learning* é um conjunto de algoritmos baseados em inteligência artificial que permite aos sistemas computacionais adquirirem conhecimento e aprimorarem seu desempenho por meio da análise de dados. O algoritmo utiliza uma abordagem indutiva, na qual os sistemas

aprendem a partir de exemplos e experiências prévias, em vez de serem programados explicitamente para realizar uma determinada tarefa (BATTA, 2018). Isso possibilita que as máquinas se adaptem e evoluam à medida que são expostas a novos dados sendo essenciais na modelagem de conjuntos de dados complexos.

Considerando a relevância crucial da água no contexto dos sistemas de irrigação no setor sucroenergético, juntamente com a análise das vantagens e desvantagens dos distintos métodos de estimativa da evapotranspiração, e diante do notável potencial apresentado pelos algoritmos de *machine learning*, o presente estudo tem como principal propósito investigar a capacidade desses algoritmos em prever a evapotranspiração atual na cultura da cana-de-açúcar. Nesse sentido, o objetivo fundamental desta tese consiste em empregar técnicas de *machine learning* para a estimativa da evapotranspiração no contexto da cana-de-açúcar, valendo-se de dados provenientes do sensoriamento remoto, tanto ativo quanto passivo, bem como de diversas fontes de informação. Para atingir tal propósito, a pesquisa foi subdividida em três capítulos distintos, cada um deles direcionado a um objetivo específico.

O Capítulo 1 aborda a estimativa da evapotranspiração atual na cana-de-açúcar utilizando dados do sensor MSI embarcado no Sentinel-2. Solução proposta para sensores que não possuem banda termal.

No Capítulo 2, aborda a estimativa da evapotranspiração atual com os mesmos dados do sensor MSI, além de informações meteorológicas, climatológicas e astronômicas. Nesse caso, não foi necessário utilizar a banda termal ou utilizar dados de uma estação meteorológica em campo.

Por fim, o Capítulo 3 busca solucionar o desafio da presença de nuvens, que impossibilita a utilização de dados de sensores ópticos orbitais. Nesse sentido, algoritmos de *machine learning* são empregados juntamente com dados de radar do Sentinel-1 e informações meteorológicas, climatológicas e astronômicas para estimar a evapotranspiração atual na cana-de-açúcar.

2 Artigos científicos

2.1 Artigo 1

Modelagem da evapotranspiração atual com imagens MSI-Sentinel e algoritmos de *machine learning*

Resumo

Os avanços nos estudos da evapotranspiração das culturas agrícolas por sensoriamento remoto têm-se intensificado nas últimas décadas com a modernização dos recursos computacionais e aplicação de algoritmos de inteligência artificial. Nesse sentido, esta pesquisa propôs a aplicação dos algoritmos de *machine learning* para estimar a ET_{rF} (fração evapotranspirativa) do modelo METRIC (*Mapping Evapotranspiration at High Resolution with Internalized Calibration*) a partir dos dados da constelação de satélites Sentinel-2 para a cultura da cana-de-açúcar. Para atingir tal objetivo, foram adquiridas imagens do sensor MSI (*MultiSpectral Instrument*), do Sentinel-2, e dos sensores OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*), do Landsat-8, que imagearam a lavoura da cana-de-açúcar aproximadamente no mesmo instante entre os anos de 2018 a 2020. As imagens dos sensores OLI e TIR foram destinadas para calcular a ET_{rF} através do METRIC (variável alvo), enquanto das imagens do sensor MSI foram extraídas as variáveis explicativas em duas abordagens, utilizando resolução espacial de 10 m (abordagem-1) e 20 m (abordagem-2). Os resultados mostraram que os algoritmos foram capazes de identificar padrões nos dados do sensor MSI para prever a ET_{rF} do modelo METRIC, e entre as duas abordagens a 2 foi a que teve melhores resultados, devido ao maior número de variáveis preditoras em relação a abordagem-1. Para abordagem-1 os algoritmos que obtiveram melhor predição foram o XgbLinear ($R^2 = 0.80$; RMSE=0.15) e XgbTree ($R^2 = 0.80$; RMSE=0.15), enquanto para abordagem-2 o algoritmo que apresentou superioridade foi o XgbLinear ($R^2 = 0.91$; RMSE=0.10), respectivamente. Assim, fica evidente que os algoritmos de *machine learning* aplicados aos dados sensor MSI foram capazes de estimar a ET_{rF} em um modo mais simplificado do que aquela que envolve balanço de energia com banda termal usado no modelo METRIC.

Palavras-chave: Manejo da irrigação. Cana-de-açúcar. Sensoriamento remoto. Inteligência artificial.

1 Introdução

A evapotranspiração consiste no fenômeno de transferência de água do estado líquido para a atmosfera a partir da evaporação da água do solo e corpos d'água, assim como pela transpiração através das folhas das plantas através dos estômatos (ALLEN, R. G., PEREIRA, L, S., RAES, D., SMITH, 2006). De acordo com Allen *et al.* (2011), quantificar a utilização de água em extensas áreas agrícolas irrigadas é de grande relevância para o planejamento e gestão dos recursos hídricos, mitigação de impactos na vazão dos corpos d'água, estabelecimento de direitos de uso e regulação do consumo, além de evitar conflitos no uso da água.

A evapotranspiração pode ser determinada diretamente em campo por técnicas lisimétricas que são bastantes confiáveis, mas oneroso ou estimada com (1) modelos totalmente físicos baseados aos princípios de conservação de massa e energia; (2) modelos semi-físicos que utilizam a conservação de massa ou energia; e (3) modelos de caixa preta baseados em redes neurais artificiais, relações empíricas, e algoritmos fuzzy, genéticos (SRIVASTAVA *et al.*, 2017) e *machine learning*. O sensoriamento remoto pode ser incluído nos modelos de base física, pois utiliza da radiação refletida e emitida pela superfície da Terra, que quando submetido a modelos de interação da radiação com a superfície, geram estimativas das propriedades da superfície terrestre. A cobertura vegetal, biomassa e índices para cada *pixel* da imagem também podem ser modelados espacialmente e temporalmente (BASTIAANSEN, 1999; TEIXEIRA *et al.*, 2009). Diversos modelos de estimativa de evapotranspiração baseados em sensoriamento remoto foram desenvolvidos e estão sendo amplamente aplicados, principalmente em pesquisas científicas voltadas a área de planejamento hídrico. Dentre estes modelos, destacam-se o *Surface Energy Balance Algorithm for Land* – SEBAL (BASTIAANSEN *et al.*, 1998), *Mapping Evapotranspiration at High Resolution with Internalized Calibration* – METRIC (ALLEN; TASUMI; TREZZA, 2007) e o *Simple Algorithm for Evapotranspiration Retrieving* – SAFER (TEIXEIRA, 2010). Contudo, estes modelos estimam a fração evapotranspirativa que corresponde a um coeficiente que deve ser multiplicado a evapotranspiração de referência para que se consiga a evapotranspiração atual.

O modelo METRIC foi desenvolvido por ALLEN, TASUMI e TREZZA (2007) a fim de estimar a fração evapotranspirativa (ET_rF), dada pela razão entre a evapotranspiração instantânea (ET_{ints}) e a evapotranspiração potencial de referência da alfafa (ET_r) ou grama (ET_o) em escala diária. Este modelo apresenta elevada complexidade pois necessita, *a priori*, estimar os parâmetros que envolvem o balanço de energia, os quais podem ser fonte de vários erros, uma vez que este modelo apresenta um método iterativo entre *pixels* quentes e frios para o cálculo da resistência aerodinâmica ao transporte de calor, que é derivado do modelo SEBAL.

Os modelos supracitados foram desenvolvidos utilizando dados dos sensores da constelação de satélites Landsat, com dependência dos comprimentos de onda no visível e infravermelho médio, próximo e termal. Contudo, os satélites da série Landsat disponibilizam informações da superfície terrestre em resolução temporal de 16 dias e resolução espacial de 30 m. A resolução temporal restringe a aplicação dos modelos, uma vez que apresenta número limitado de imagens disponíveis pelos sensores embarcados nestes satélites. Em contrapartida, a constelação de satélites Sentinel-2 apresenta-se como alternativa à esta limitação dos satélites Landsat. Composta por dois satélites, Sentinel-2A e Sentinel-2B, com sensores de iguais características esta constelação permite uma disponibilidade de imagens a cada 5 dias com resolução espacial variando entre 10, 20 e 60 m, produzindo, deste modo, maior número de imagens da superfície terrestre ao se comparar a constelação Landsat. O sensor que pertence a esta constelação é o MSI (*MultiSpectral Instrument*) que tem a capacidade de adquirir informações da superfície terrestre nos comprimentos de onda do visível, borda do vermelho e infravermelho próximo e médio, contudo não possui sensor na região espectral do infravermelho termal.

Considerando as diferenças entre os satélites mencionadas anteriormente, destaca-se que o desenvolvimento de modelos com capacidade preditiva da ET_a (evapotranspiração atual) da cultura agrícola utilizando sensor MSI será de grande valia para comunidade científica e profissionais de campo, possibilitando a determinação da ET_a com maior resolução espacial e temporal ao longo do ciclo da cultura, e, ainda, pode integrar a outros sensores embarcados em outros satélites formando os multi-sensores, que permitem que, na ausência de informação de um sensor o outro o complementarará similar ao descrito por Filgueiras *et al.* (2020). Outro ponto importante é que o MSI não apresenta o infravermelho termal, portanto a ET_a pode ser estimada sem informação da temperatura de superfície o qual pode reduzir erros do operador, pois modelos que envolve informações termais tem em sua estrutura o complexo balanço de energia para quantificar a energia latente do sistema, somado a isso as bandas termais possuem menor resolução espacial, sendo necessário, por muitas vezes, a interpolação para compatibilizar o tamanho dos *pixels* com aqueles de outras bandas, a exemplo do Landsat 5,7, 8 e 9. Por outro lado, modelar a evapotranspiração na ausência de dados da banda termal pode gerar resultados insatisfatórios devido a relação intrínseca entre temperatura de dossel da vegetação, derivado da banda termal, com a condutância estomática (BERNI *et al.*, 2009; TAIZ *et al.*, 2015).

Almejando o desenvolvimento de modelos com elevada capacidade preditiva e confiabilidade, busca-se através da aplicação dos algoritmos de *machine learning* modelar variáveis de interesse. Estes algoritmos, baseados em inteligência artificial, apresentam

robustez em sua estrutura que permitem identificar padrão de relação entre as variáveis que se deseja modelar, com as variáveis ditas preditoras (independentes). Segundo Cervantes *et al.* (2020) o *machine learning* é uma área interdisciplinar fundamentada na ciências da computação, estatísticas, matemática, otimização entre diversas outras áreas. Diversos algoritmos de *machine learning*, cada um com características diferentes, são utilizados para predição e classificação, sendo, nos últimos anos, aplicado em pesquisa voltado às ciências agrárias e sensoriamento remoto com intuito de desenvolver modelos com maior capacidade de representar o fenômeno (ADAB *et al.*, 2020; FILGUEIRAS *et al.*, 2019b; GRANATA, 2019; TIKHAMARINE *et al.*, 2019; VIRNODKAR *et al.*, 2020). Na predição da evapotranspiração esses modelos foram utilizados em dados de sensoriamento remoto e dados climáticos para estimar a evapotranspiração atual (CARTER; LIANG, 2019; XU *et al.*, 2019), bem como em apenas dados climáticos para estimar a evapotranspiração de referência (CHAI *et al.*, 2018; KADKHODAZADEH *et al.*, 2022; RAZA *et al.*, 2020; TANG *et al.*, 2018). Os diversos estudos disponíveis aplicam os algoritmos na predição da evapotranspiração de referência ou na evapotranspiração atual utilizando sensoriamento remoto com dados termiais. Desta forma, buscou-se no presente trabalho treinar algoritmos de *machine learning* para estimar a fração evapotranspirativa do modelo METRIC a partir dos dados da constelação de satélites Sentinel-2 que não utiliza dados termiais para a cultura da cana-de-açúcar, usada como estudo de caso.

2 Materiais e Métodos

2.1 Área de estudo

O estudo foi realizado na região norte do estado de Minas Gerais, Brasil, classificada com o clima AS- com seca no verão (ALVARES *et al.*, 2013) (Figura 1). O cultivo de cana-de-açúcar foi estabelecido sob 150 equipamentos de irrigação por sistema de pivô central.

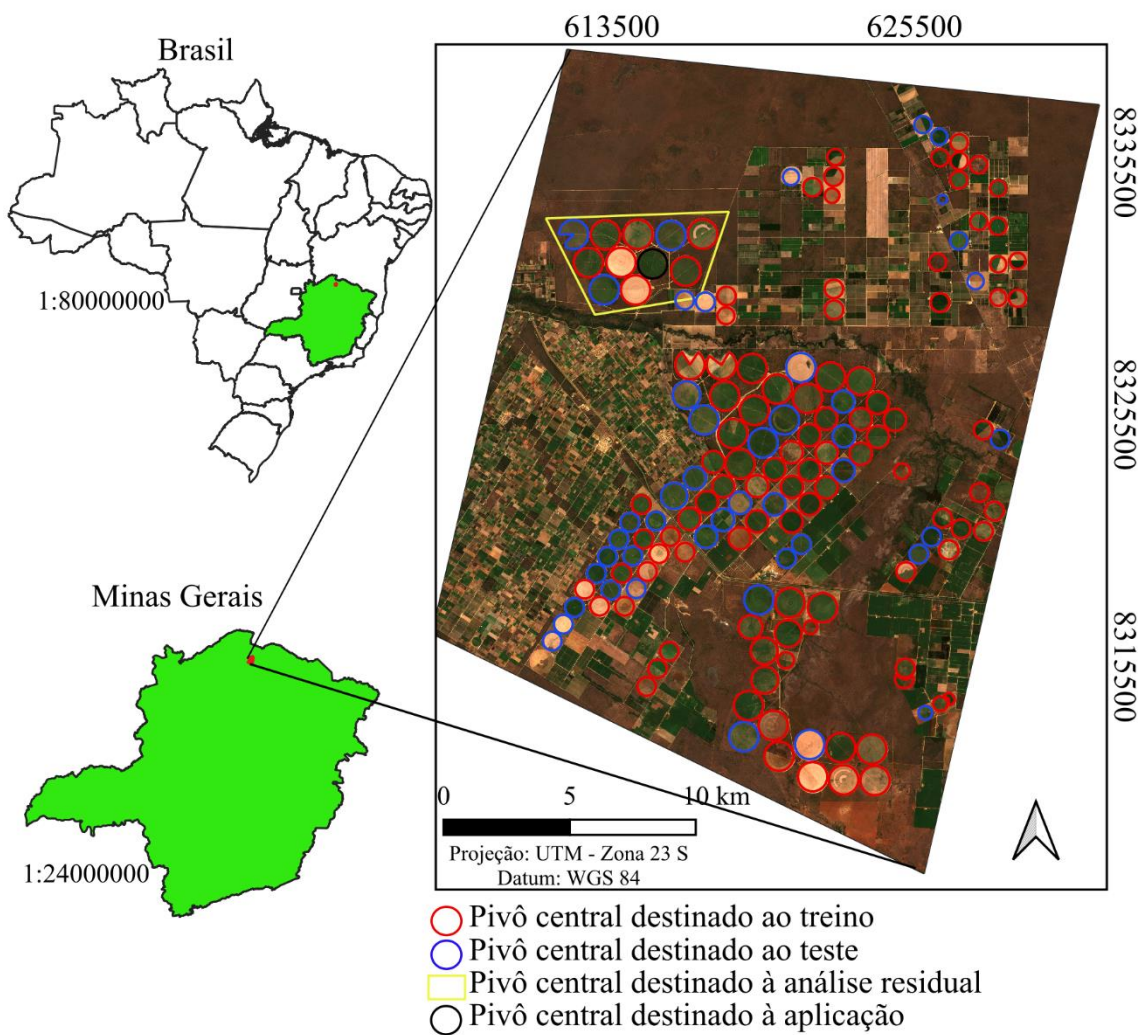


Figura 1. Localização da área estudo destacando os pivôs centrais utilizados para treino, teste e aplicação dos modelos.

Esta área foi escolhida por ser uma área de parceria de pesquisa com pouca ou nenhuma precipitação (Figura 2) que favorece a aquisição de um maior número de imagens livres de nuvens, podendo assim adquirir um volume maior de informações espectrais da área, principalmente coincidentes com Landsat-8 e Sentinel-2.

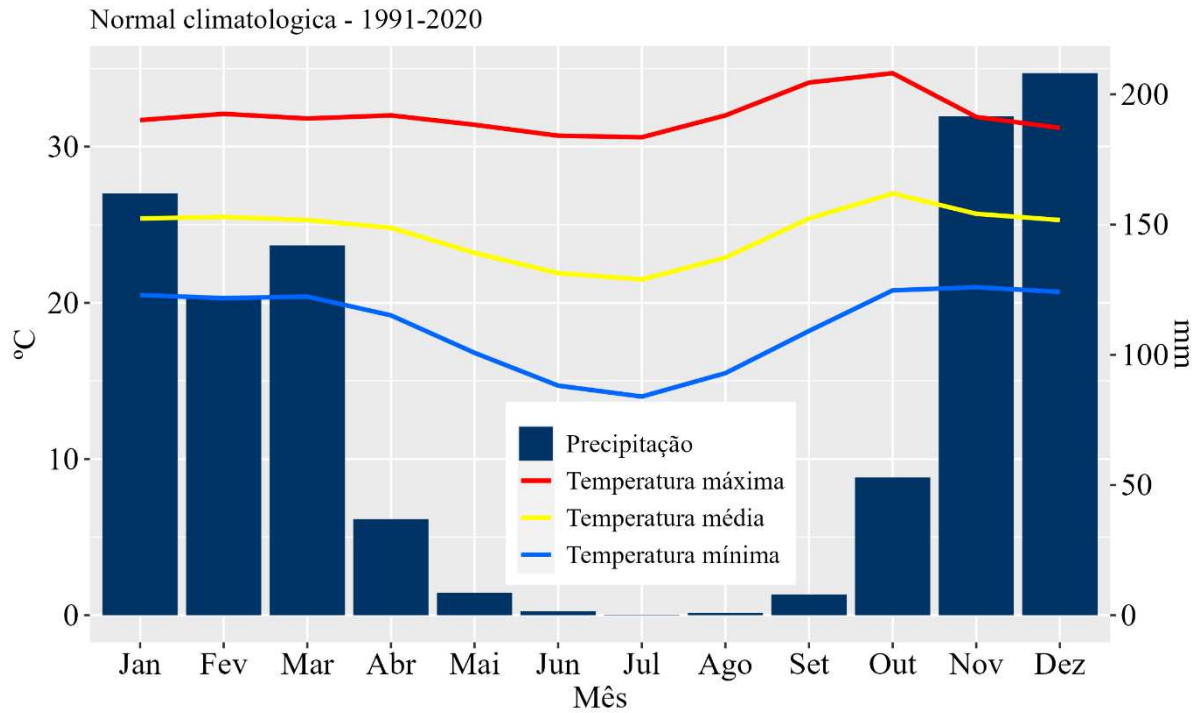


Figura 2. Normal climatológica da região de estudo extraída da estação A539 do INMET (Instituto Nacional de Meteorologia).

2.2 Dados Landsat-8 e Sentinel-2

Os dados dos sensores a bordo dos satélites Landsat-8 e Sentinel-2 foram adquiridos para gerar a variável observada através do modelo METRIC, e as variáveis explicativas utilizadas para treinar e testar os algoritmos de *machine learning* e podem ser visualizados de forma simplificada no fluxograma na Figura 3.

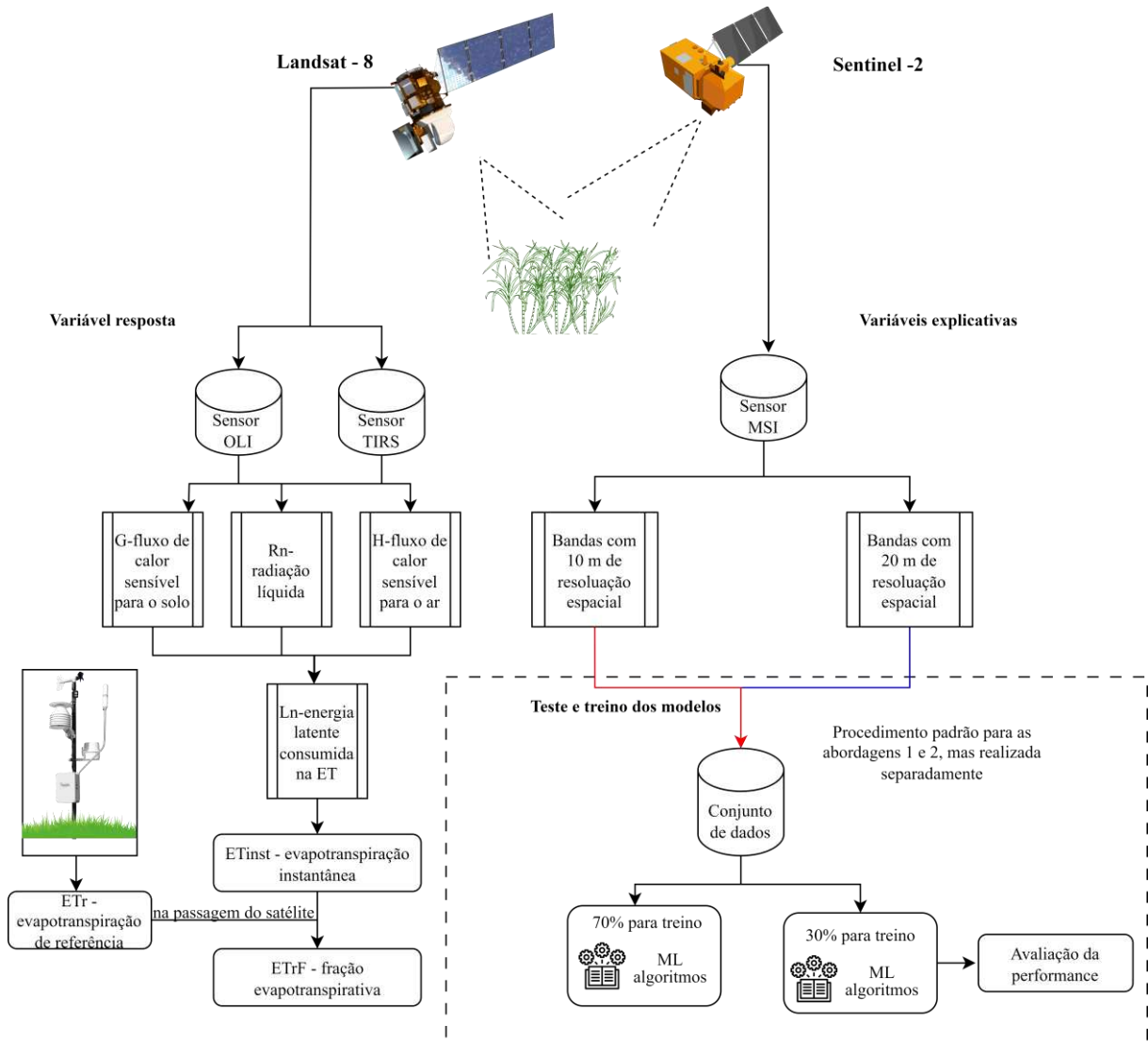


Figura 3. Aquisição de dados de sensores a bordo do Landsat-8 e Sentinel-2 para treinamento e teste em algoritmos de *machine learning* (ML).

Foram adquiridos, no total, 56 imagens de satélites, dos quais 22 foram provenientes do sensor OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*) do satélite Landsat-8 e 34 do sensor MSI (*Multispectral Instrument*) a bordo dos satélites Sentinel-2A e 2B. As imagens do sensor OLI e TIRS foram necessárias para cômputo da ET_r através do modelo METRIC, sendo está a variável resposta, ao passo que os dados do sensor MSI foram utilizados como variáveis preditoras nos modelos de *machine learning*. As características espectrais e espaciais dos sensores utilizados estão disponíveis na Tabela 1 e 2.

Tabela 1. Características espectrais e espacial do sensor OLI e TIRS

	Banda espectral	Comprimento de onda (μm)	Resolução espacial (m)
B1	<i>Coastal aerosol (Ca)</i>	0.43 – 0.45	
B2	Azul (B)	0.45 – 0.51	
B3	Verde (G)	0.53 – 0.59	
B4	Vermelho (R)	0.64 – 0.67	
B5	Infravermelho próximo (NIR)	0.85 – 0.88	
B6	Infravermelho de ondas curtas1 (SWIR1)	1.57 – 1.65	30
B7	Infravermelho de ondas curtas2 (SWIR2)	2.11 – 2.29	
B8	<i>Panchromatic (PCh)</i>	0.50 – 0.68	
B9	<i>Cirrus (C)</i>	1.36 – 1.38	
TIRS			
B10	Infravermelho termal1 (TIRS1)	10.60 – 11.19	100*
B11	Infravermelho termal2 (TIRS2)	11.50 – 12.51	

*A resolução do sensor é de 100 m, mas as imagens são reamostradas e disponibilizada em 30 m

Tabela 2. Características espectrais e espacial do sensor MSI

	Banda espectral	Comprimento de onda (μm)	Resolução espacial (m)
B2	Azul (B)	0.459 - 0.525	
B3	Verde (G)	0.542 - 0.578	10
B4	Vermelho (R)	0.650 - 0.680	
B8	Infravermelho próximo (NIR)	0.781 - 0.887	
B5	Borda do vermelho 1 (Re1)	0.697 - 0.712	
B6	Borda do vermelho 2 (Re2)	0.733 - 0.748	
B7	Borda do vermelho 3 (Re2)	0.773 - 0.793	
B8A	Infravermelho próximo (NIR)	0.856 - 0.876	20
B11	Infravermelho de ondas curtas1 (SWIR1)	1.569 - 1.660	
B12	Infravermelho de ondas curtas2 (SWIR2)	2.115 - 2.290	
B1	<i>Coastal aerosol</i>	0.433 - 0.453	
B9	<i>Water vapor</i>	0.935 - 0.955	60
B10	<i>Cirrus</i>	1.359 - 1.390	

As imagens foram adquiridas entre os anos de 2018 e 2020. Durante esse período, oito imagens provenientes do satélite Landsat-8 e oito provenientes do sensor MSI imagearam a área de estudo exatamente no mesmo dia. As imagens dessas datas foram destinadas ao treinamento e teste dos modelos e aplicação das análises dos resíduos. Optou-se pelas imagens com datas coincidentes entre os satélites para minimizar reflectância diferentes da mesma área de estudo, haja vista que vegetação, principalmente culturas agrícolas, apresentam rápida mudança ao longo do tempo. Assim, adquirir imagens em dias diferentes para realizar a

modelagem poderia ter vícios e não representar fidedignamente o produto modelado. Na Tabela 3 estão disponíveis as imagens com data, hora (GMT-3) e Path/Row utilizada para o treinamento dos modelos.

Tabela 3. Características das oito cenas selecionadas para desenvolvimento dos modelos.

Data (dd/mm/aaaa)	Landsat-8		Sentinel-2	
	hora (hh:mm:ss)	Path/Row	hora (hh:mm:ss)	Tile Number
Treinamento				
06/07/2018	09:55:17.157	218/71	10:12:41.024	T23LPD
22/05/2019	09:55:49.714	218/71	10:12:51.024	T23LPD
04/10/2019	10:02:19.874	219/70	10:12:49.024	T23LPD
29/10/2019	09:56:34.656	218/71	10:12:49.024	T23LPD
17/01/2020	09:56:20.969	218/71	10:12:41.024	T23LPD
31/05/2020	10:01:27.427	219/70	10:12:49.024	T23LPD
19/08/2020	10:02:00.995	219/70	10:12:49.024	T23LPD
Análise dos resíduos				
12/02/2020	09:56:32.117	218/71	10:12:41.024	T23LPD

Entre 23/11/2019 até 04/09/2020 foram adquiridas 17 imagens do sensor OLI e TIRS (incluindo 4 imagens da Tabela 3) e 30 do sensor MSI (incluindo 4 imagens da Tabela 3). Estas imagens foram utilizadas para determinar o K_c (coeficiente de cultura) e ET_a ao longo do ciclo da cana-de-açúcar. Essa estimativa foi realizada tanto pelo modelo METRIC quanto pelo modelo proposto.

2.3 Variável resposta

A variável resposta, ET_rF , foi obtida em cada *pixel* com resolução espacial de 30 x 30 m a partir do modelo METRIC. O cômputo da ET_rF se baseia na razão entre a ET_{inst} (evapotranspiração instantânea) em cada *pixel* pela ET_r (evapotranspiração de referência alfafa) horária, calculada pela estação meteorológica. Allen *et al.* (2007) padronizaram a ET_r para a alfafa com altura de 0.5 m e, segundo os autores, ao se utilizar essa condição a ET_rF pode ser considerada equivalente ao K_c , além de permitir a extrapolação da evapotranspiração atual da cultura durante o momento da passagem do satélite para o nível diário de 24 h. Assim, a ET_rF é determinada a partir da Eq. (1).

$$ET_rF = \frac{ET_{inst}}{ET_r} \quad (1)$$

onde ET_{inst} é a evapotranspiração instantânea ($mm.h^{-1}$) e ET_r é evapotranspiração de referência ($mm.h^{-1}$) padronizada para alfafa a 0.5 m de altura no momento da passagem do satélite.

A ET_{inst} é calculada a partir da energia latente consumida no processo de evapotranspiração (ET) e calor latente de vaporização Eq. (2)

$$ET_{\text{inst}} = 3600 \frac{LE}{\lambda \rho_w} \quad (2)$$

onde 3600 é a conversão de segundos, intervalo de tempo da passagem do satélite, para hora; LE é a energia latente consumida na ET ($W m^{-2}$); ρ_w é densidade da água ($\sim 1000 \text{ kg m}^{-3}$); e λ é o calor latente de vaporização ($J \text{ kg}^{-1}$).

O λ representa o calor absorvido quando um kg de água é evaporada e foi calculado pela Eq. (3).

$$\lambda = [2.501 - 0.00236(T_s - 273.15)] \times 10^6 \quad (3)$$

onde T_s é a temperatura de superfície ($^{\circ}K$) determinada pela banda 10 do sensor TIRS (Tabela 1).

A LE, por sua vez, é calculada a partir do balanço de energia da superfície terrestre, Eq. (4), que envolve a radiação líquida (R_n), fluxo sensível do calor transferido para o solo (G) e fluxo sensível do calor convectado para o ar (H). Esses três componentes responsáveis por determinar a LE são expressos, geralmente, em $W m^{-2}$.

$$LE = R_n - G - H \quad (4)$$

A R_n é a energia radiante da superfície que é particionada em H , G e LE, e é determinada pela Eq. (5).

$$R_n = R_{s\downarrow} - \alpha R_{s\downarrow} + R_{L\downarrow} - R_{L\uparrow} - (1 - \epsilon_0) R_{L\downarrow} \quad (5)$$

onde $R_{s\downarrow}$ é a entrada da radiação de ondas curtas ($W m^{-2}$); α é o albedo de superfície (adim.), determinado pelas bandas 2, 3, 4, 5, 6 e 7 do Landsat-8 (Tabela 1); $R_{L\downarrow}$ e $R_{L\uparrow}$ é a entrada e saída de ondas longas ($W m^{-2}$), respectivamente; e ϵ_0 é a emissividade térmica da superfície.

G é o fluxo de armazenamento de calor no solo devido a condução do calor. Quando há presença de vegetação o G tende a ter menores valores, ou seja, menor é a taxa de armazenamento de calor no solo. O modelo METRIC disponibiliza duas metodologias para quantificar G , a desenvolvida por Bastiaanssen (1999) e por Tasumi (2003), detalhes podem ser vistos em Allen *et al.* (2007). Para este trabalho optou-se pela metodologia de Tasumi (2003), pois foi desenvolvido em áreas irrigadas. Assim, a G foi quantificado pela Eq. (6a e 6b).

$$\frac{G}{R_n} = 0.05 + 0.18e^{-0.521 \text{ LAI}} \quad \text{para LAI} \geq 0.5 \quad (6a)$$

$$\frac{G}{R_n} = \frac{1.80(T_s - 273.15)}{R_n + 0.084} \quad \text{para LAI} < 0.5 \quad (6b)$$

onde LAI é o índice da área foliar ($\text{m}^2 \text{m}^{-2}$), o qual foi estimado pela metodologia aplicada por Allen *et al.* (2011).

Por fim, o H, que representa a taxa de perda de calor para o ar por convecção e condução influenciado por uma diferença de temperatura, foi calculado pela Eq. (7).

$$H = \rho_{\text{air}} C_p \frac{dT}{r_{ah}} \quad (7)$$

onde ρ_{air} é a densidade do ar (kg m^{-3}); C_p é o calor específico do ar a pressão constante ($\text{J kg}^{-1} \text{K}^{-1}$); dT é a diferença de temperatura entre duas alturas ($z_1 - 0.1 \text{ m}$ e $z_2 - 2 \text{ m}$) em uma área próxima à superfície do solo; e r_{ah} é a resistência aerodinâmica (s m^{-1}) entre essas duas alturas.

Todo o cálculo até chegar na ET_iF descrito nesse tópico foi realizado utilizando o pacote *Water*, desenvolvido por Olmedo *et al.* (2018) para ambiente de linguagem R (R CORE TEAM, 2020).

2.4 Extração dos dados para treinamento

As imagens do sensor MSI foram separadas em dois grupos de acordo com sua resolução espacial. A partir desta distinção, duas abordagens foram aplicadas: Abordagem-1 utilizou-se resolução espacial de 10 m e a abordagem-2 resolução espacial de 20 m. Essa divisão teve por objetivo conhecer o efeito do número de variáveis preditoras e da resolução espacial no desempenho dos modelos preditivos.

As variáveis primárias selecionadas para a abordagem-1 foram reflectância das bandas azul (ρ_B), verde (ρ_G), vermelho (ρ_R) e infravermelho próximo (ρ_{NIR}), enquanto a abordagem-2 foram reflectância das bandas azul (ρ_B), verde (ρ_G), vermelho (ρ_R), borda do vermelho 1 (ρ_{Re1}), borda do vermelho 2 (ρ_{Re2}), borda do vermelho 3 (ρ_{Re3}), infravermelho próximo (ρ_{NIR}), infravermelho de ondas curtas1 (ρ_{SWIR1}) e infravermelho de ondas curtas2 (ρ_{SWIR2}). Devido as abordagens-1 e 2, possuem resolução espacial de 10 e 20 m, respectivamente, foi necessário compatibilizar suas dimensões com a variável dependente, ou seja, elevar seus valores para resolução de 30 x 30 m. Para isso, utilizou-se a função *resample* com o método bilinear do pacote raster (HIJMANS *et al.*, 2020) para linguagem R, tendo como base as imagens da variável dependente. Com as imagens em igual resolução espacial, procedeu-se com recorte dos campos irrigados com sistemas de pivôs centrais dentro de cada cena utilizando arquivo tipo *shapfile* e em seguida foi realizado separação dos pivôs em dois grupos, em que um grupo possuía 70% do total dos pivôs, o qual foi destinado ao treino do

modelo, e o outro com 30% que foi destinado ao teste do modelo após o treino (Figura 1), assim produziu-se dois grupos de imagens, uma com 70% dos pivôs e o outro contendo 30%.

Após redimensionamento dos *pixels* e separação dos pivôs, extraiu-se as informações espectrais referente a cada banda contida na abordagem-1 e 2, assim como a ET_rF tanto para o grupo de imagens que continha 70% dos pivôs quanto para aquele com 30%. Em função da limitação computacional não foi possível utilizar todos os *pixels* dos pivôs, assim foram extraídos, aleatoriamente, para o conjunto de 70% o equivalente a 25.000 informações totalizando 175.000 dados, lembrando que foram utilizadas 7 imagens (Tabela 3). Já para o grupo com 30% extraiu-se, também aleatoriamente, 10.714 dados de cada imagem, totalizando 74.998 *pixels*, o qual mantém uma proporção de 30% dos dados destinados ao treinamento. Posteriormente, foram filtrados valores de ET_rF menores que 0 que eventualmente continha em alguns *pixels* gerado pelo ruído em algumas imagens. Ao fim, obteve-se arquivos do tipo *data frame* contendo coordenadas geográficas, bandas espectrais e ET_rF .

2.5 Treinamento e avaliação estatística dos modelos

2.5.1 Produção e seleção das variáveis preditoras

Com intuito de aumentar o número de variáveis preditoras, aplicou-se a técnica NRPB (*normalized ratio procedure between bands*), sobre as variáveis preditoras primárias tanto para abordagem-1, quanto para abordagem-2. A NRPB realizou a normalização entre todas as variáveis primárias, aumentando dessa forma o número variáveis explicativas para os modelos (Eq. 8), com isso o número de variáveis produzidas em cada abordagem é superior as primárias, principalmente na abordagem-2 que possui maior número de variáveis.

$$NRPB = \frac{\rho_x - \rho_y}{\rho_x + \rho_y} \quad (8)$$

onde ρ_x e ρ_y correspondem as reflectâncias de superfície dos comprimentos de ondas do sensor MSI.

O processo de NRPB foi realizado utilizando a função *band_ratio* do pacote *labgeo* (Fernandes Filho 2019) para linguagem R, sendo aplicado com sucesso por (FILGUEIRAS *et al.*, 2019b, 2019a).

Dentro das variáveis produzidas pela NRPB realizou-se a seleção, em duas etapas, das variáveis de maior importância para os modelos. A primeira foi a remoção de variáveis com autocorrelação, assim variáveis explicativas que apresentaram correlação acima de 95% entre si foram removidas diminuindo, desta forma, a redundância de informações. A segunda etapa foi a remoção de variável de menor importância e para tal aplicou-se o algoritmo RFE (*recursive*

feature elimination) através do pacote *caret* (KUHNS *et al.*, 2020) em linguagem R. Tal algoritmo remove as variáveis preditoras de menor importância para o modelo, e para isso ele constrói um modelo base com todas as preditoras e em seguida calcula a importância de cada preditora para esse modelo base realizando, posteriormente, a classificação dessas preditoras e removendo-as aquelas de menor importância, restando assim, menor número de variáveis para as etapas de treinamento. Este processo é fundamental pois reduz variáveis desnecessárias aos modelos, facilitando tanto o treinamento quanto a aplicação futura.

2.5.1 Treinamento e teste

O treinamento dos modelos de *machine learning* consistiu na inserção das variáveis preditoras, que restaram após os processos de seleção das variáveis, em quatro algoritmos de regressão, sendo eles: *Linear regression* (LM); *Cubist*; *eXtreme Gradient Boosting- método linear* (XgbLinear); e *eXtreme Gradient Boosting- método tree* (XgbTree). Esses modelos foram escolhidos por apresentarem elevado potencial preditivo e serem mais rápidos no treinamento, o qual foi realizado com auxílio do pacote *caret* (KUHNS *et al.*, 2020). Finalizado o treinamento, obteve-se 8 previsões possíveis, sendo 4 para abordagem-1 e 4 para abordagem-2, os quais foram submetidos ao teste para avaliação da assertividade e erro estatístico.

O teste dos algoritmos treinados foi realizado com os dados extraídos dos 30% dos pivôs destinados para essa finalidade. A análise estatística para o teste foi elaborada através do coeficiente de determinação (R^2), Eq. (9), raiz do erro quadrático médio (RMSE), Eq. (10), erro médio absoluto (MAE), Eq. (11), e erro médio de bias (MBE), Eq. (12).

$$R^2 = \frac{[\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})]^2}{[\sum_{i=1}^n (P_i - \bar{P})^2][\sum_{i=1}^n (O_i - \bar{O})^2]} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (11)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (12)$$

onde P_i é o valor previsto pelo modelo; O_i é o valor observado; \bar{O} é o valor médio observado e n é o número de pares de dado.

2.5.3 Análise residual

Além das métricas estatísticas, realizou-se a análise residual para avaliar o comportamento do erro dos valores preditos para todos os modelos treinados para as duas abordagens. Para isso, destinou-se exclusivamente uma imagem para esta finalidade (Tabela 3),

nesta imagem e dentro da área de estudo (Figura 1), 11 pivôs foram selecionados (destaque em amarelo) para análise dos resíduos entre valores observado e predito. A escolha dos 11 se deu por estarem próximo, o que facilita a elaboração de mapas temáticos, e por não haver presença de nuvem próximos a eles. É importante mencionar que a imagem destinada para análise residual não esteve envolvida no processo de treinamento, assim, os dados desta imagem são informações novas para os modelos já treinados. Desta forma, aplicou-se os modelos, posteriormente extraiu-se a ET_{rF} predita (abordagem-1 e 2) e a ET_{rF} observada (METRIC) dos 11 pivôs (Figura 1) para o cômputo dos resíduos (Eq. 13).

$$\text{Residual} = \text{Observed value} - \text{Predicted value} \quad (13)$$

2.6 Aplicação

Dentre os pivôs selecionados (Figura 1), um pivô (destacado em preto) foi utilizado para aplicação dos modelos treinados. A escolha deste pivô é devida maior número de imagens Sentinel-2 durante o ciclo da cultura. A aplicação consistiu em quantificar o K_c e ET_a ao longo do ciclo da cultura da cana-de-açúcar para o melhor modelo, e para o modelo METRIC. Além do K_c também se estimou o consumo hídrico total durante o ciclo da cultura da cana-de-açúcar. Para isso, utilizou-se das 29 imagens do sensor MSI desde 23/11/2019 (DAE-dias após emergência 05) até 03/09/2020 (DAE 290) e 17 imagens do sensor OLI e TIRS entre 07/12/2019 (DAE 19) até 04/09/2020 (DAE 291).

Nas 29 imagens do sensor MSI aplicou-se o modelo que obteve melhores resultados estatísticos nas abordagens 1 e 2, sendo possível a determinação da ET_{rF} . De acordo com Allen *et al.* (2007) a ET_{rF} é equivalente ao K_c quando se utiliza a ET_r da alfafa, situação em que a ET_{rF} foi modelada. Já as 17 imagens do sensor OLI e TIRS foram utilizadas para computar o K_c através do modelo METRIC. Após quantificar o K_c , estabeleceu-se a evapotranspiração atual da cultura. Para tal, calculou-se a evapotranspiração de referência acumulada em um período de 24 h (ET_{r-24}) no dia em que cada imagem foi adquirida.

A ET_{r-24} , em cada data, foi estimada utilizando a equação Penman-Monteith padronizado pela ASCE (*American Society of Civil Engineers*) (ALLEN *et al.*, 2005) e os dados meteorológicos foram adquiridos na estação A539 – Mocambinho que pertence ao INMET (Instituto Nacional de Meteorologia). De posse de todas ET_{r-24} e do K_c , estabeleceu-se a ET_a em cada *pixel* pela Eq. (14).

$$ET_a = K_c \times ET_{r-24} \quad (14)$$

Por fim, determinou-se evapotranspiração total da cultura da cana durante o ciclo pela abordagem-1 e 2 e pelo METRIC aplicando função integral ao longo do tempo, em que x era as datas referente as imagens e y a ET_a de cada data. Assim, calculou-se a área abaixo dessa curva que correspondeu a evapotranspiração total durante o ciclo. Este processo foi realizado utilizando a função `auc` do pacote `MESS` (EKSTRØM, 2020) para linguagem R.

3 Resultados e Discussão

3.1 Modelos

Na abordagem-1 foram gerados 11 preditores ao aplicar a normalização entre as bandas espectrais de resolução espacial de 10 m, desse total restaram apenas 7 ao remover preditores com correlação acima de 95% e 5 ao utilizar o RFE (*recursive feature elimination*). Na Figura 4 são apresentados os resultados após a aplicação do algoritmo RFE, evidenciando que, potencialmente, o número de variáveis predictoras ideais são 5, acima disso pode ser removido, pois apresenta baixa importância ao modelo e cuja exclusão terá o menor efeito sobre os erros nos modelos treinados, porém número de preditores menores que 5 afetam negativamente a predição do modelo.

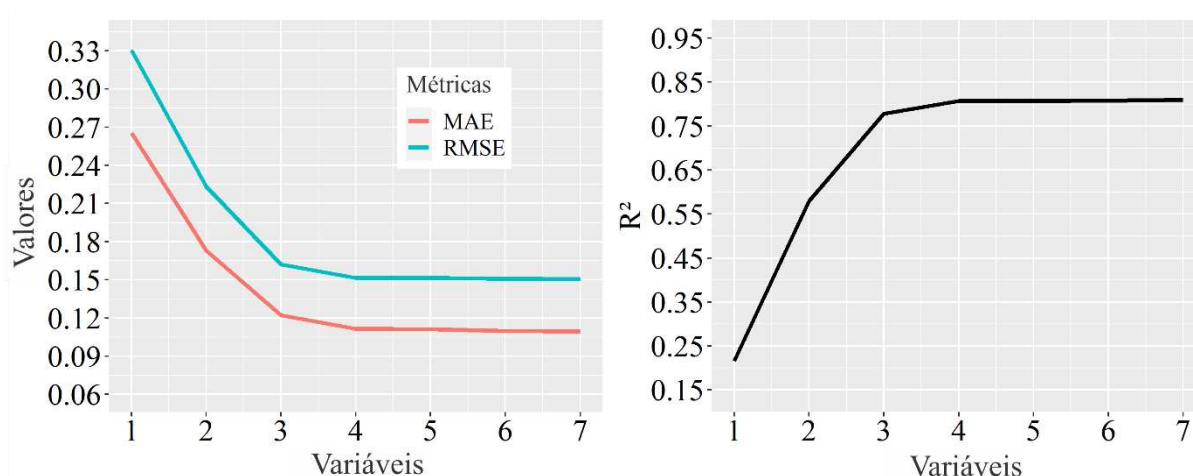


Figura 4. Seleção das variáveis predictoras ao aplicar o RFE na abordagem-1.

Na abordagem-2 o número de preditores gerados pela normalização entre bandas foram iguais a 45, superior a abordagem-1 devido ao maior número de bandas espectrais para a resolução espacial de 20 m. Estabelecendo o patamar de 95% de correlação, restaram 22 variáveis das quais foram reduzidas a 12 após ser aplicando o RFE. Na Figura 5 pode ser visualizado a estabilidade das métricas estatísticas utilizando 12 variáveis predictoras para treinar o modelo.

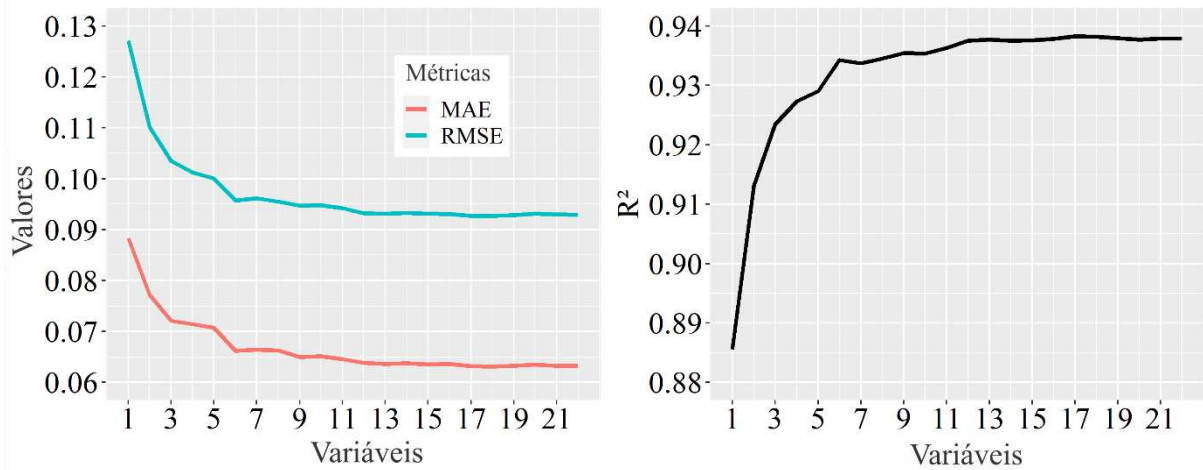


Figura 5. Seleção das variáveis predictoras ao aplicar o RFE na abordagem-2.

Na Tabela 4 é observado as 5 variáveis predictoras que foram selecionadas para abordagem-1 e as 12 selecionadas para abordagem-2. É importante mencionar que as predictoras desta tabela não estão ranqueadas por importância, pois cada algoritmo atribui importância diferente após o treinamento.

Tabela 4. Variáveis predictoras selecionadas para treinamento dos modelos para abordagem-1 e 2.

Nº	Abordagem-1	Abordagem-2
1	B2	B6
2	B4	B8
3	B8	B12
4	(B2-B3) / (B2+B3)	(B2-B3) / (B2+B3)
5	(B2-B4) / (B2+B4)	(B2-B4) / (B2+B4)
6	-	(B2-B5) / (B2+B5)
7	-	(B2-B12) / (B2+B12)
8	-	(B5-B11) / (B5+B11)
9	-	(B5-B12) / (B5+B12)
10	-	(B6-B8) / (B6+B8)
11	-	(B8-B12) / (B8+B12)
12	-	(B11-B12) / (B11+B12)

Para a abordagem-1 os comprimentos de onda que mais aparecem entre as predictoras foram a B2 (azul) e B4 (vermelho) sendo ambas 2 vezes presente contabilizando individualmente na NRPB. Os dois comprimentos de onda que se destacaram nesta abordagem são absorvidos em maior intensidade pelas plantas a partir dos pigmentos fotossintetizantes (TAIZ *et al.*, 2015). Ainda de acordo com esses autores, estes comprimentos de onda estão relacionados com o espectro de ação da liberação de oxigênio a partir da fotossíntese, ou seja, são neles que ocorrem a magnitude da liberação desse elemento. Assim, quanto maior absorção dos comprimentos de ondas azul e vermelho maior liberação de oxigênio e por consequência maior liberação de vapor de água, uma vez que o vapor de água é liberado juntamente com o

oxigênio durante a fotossíntese, daí o possível motivo dessas bandas espectrais possuírem destaques. Os comprimentos de onda B2, B3 (verde) e B4 correspondem aos pigmentos das folhas, enquanto a B8 (infravermelho próximo) tem relação com a parte estrutural, assim, nota-se que as preditoras selecionadas tem importância tanto para o processo fotossíntese quanto a estrutura dos componentes celulares das plantas estudadas, sendo assim, 3 comprimentos com características relacionado aos pigmentos fotossintetizantes e 1 relacionado a estrutura. Na abordagem-2 as bandas que se destacaram foram a B12 (infravermelho de ondas curtas) aparecendo 5 vezes e a B2 aparecendo 4 vezes, a primeira é o comprimento de onda curtas que possui relação com conteúdo de água o qual é objeto de estudo neste trabalho a partir do processo físico-biológico da evapotranspiração. Tal fato pode ser notado em curvas espectrais geradas em folhas verdes da cana-de-açúcar (SANTOS; DEMATTÊ; SILVERO, 2020). Além destes, destaca-se também as bandas B5 e B6 que são bordas do vermelho, que representam uma transição entre a região visível e o infravermelho próximo, a B8 e a B11 que é outro infravermelho de ondas curtas. Assim, nota-se que nesta abordagem as preditoras selecionadas tem relação não apenas com a parte fotossintética e estrutural, mas também com conteúdo de água. Com isso, pode ser inferir que foram atribuídos 3 comprimentos de ondas com características fotossintéticas, 1 estrutural, 2 que se encontra na transição do fotossintético para o estrutural e 2 com conteúdo de água.

O desempenho do teste da abordagem-1 para cada modelo de *machine learning* pode ser visualizado na Figura 6.

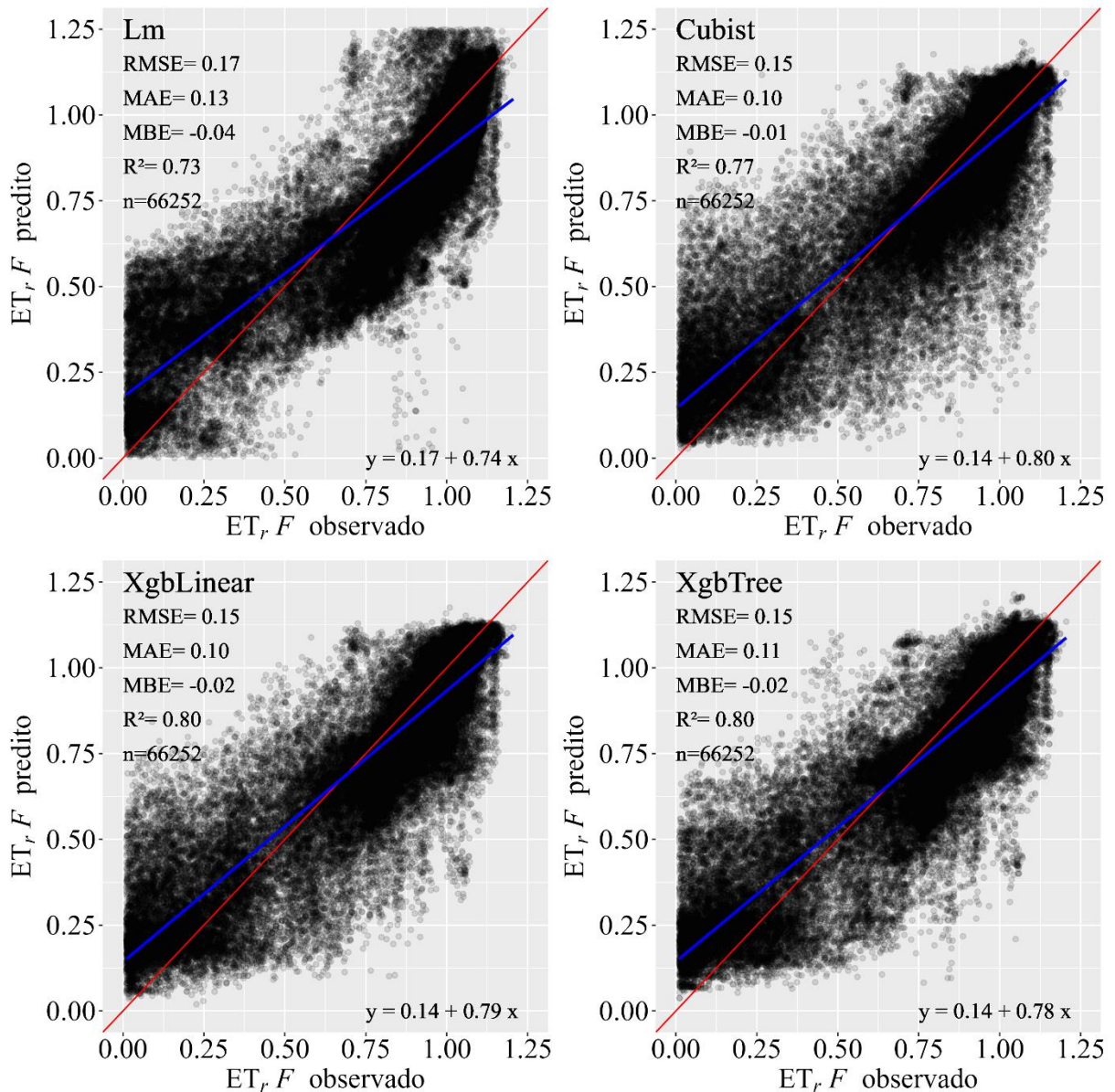


Figura 6. Comparação entre $ET_r F$ observada e predita com as respectivas estatísticas para o teste da abordagem-1

Entre os 4 modelos escolhidos, nesta abordagem, o XgbLinear e o XgbTree obtiveram os melhores resultados estatísticos, com R^2 igual 0.80 e RMSE de 0.15, seguido do Cubist com R^2 igual 0.77 e RMSE de 0.15 e por último com pior resultado a regressão linear múltipla (Lm) com o menor R^2 (0.73) e o maior RMSE (0.17). Independente do modelo, chama atenção a elevada dispersão apresentada no teste, principalmente na porção de valores inferiores a 0.6 (Figura 6). Essa dispersão pode estar relacionada ao número de variáveis predictoras, uma vez que foram utilizadas 5 variáveis para explicar o fenômeno. Associado a isso, valores de $ET_r F$ menores que 0.6, correspondem a baixa cobertura vegetal, estando associados a presença de solo exposto, o que pode ter induzido à elevada dispersão para valores baixos de $ET_r F$. (KE *et al.*, 2017) ao estimar a evapotranspiração utilizando *machine learning* em dados do Landsat-8

e MODIS para ambientes heterogêneos, constataram que área com culturas agrícolas houve maior dispersão entre valores de evapotranspiração predito vs evapotranspiração observado quando comparado com áreas de floresta, pastagem e arbustos. Assim, os resultados destes autores vão de encontro com os resultados aqui encontrado, pois áreas destinadas a culturas agrícolas há maior movimentação da superfície, sendo, desta forma mais dinâmico, tanto no que se refere a cobertura vegetal, quanto no que se relaciona a umidade do solo.

Na Figura 7 é possível observar os resultados estatísticos do teste para a abordagem-2, aquela que usou apenas resolução espacial de 20 m como fonte de variáveis preditoras.

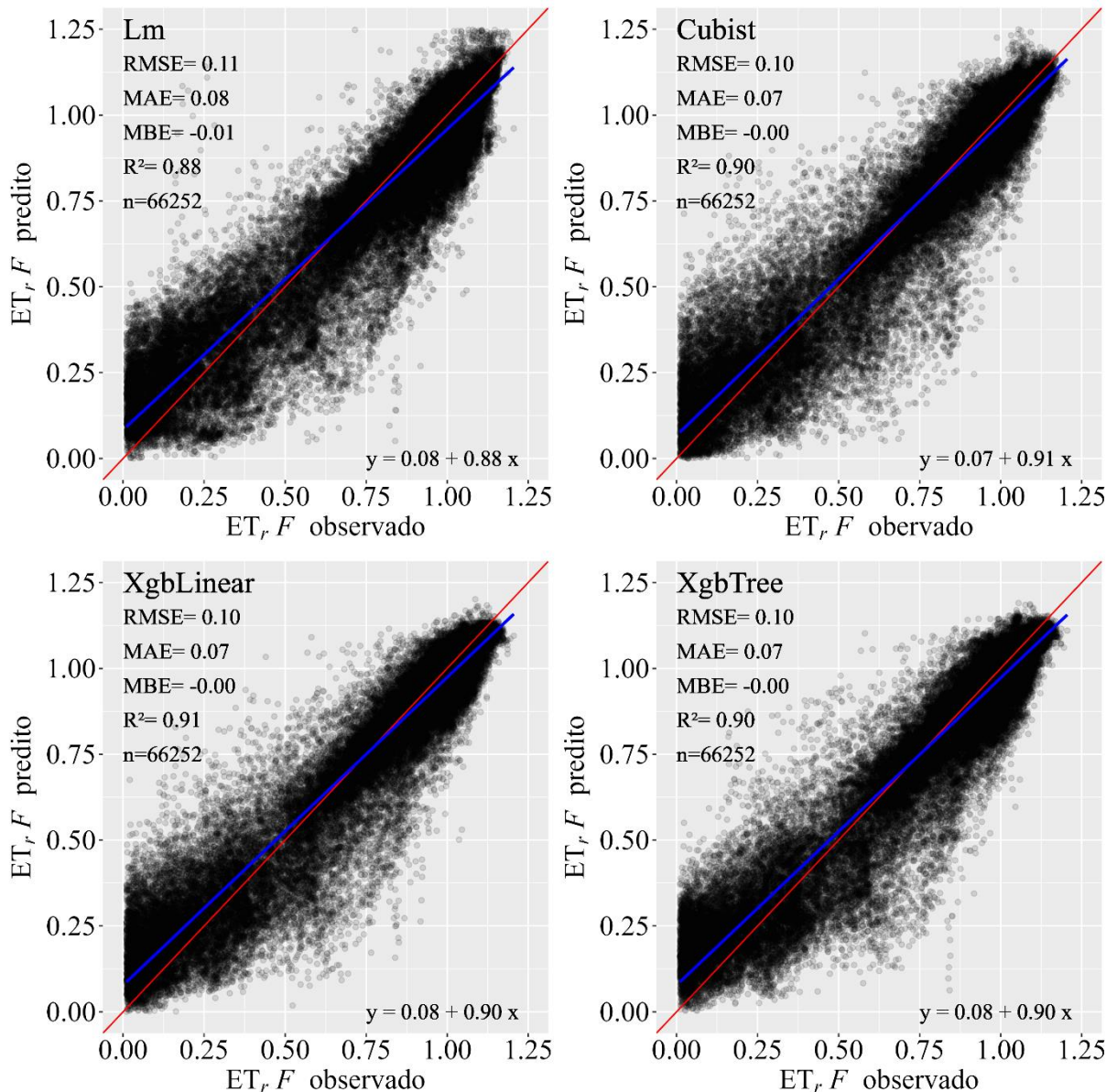


Figura 7. Comparação entre $ET_r F$ observada e predita com as respectivas estatísticas para o teste da abordagem-2

Observa-se na Figura 7 que a Lm, assim como na abordagem-1, obteve resultado inferior aos demais, enquanto Cubist, XgbLinear e o XgbTree foram que melhores conseguiram explicar a variável resposta, apresentando menor RMSE (0.10) e maior R^2 (≥ 90). Conquanto, para o R^2 , assim como na abordagem-1, diferentes valores foram calculados entre esses três últimos modelos, sendo de 0.91 para o XgbLinear e de 0.90 para o Cubist e o XgbTree. Ao comparar as Figuras 6 e 7, nota-se que a abordagem-2 obteve resultados mais satisfatórios que a abordagem-1, apresentando modelos com R^2 superior a 0.88 e RMSE inferior a 0.11, enquanto a abordagem-1 apresentou R^2 inferior a 0.80 e RMSE superior a 0.15. Ainda, constata-se que a abordagem-2 reduziu significativamente a dispersão observada na Figura 6. A melhoria encontrada na abordagem-2 está relacionada ao número de variáveis explicativas, na qual

utilizou-se 12 preditoras. Outro fato, é que os comprimentos de ondas bordas do vermelho e as infravermelhos de ondas curtas apresentam forte relação com conteúdo de água na vegetação, elevando a capacidade preditiva dos modelos treinados. Zhang e Zhou (2019) mencionam que índices de vegetação utilizando borda dos vermelhos tem maior sensibilidade ao conteúdo de água em cultura de milho, com base neste fato é interessante notar que a Tabela 4, para esta abordagem, apresenta 4 preditoras que são índices de vegetação com a presença desse comprimento de onda. Já Santos *et al.* (2020) estudando estresse hídrico e resposta espectral na cultura da cana-de-açúcar identificaram que o comprimento de onda na região do infravermelho de ondas curtas foi sensível ao conteúdo de água nas folhas, assim como Chandel *et al.* (2020) obtiveram resultados parecidos com a cultura do trigo. Desse modo, fortalece as evidências da importância desses comprimentos de onda na quantificação do conteúdo de água, ainda mais quando eles são normalizados com outras bandas através do NRPB.

3.2 Análise residual

A partir da aplicação dos modelos elaborados na imagem do dia 12/02/2020 (Tabela 3), a qual foi destinada à análise residual, gerou-se os resultados presente na Figura 8.

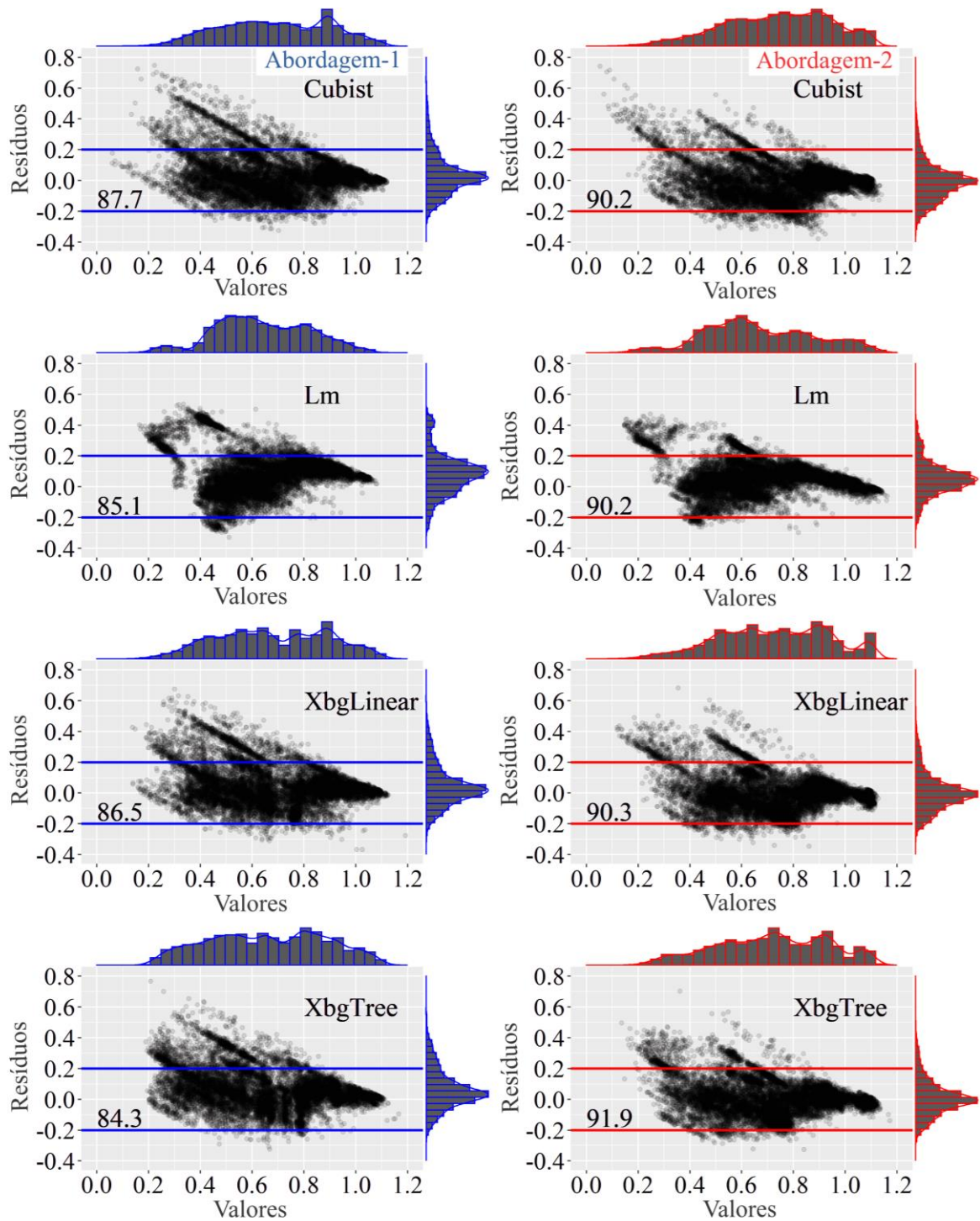


Figura 8. Análise residual da ET_rF estimado através dos modelos de *machine learning*.

Nota-se que a Figura 8 mantém o padrão encontrado na Figura 6 e 7, em que a abordagem-1 apresentou resultado inferior quando comparado à abordagem-2, devido à menor concentração dos pontos residuais próximos ao valor zero. Observou-se bom desempenho para abordagem-2, haja visto que menos de 10.3 % dos valores se afastaram de -0.2 a 0.2, o que indica um baixo erro na predição, com destaque para a XbgLinear. Nota-se que em ambas as abordagens a predição da variável explicativa tendem a ser menos precisa quando o valor de

ET_rF é baixo, sobretudo valores inferiores a 0.6 como discutido anteriormente, tal fato pode ser notado com maior clareza na Figura 9.

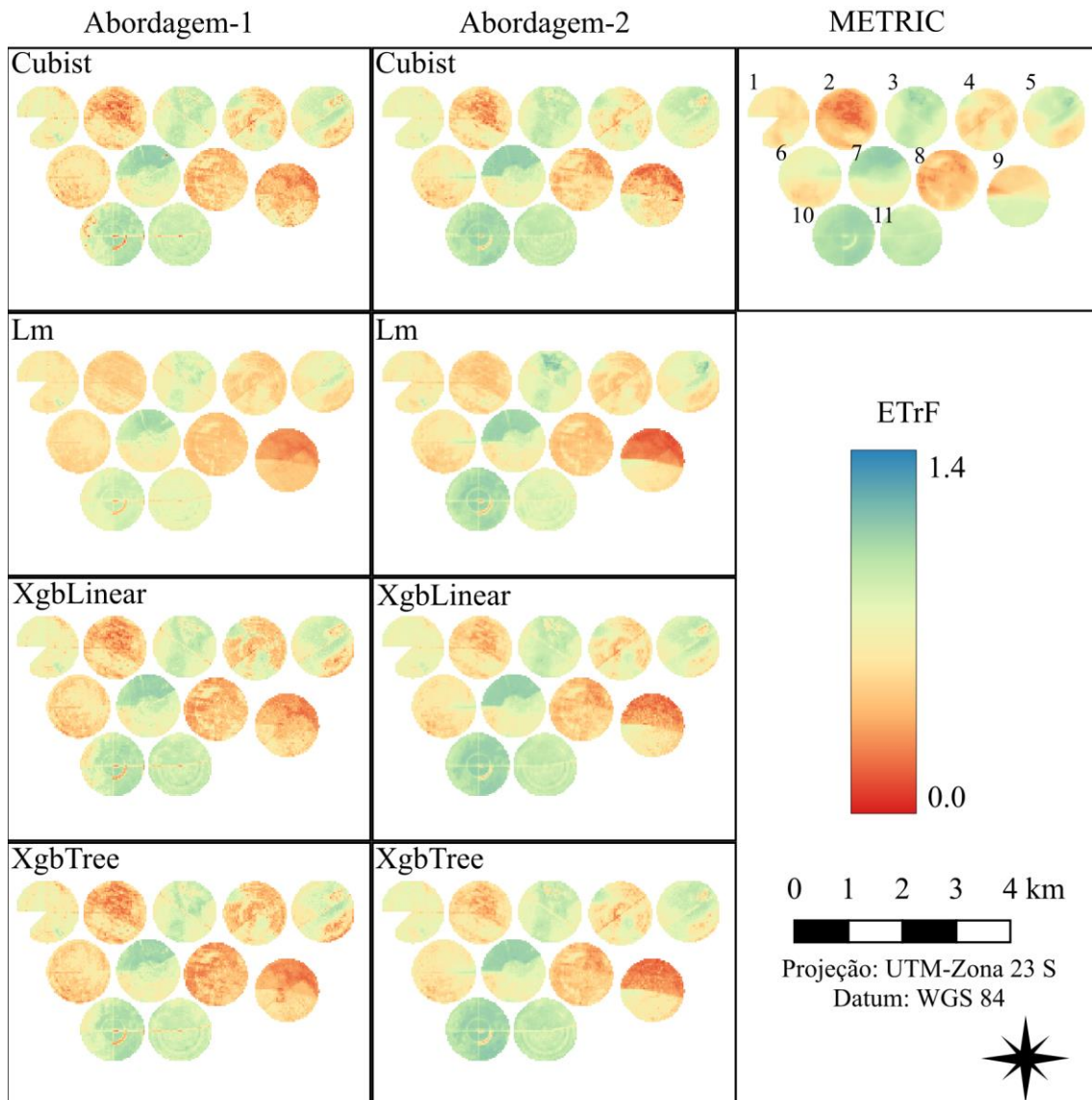


Figura 9. Variabilidade espacial da ET_rF para a abordagem-1, 2 e METRIC para o dia 12/02/2020.

Na Figura 9 é possível notar com maiores detalhes à baixa precisão durante a predição para valores observados menores que 0.6, com destaque para o pivô 9. Nesse pivô, quando o METRIC apresentou valores próximo de 0.6, as abordagens-1 e 2 obtiveram valores preditos próximo a zero, o qual fica mais evidente na Figura 8 das análises dos resíduos. Contudo, esta Figura também revela boa assertividade dos modelos, a qual pode ser visualizado principalmente no pivô 7 e 10, onde os modelos de predição conseguiram capturar riquezas de detalhes que o METRIC estimou. Na região sudeste do pivô 10 apresenta baixo valores de ET_rF pelo METRIC, o qual também foi registrado pelos modelos de *machine learning*. No entanto, para este pivô, foi mais proeminente ao se utilizar as imagens do Sentinel-2 pois elas apresentam

maior resolução espacial, além disso a banda termal no Landsat-8 possui resolução mais grosseira (100 m) que pode mascarar nuances da superfície da área monitorada.

3.3 Aplicação dos modelos

Na Figura 10 está apresentado a distribuição do coeficiente de cultivo (K_c) ao longo do ciclo da cultura da cana-de-açúcar determinado pelo modelo XgbLinear, o qual apresentou as melhores métricas, para as abordagens 1 e 2, pelo modelo METRIC e pela recomendação do boletim da FAO-56 (Allen *et al.*, 1998).

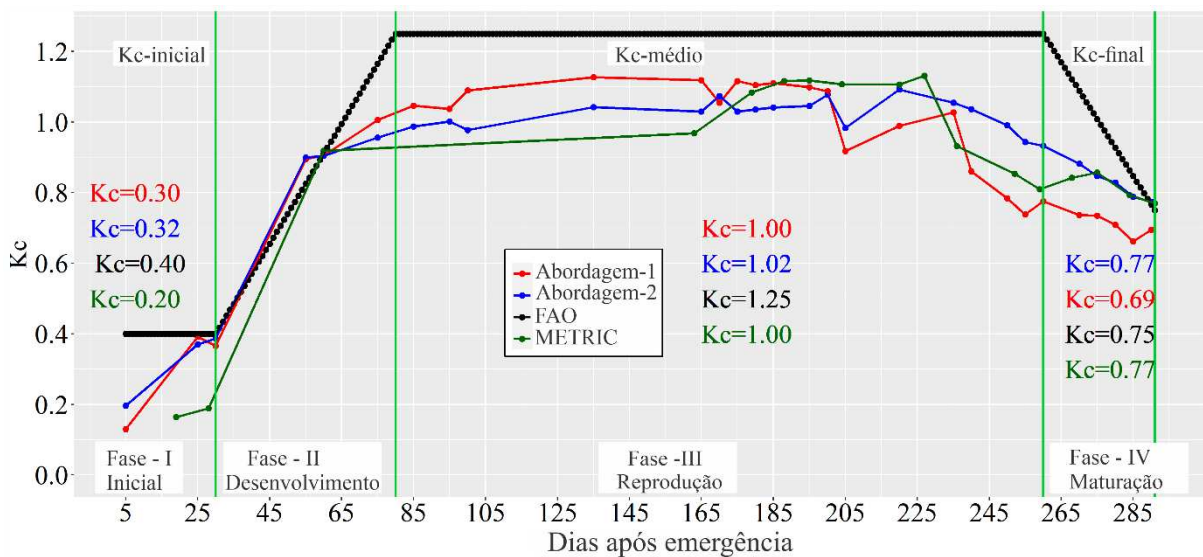


Figura 10. Comparação dos valores do K_c entre as abordagens-1 e 2, METRIC e recomendado pelo boletim 56 da FAO.

A Figura 10 evidencia o que já foi discutido anteriormente, em que na fase inicial da cultura, onde tem-se baixa densidade de cobertura vegetal, os valores da predição pelo modelo de *machine learning* apresentam incertezas em relação ao METRIC, porém à medida que a cultura se desenvolve esse erro tende a reduzir. Para fase I, quando ocorre emergência da cultura, o K_c pela abordagem-1 e 2 foram de 0.30 e 0.32 respectivamente, nessa ordem, ao passo que o METRIC estimou um K_c de 0.20, enquanto o K_c determinado pelo boletim 56 da FAO (ALLEN, R. G., PEREIRA, L. S., RAES, D., SMITH, 2006) é de 0.40. Na fase II, nota-se que as curvas do K_c nas abordagens 1 e 2, vão se aproximando a do METRIC à medida que o estágio fenológico se avança, assim como a curva de K_c -FAO tem proximidade nas duas abordagens e pequeno afastamento da curva do METRIC. Para a fase III, crescimento da cultura, o K_c médio foi de 0.98 para abordagem-1 e 1.00 para o METRIC, enquanto na abordagem-2 foi de 1.02, já o boletim FAO 56 para esta fase recomenda utilizar o valor de 1.25. Ainda na fase III, a curva do METRIC tem comportamento similar com a Abordagem-2, o que reforça os resultados encontrado no teste dos modelos. A fase IV, maturação, a abordagem-1 foi a que mais se afastou,

tanto do K_c FAO, quanto do METRIC, tendo seu valor igual a 0.69, ao passo que a abordagem-2 foi de 0.76 e o METRIC de 0.77. Dingre e Gorantiwar (2020) estimando o K_c para a cana-de-açúcar pelo método do balanço hídrico encontraram valores médios para fase I, III e IV iguais a 0.36, 1.20 e 0.78. Já Silva *et al.* (2012) para região semiárido brasileira encontraram K_c na cana soca para fase I, III e IV iguais a 0.65, 1.10 e 0.85, sendo diferente do recomendado pela FAO. Com isso nota-se divergência em valores de K_c presente em literaturas em relação ao recomendado e isso pode ser explicado pela especificidade do clima do local onde cada estudo é realizado. A fase III, para este trabalho, atingiu valor máximo de 1.09 contudo tendo variações ao longo de toda fase e tal variação pode estar relacionado com a condições reais que está acontecendo em campo, em outra palavras, o METRIC, assim como o modelo treinado, quantifica não só o K_c , mas o produto entre K_c e K_s (coeficiente de estresse), no entanto é referido em Allen *et al.* (2007) como K_c .

Ainda na Figura 10, é perceptível que o Sentinel-2, em função da sua melhor resolução temporal, foi possível obter maior número de informações do K_c ao longo do ciclo da cultura em relação ao Landsat-8, o qual torna-se importante para obtenção de maior variabilidade temporal deste coeficiente, podendo acarretar um manejo da irrigação mais assertivo em relação a utilização apenas dos dados do Landsat-8. Saleem e Awange (2019) mencionam que o Sentinel-2 representa a nova era para a obtenção de informações mais precisas da superfície terrestre devido seu sensor possuir maior resolução espacial e temporal entre os satélites que disponibilizam as imagens de forma gratuita. Contudo, a informação referente ao K_c pode ser ampliada quando ao utilizar a união das duas plataformas orbitais, uma vez que será obtido maior número de informações com maior frequência de imagens, como destacado em (FILGUEIRAS *et al.*, 2020).

A evapotranspiração atual da cultura em nível espacial de 10, 20 e 30 metros para a abordagem-1, abordagem-2 e METRIC, pode ser observado na Figura 11.

Na Figura 11 chama atenção o DAE 030 em que é observado um raio com ET_a maior que no restante do pivô, pois a cultura está em fase inicial de emergência e o solo encontra-se em sua maior área exposto, o qual recebe água proveniente da irrigação e este raio corresponde ao solo exposto úmido em que é notado deslocamento no sentido horário do equipamento de irrigação, ou seja, as informações com 10 e 20 m de resolução permitem visualizar detalhes que o modelo METRIC, em função do uso das imagens termais, não permitiu. Informações com detalhes espaciais, como essas vistas pelas imagens Sentinel-2 na Figura 11 são de grande valia para profissionais do campo. As datas coincidentes entre Sentinel-2 e Landsat-8 mostraram elevada similaridades na ET_a espacial entre a abordagem-1 e 2 com o METRIC, com destaque para a abordagem-2 em que teve maior proximidade espacial. Tal proximidade é evidenciado na Tabela 5 em que as médias da ET_a na abordagem-2 possuem menores diferenças das médias estimadas pelo modelo METRIC. Além disso o desvio padrão para a abordagem-2, também foi menor em relação a abordagem-1, porém maior que o modelo METRIC. Esse fato pode estar atrelado a resolução espacial mais detalhadas das abordagens 1 e 2 em relação ao método que utiliza o modelo METRIC.

Tabela 5. Médias da evapotranspiração atual (ET_a) para datas coincidentes entre Sentinel-2 e Landsat-8.

DAE	Abordagem-1 (mm)	Abordagem-2 (mm)	METRIC (mm)	Diferença (%)	
				Abor-1	Abor-2
060	6.71±0.44	6.63±0.36	6.75±0.25	0.59	1.77
195	4.52±0.22	4.48±0.09	4.78±0.06	5.44	6.28
220	4.25±0.21	4.68±0.11	4.79±0.08	11.27	2.30
275	3.75±0.36	4.32±0.20	4.44±0.17	16.54	2.77
ET_a -Total	1417.77	1474.26	1544.11	8.18	4.52

Após realizar a integral da ET_a , constatou-se que a demanda hídrica da cana-de-açúcar foi de 1417.77 mm para a abordagem-1, 1474.26 mm para a 2 e 1544.11 mm para o METRIC. É perceptível que a evapotranspiração total estimada pelas abordagens-1 e 2 foram próximas ao valor evapotranspirado total pelo METRIC, onde as diferenças percentuais foram de 8.18 da abordagem-1 e 4.52 para a abordagem-2 em relação ao METRIC. A abordagem-2 teve seu valor mais próximo do valor estimado pelo METRIC, indo de encontro dos valores encontrados para as datas destacadas na Figura 10. A evapotranspiração total da cana-de-açúcar durante seu ciclo encontrado por Dingre e Gorantiwar (2020) foi de 1388 mm, já o valor encontrado por Silva *et al.* (2015) para condições do Nordeste brasileiro foi 1600 mm. Assim, nota-se que a evapotranspiração total encontrada nesse trabalho se aproxima do valor encontrado em outras regiões.

4 Conclusão

O uso do sensor MSI combinado com técnicas de *machine learning* se mostraram uma alternativa nova e simples para estimar a ET_rF através de informações espectrais, complementando a estimada pelo modelo METRIC utilizando os sensores OLI e TIRS, aumentando a frequência em que informações são geradas para áreas de interesse.

A abordagem-2, em função do maior número de preditoras, apresentou resultados estatísticos superiores a abordagem-1, principalmente para o modelo XgbLinear que obteve R^2 de 0.91 e RMSE de 0.10, já as métricas para o mesmo modelo, considerando a abordagem-1, foi de 0.80 e 0.15 para o R^2 e RMSE, respectivamente. Contudo, todos os modelos desenvolvidos mostraram limitações quando a variável dependente apresenta valores de ET_rF inferiores a 0.6.

A aplicação dos modelos suporta evidência que a abordagem-2 foi melhor avaliada estatisticamente com a variável resposta, em que suas médias apresentaram resultados próximos ao valor médio estimado pelo modelo METRIC para datas coincidentes. Portanto, a abordagem 2 é a recomendada para estimar a ET_rF na cana-de-açúcar na área onde o modelo foi treinado.

O K_c da cultura, identificado para a fase vegetativa I, III, e IV, na abordagem-1 foram de 0.30, 0.98 e 0.69, respectivamente, enquanto pela abordagem-2 foram de 0.32, 1.02 e 0.76, nessa ordem.

Por fim, este trabalho traz uma possibilidade da utilização de outros sensores remotos para computar espacialmente e temporalmente a evapotranspiração sendo enriquecedor para o debate científico para aplicação na irrigação de precisão.

2.2 Artigo 2

Evapotranspiração atual por meio de sensoriamento remoto e técnicas de *machine learning*

Resumo

Os avanços tecnológicos recentes têm permitido produzir muitos estudos sobre evapotranspiração, resultando em melhorias nas estimativas da evapotranspiração de referência e nos coeficientes de cultivo com dados via sensoriamento remoto. Entretanto, muitas vezes essas duas áreas de pesquisa funcionam independentemente, produzindo estudos valiosos, mas sem uma integração efetiva para prever a evapotranspiração atual diretamente, sem a necessidade de estações meteorológicas. Dessa forma, este trabalho teve como objetivo modelar a evapotranspiração atual na cultura da cana-de-açúcar usando técnicas de *machine learning*, independentemente de estações meteorológicas e dados de sensores térmicos. Para alcançar esse objetivo, foram usados dados dos sensores OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*) embarcados nos satélites Landsat-8 e 9 para produzir a variável observada a partir do modelo METRIC, e dados dos satélites Sentinel-2A e 2B, Nasa Power, WorldClim e variáveis astronômicas, latitude, elevação, dia do ano e mês para gerar as variáveis explicativas e alimentar 13 modelos de *machine learning* em três biomas diferentes: Mata Atlântica, Caatinga e Cerrado. Os resultados indicaram que o modelo brnn (*bayesian regularized neural networks*) com R^2 e RMSE de 0.73 e 1.10, respectivamente, e o XgbLinear (*extreme gradient boosting* - método linear) que obteve valores de 0.74 e 1.25 para essas métricas, nessa ordem, apresentaram o melhor desempenho geral. Análises específicas indicaram que o brnn foi superior para áreas cultivadas no bioma Mata Atlântica e Caatinga, enquanto o XgbLinear foi superior no bioma Cerrado. Esses resultados mostram que os algoritmos de *machine learning* têm a capacidade de prever a evapotranspiração atual sem a necessidade do uso de estações meteorológicas e dados térmicos.

Palavras-chave: Cana-de-açúcar. Irrigação. Inteligência artificial.

1 Introdução

O processo evapotranspirativo, que representa a perda de água através das plantas, solo e corpos hídricos, é amplamente estudado em diversas áreas, incluindo hidrologia, climatologia, meteorologia, ciências ambientais e agronomia. Desde 1948, autores como Penman (1948) e Thornthwaite (1948) se dedicaram ao estudo desse tema e são considerados referências na área. Desde então, muitos estudos foram conduzidos para expandir e aprofundar o conhecimento sobre o processo evapotranspirativo.

A evapotranspiração atual (ET_a) das culturas agrícolas é uma variável importante que permite o estudo do balanço hídrico do solo e a estimativa do teor de umidade. Especificamente em áreas agrícolas irrigadas, essa evapotranspiração permite quantificar a adequada reposição das lâminas de irrigação. No entanto, para estimar a ET_a comumente são necessárias estações meteorológicas em campo para computar a evapotranspiração de referência da grama (ET_o) ou alfafa (ET_r). Normalmente, se utiliza a ET_o multiplicado por um coeficiente de cultivo (K_c), que é tabelado (ALLEN *et al.*, 1998) ou determinado experimentalmente, e por um coeficiente de estresse (K_s), que está relacionado à umidade do solo, para encontrar a ET_a (ABOU ALI *et al.*, 2023; ALLEN *et al.*, 1998; KIMBALL *et al.*, 2023; KOCIAN *et al.*, 2023; XIANG *et al.*, 2020). Porém, as estações meteorológicas precisam de sensores de radiação solar, umidade relativa, temperatura do ar e velocidade do vento para determinar a ET_o , que é adotada como padrão pelo boletim 56 da FAO (*Food and Agriculture Organization*) (ALLEN *et al.*, 1998). Por conta disso, essas estações podem ser caras e requerem manutenção e reparos em seus sensores, além da demanda técnica para análise e interpretação dos dados. Os K_c são valores fixos, geralmente em três estágios da cultura: inicial, reprodutivo e final, com interpolação para o estágio de desenvolvimento (ALLEN *et al.*, 1998). Esses valores não são constantes, mesmo assim são adotados por questão de praticidade da aplicação em campo. Por fim, o coeficiente de estresse é determinado por modelos matemáticos em função da disponibilidade de água no solo.

Devido às dificuldades e aos custos das estações meteorológicas, muitos estudos buscam simplificar a obtenção da evapotranspiração de referência por meio do uso de um menor número de variáveis meteorológicas. Com os avanços computacionais, algoritmos de inteligência artificial estão sendo cada vez mais utilizados para a obtenção mais precisa da evapotranspiração de referência com variáveis reduzidas, o qual pode ser observado em diversos estudos (ALTHOFF *et al.*, 2019, 2020; BAYRAM; ÇITAKOĞLU, 2023; BELLIDO-JIMÉNEZ; ESTÉVEZ; GARCÍA-MARÍN, 2021; DONG *et al.*, 2022; MAKWANA; TIWARI; DEORA, 2023; SAGGI; JAIN, 2019; SATTARI *et al.*, 2021).

O coeficiente de cultura, K_c , também tem sido aprimorado para evitar o uso de dados tabelados. As melhorias foram alcançadas por meio da utilização do sensoriamento remoto, que pode empregar tanto índices de vegetação para estimar empiricamente, como é observado em diversos estudos (DINGRE; GORANTIWAR; KADAM, 2021; GOFFIN *et al.*, 2022; KAMBLE; KILIC; HUBBARD, 2013; PELTA *et al.*, 2022), quanto modelos físicos baseados em dados nas faixas do comprimento de onda do visível, infravermelho próximo, de ondas curtas e termal (ALLEN; TASUMI; TREZZA, 2007; ANDERSON *et al.*, 2004; BASTIAANSEN *et al.*, 1998; MECIKALSKI *et al.*, 1999; NORMAN *et al.*, 2000; SENAY; BUDDE; VERDIN, 2011; TEIXEIRA, 2010; YANG *et al.*, 2015). No entanto, o uso do sensoriamento remoto para quantificar o K_c , baseados em índices de vegetação, podem ter os parâmetros de ajuste da equação enviesado para o local de desenvolvimento. Enquanto os modelos físicos mais complexos necessitam para sua aplicação dados de sensores termais capazes de registrar a emitância da superfície.

Os sensores térmicos têm um grande potencial de aplicação na identificação de áreas com estresse hídrico em culturas (ZHOU *et al.*, 2021) e são amplamente utilizados em modelos para determinar a evapotranspiração. Contudo, apenas alguns satélites têm esse tipo de sensor embarcado com resolução espacial adequada para a agricultura, o que dificulta o uso mais frequente desses dados. Em função dessa resolução, é comum a aplicação e estudos utilizando dados do satélite Landsat-8, visto que ele fornece dados nos comprimentos de onda do visível ao infravermelho termal em uma resolução espacial de 30 metros, porém com resolução temporal de 16 dias. Com recém lançamento do Landsat-9 que possui os mesmos sensores do Landsat-8 a disponibilidade da informação espectral passou a ser de 8 dias o que representa significativa evolução. Apesar do avanço, ainda é insuficiente para obter informações com elevada frequência temporal da área monitorada, especialmente em áreas de agricultura irrigada que requerem estimativas frequentes de evapotranspiração. Para superar essa limitação, é possível integrar os dados dos satélites Landsat-8 e 9 com os dados dos satélites Sentinel-2A e 2B, que disponibiliza informação espectral a cada 5 dias. A harmonização entre os dados Landsats e Sentinels favorece a obtenção de um maior número de informações da área de interesse (JAAFAR; MOURAD; SCHULL, 2022; STUMPF; MICHÉA; MALET, 2018), o qual pode permitir estimativas mais precisas da evapotranspiração.

Com base no exposto acima, percebe-se que apesar dos avanços com sensoriamento remoto para estimar a ET_a , ainda apresenta desafios quanto maior resolução temporal para garantir estimativa com maior frequência, além disso muitos estudos estão sendo realizados isoladamente para melhorar a estimativa do coeficiente de K_c e outros para melhorar a

estimativa da ET_o . Em contraste a isso, este estudo apresenta uma abordagem inovadora, pois tem como objetivo quantificar a evapotranspiração atual da cultura da cana-de-açúcar, sem a necessidade de estações meteorológicas ou coeficientes de cultivo tabelados, ou seja, a proposta é determinar diretamente a ET_a sem a necessidade de estimar K_c , K_s nem ET_o . Para tal, os dados espectrais do Sentinel-2 são fundamentais nesse estudo, pois possui sensor com capacidade de adquirir dados da superfície terrestre em nove comprimento de onda com resolução de 20 metros e a cada 5 dia. Dessa forma, é possível reduzir os custos operacionais em campo com a instalação e manutenção de estações meteorológicas, além de reduzir erros proporcionados pelos dados tabelados e ainda tem o ganho da harmonização com Landsat-8 e 9, o qual proporciona maior a frequência temporal da ET_a .

Os satélites são tecnologia excelentes para modelagem, já que geram uma grande quantidade de dados que, quando adequadamente modelados, se convertem em informações úteis. Para lidar com grandes volumes de dados (*big data*), são necessários modelos robustos com capacidade de modelar a variável de interesse a partir desse banco de dados. Nos dias de atuais, os algoritmos de *machine learning* estão ganhando notoriedade devido sua capacidade em realizar previsões com maior velocidade de processamento e menor erro utilizando *big data* (DI NUNNO; GRANATA, 2023; REICHSTEIN *et al.*, 2019; SHANG *et al.*, 2023). Esses algoritmos são conjuntos de regras e procedimentos matemáticos que permitem que um sistema de computador aprenda com dados e melhore seu desempenho em uma tarefa específica sem ser programado explicitamente.

A partir da capacidade preditiva dos algoritmos de *machine learning* e da necessidade de se obter modelos precisos e independentes de estações meteorológicas para estimar a evapotranspiração atual da cana-de-açúcar em alta resolução temporal, este trabalho tem como objetivo geral utilizar algoritmos robustos de *machine learning* para modelar a evapotranspiração atual da cana-de-açúcar em todo o Brasil. Os objetivos específicos deste estudo são: 1) Avaliar o desempenho de cada modelo e selecionar os melhores; 2) Analisar o comportamento dos melhores modelos nas áreas onde os dados foram coletados; e 3) Aplicar o melhor modelo.

2 Materiais e Métodos

2.1 Área de estudo

O estudo foi realizado em três biomas brasileiros: a Caatinga, o Cerrado e a Mata Atlântica. O primeiro possui área de cobertura terrestre equivalente a 844453 km², o segundo possui 2036448 km² e o último possui 1110182 km², sendo estes os principais biomas

cultivados com cana-de-açúcar no país (Figura 1). A latitude mínima considerando todos os biomas do estudo é de -15.16 sul, referente ao bioma Caatinga, e -21.68 sul, referente à Mata Atlântica, o que resulta em uma elevada variação climática na área de estudo.

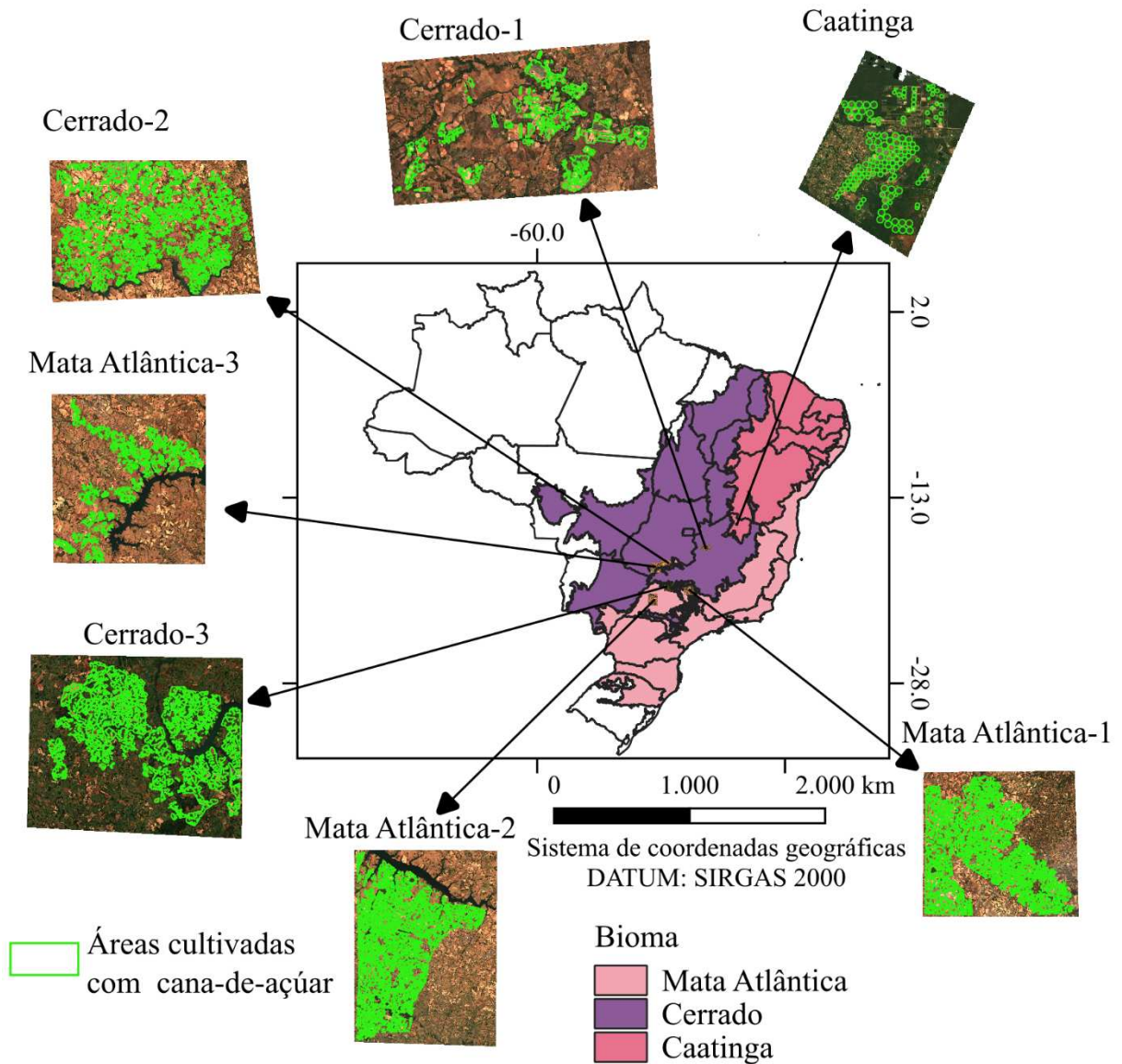


Figura 1. Localização da área de estudo

A escolha dessas áreas ocorreu por apresentar maior compatibilidade de imagens entre Landsat-8 e 9 com Sentinel-2A e 2B, assim como grandes variações climatológicas dentro de cada área como pode ser observado na Figura 2.

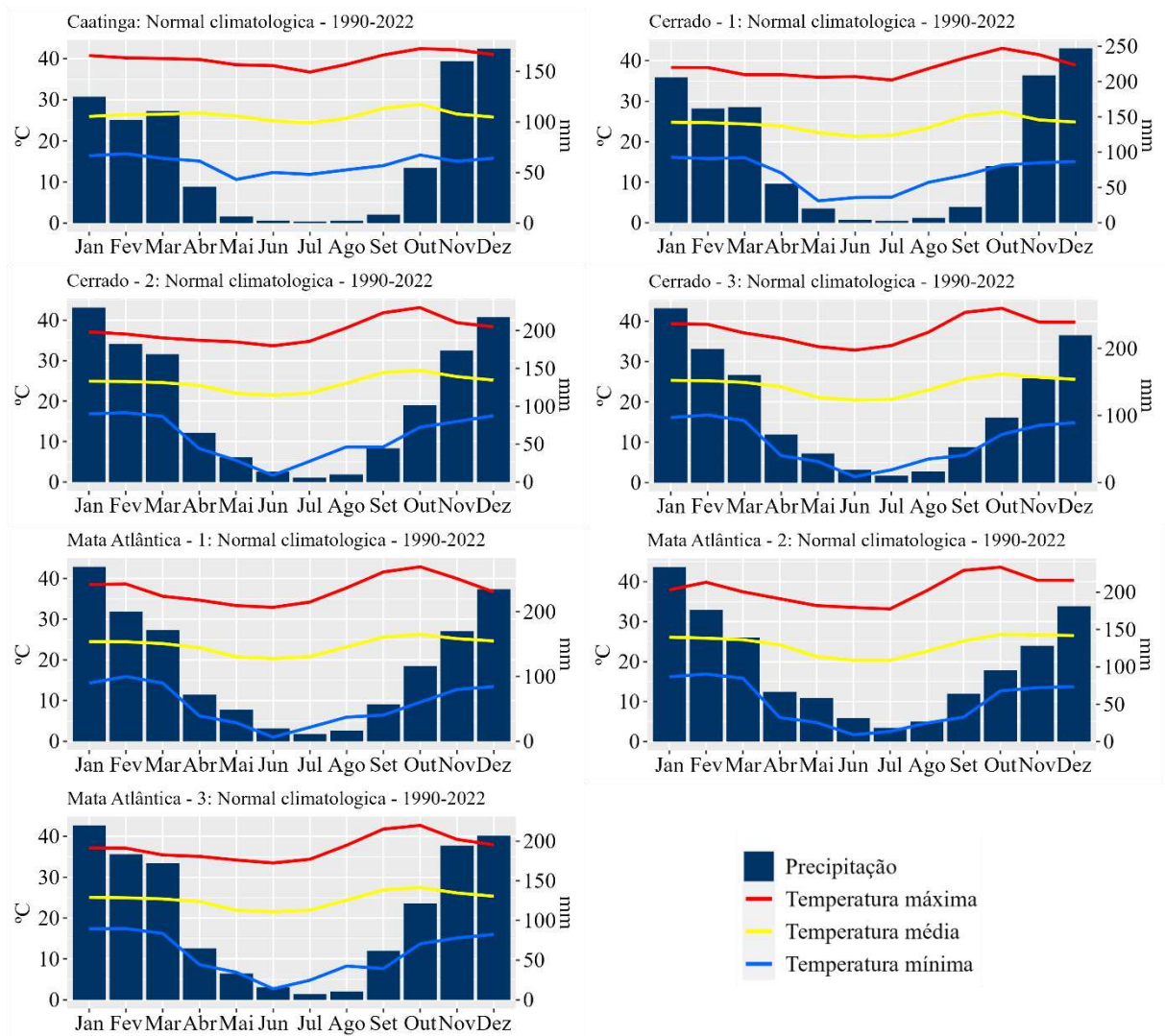


Figura 2. Climatologia com base nos produtos Nasa Power das áreas selecionadas para o estudo.

Nesses três biomas, foram coletados dados amostrais disponibilizados dos sensores *Operational Land Imager (OLI)* e *Thermal Infrared Sensor (TIRS)*, a bordo dos satélites *Landsat-8* e *Landsat-9*, os quais foram utilizados para gerar a variável resposta. Além disso, na mesma área, foram coletados dados do sensor *MultiSpectral Instrument (MSI)*, presente nos satélites *Sentinel 2A* e *2B*. Esses dados foram combinados com variáveis meteorológicas e climatológicas para produzir as variáveis preditivas. De maneira simplificada, todo esse procedimento pode ser visualizado na Figura 3.

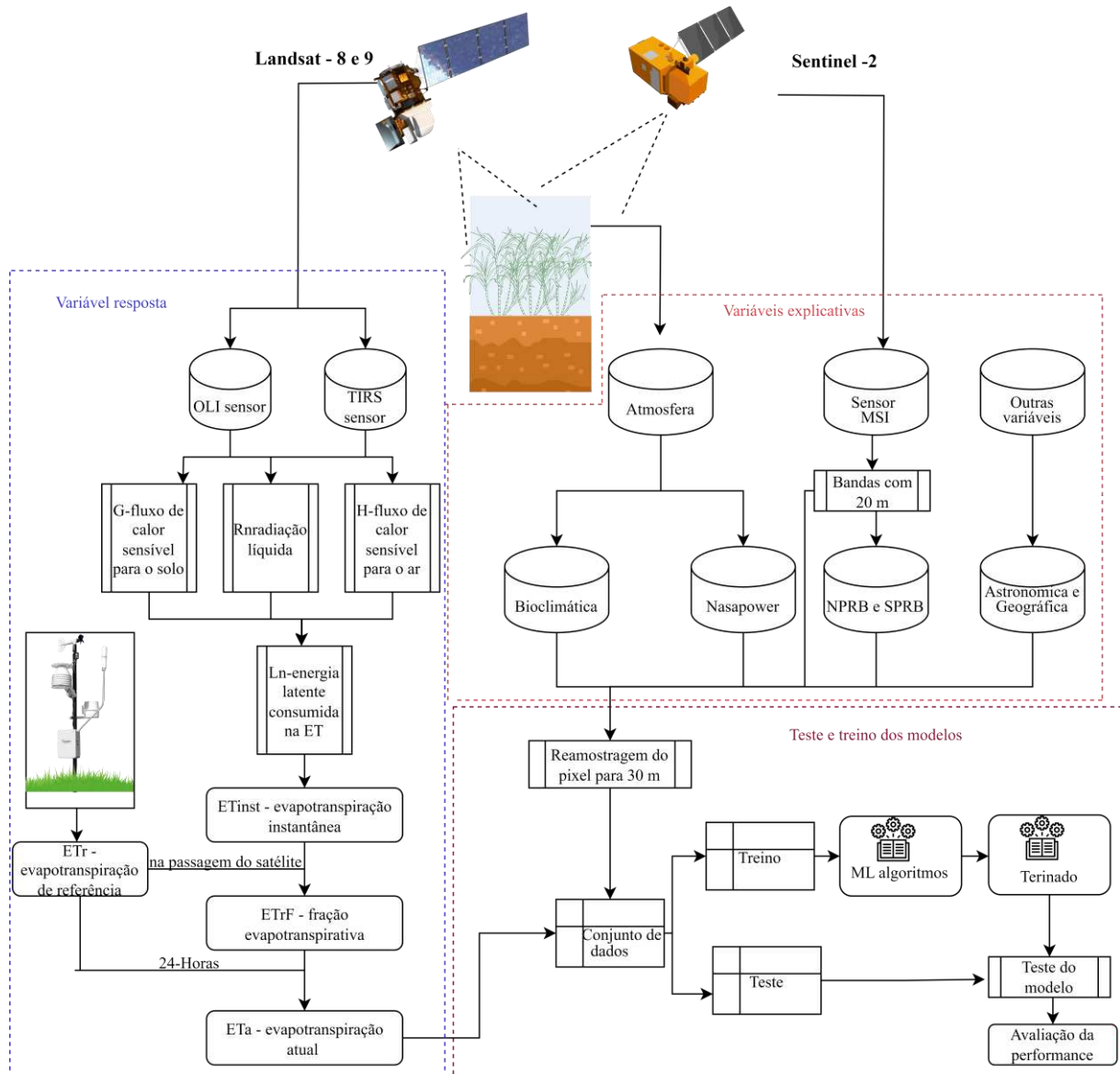


Figura 3. Fluxograma com procedimento metodológico simplificado

2.2 Variável resposta - Evapotranspiração atual

Utilizou-se como variável a ser modelada a ET_a da cana-de-açúcar para as áreas da Figura 1. Para tal, aplicou-se o método do balanço de energia desenvolvido por Allen *et al.* (2007) conhecido como *Mapping EvapoTranspiration at high Resolution with Internalized Calibration* (METRIC) e esse utiliza uma série de complexos cálculos baseados em reflectância e emitância da superfície terrestre adquirido via sensoriamento remoto, sendo, majoritariamente, através dos sensores a bordo dos satélites Landsats e de uma estação meteorológica dentro de cada cena auferidas dos sensores orbitais.

2.2.1 Aquisição dos dados espectrais e meteorológicos

Foram adquiridos, no total, 53 imagens de satélites entre os anos de 2017 a 2022 (Tabela 1), para as diferentes localidades da Figura 1 e diferentes datas, do sensor OLI e TIRS

embarcado no satélite Landsat-8 e Landsat-9 através da plataforma *United States Geological Survey* (USGS) (<https://earthexplorer.usgs.gov/>), além de dados meteorológicos. Para ambos os sensores, foram adquiridos os produtos da coleção-2 em nível-2, em resolução espacial de 30 m.

Tabela 1. Características das imagens e a área onde elas foram adquiridas.

Bioma	Área	Nº de imagens	Path/Row	Ano
Caatinga	Caatinga	11	218/071	2018-2022
	Cerrado-1	6	220/072	2017-2022
Cerrado	Cerrado-2	9	222/073	2017-2022
	Cerrado-3	4	221/074	2018-2022
Mata Atlântica	Mata Atlântica-1	6	220/074	2018-2022
	Mata Atlântica-2	7	222/075	2018-2022
	Mata Atlântica-3	10	222/073	2017-2022

Os dados meteorológicos adquiridos foram temperatura máxima e mínima do ar, radiação solar, umidade relativa média e velocidade do vento médio nas estações meteorológicas da Tabela 2. Adquiriu-se estes dados em intervalo de tempo horário (GMT-3) para as respectivas datas de passagem de satélites na base de dados do Instituto Nacional de Meteorologia (INMET).

Tabela 2. Estações meteorológicas com suas localidades e código.

Bioma	Área	Código-INMET	Lat/long	Path/Row
Caatinga	Caatinga	A539	-15.09 / -44.02	218/071
	Cerrado-1	A542	-16.55 / -46.88	220/072
Cerrado	Cerrado-2	A035	-18.41 / -49.19	222/073
	Cerrado-3	A748	-20.56 / -48.54	221/074
Mata Atlântica	Mata Atlântica-1	A753	-20.36 / -47.78	220/074
	Mata Atlântica-2	A734	-21.32 / -50.93	222/075
	Mata Atlântica-3	A011	-18.97 / -50.63	222/073

2.2.2 Determinação da ET_a

Os cálculos para alcançar o valor da ET_a dependem das energias de radiação líquida (R_n), fluxo sensível do calor transferido para o solo (G) e fluxo sensível do calor convectado para o ar (H), os quais envolvidos é possível extrair a LE que é a energia latente consumida no processo evapotranspirativo da superfície, assim, a LE, em $W m^{-2}$, pode ser determinada pela Eq.1.

$$LE = R_n - G - H \quad (1)$$

A R_n é a energia radiante da superfície que é particionada em H , G e LE, e é determinada pela Eq. (2)

$$R_n = R_{s\downarrow} - \alpha R_{s\downarrow} + R_{L\downarrow} - R_{L\uparrow} - (1 - \varepsilon_0) R_{L\downarrow} \quad (2)$$

onde $R_{s\downarrow}$ é a entrada da radiação de ondas curtas (W m^{-2}); α é o albedo de superfície (adim.), determinado pelas bandas 2, 3, 4, 5, 6 e 7 do Landsat-8 e 9; $R_{L\downarrow}$ e $R_{L\uparrow}$ é a entrada e saída de ondas longas (W m^{-2}), respectivamente; e ε_0 é a emissividade térmica da superfície.

O calor transferido para o solo foi determinado utilizando o modelo de Tasumi (2003) que envolve a R_n , índice de área foliar (LAI), estimado pela metodologia aplicada por Allen *et al.* (2011), e temperatura de superfície (T_s). Este parâmetro necessita de uma condicional, onde se a LAI for igual ou superior a 0.5, então faze-o uso da Eq. (2a), senão, faz-se o uso da Eq. (2b).

$$\frac{G}{R_n} = 0.05 + 0.18e^{-0.521 \text{ LAI}} \quad \text{para LAI} \geq 0.5 \quad (2a)$$

$$\frac{G}{R_n} = \frac{1.80(T_s - 273.15)}{R_n + 0.084} \quad \text{para LAI} < 0.5 \quad (2b)$$

O fluxo sensível do calor convectado para o ar é o processo mais complexo nos modelos de balanço de energia, fruto, por algumas vezes, de erro na estimativa da ET_a . Para estimá-lo utiliza-se de uma função aerodinâmica, Eq. (3), que considera a resistência aerodinâmica em duas diferentes alturas (geralmente 0.1 m e 2.0 m do solo) que, por sua vez, é calculado usando a velocidade do vento extrapolada de alguma altura acima da superfície do solo (normalmente 100 para 200 m) e um esquema de correção de estabilidade interativa com base nas funções Monin-Obhukov. Além da resistência aerodinâmica, outro parâmetro é a diferença de temperatura para as alturas 0.1 e 2.0 m, o qual utiliza a temperatura de superfície proveniente de imagens de satélite corrigida por um modelo linear (Allen *et al.*, 2007).

$$H = \rho_{\text{air}} C_p \frac{dT}{r_{ah}} \quad (3)$$

onde ρ_{air} é a densidade do ar (kg m^{-3}); C_p é o calor específico do ar a pressão constante ($\text{J kg}^{-1} \text{K}^{-1}$); dT é a diferença de temperatura entre duas alturas (z_1 e z_2) em uma área próxima à superfície; e r_{ah} é a resistência aerodinâmica (s m^{-1}) entre essas duas alturas.

Após calculado a LE, determina-se a evapotranspiração instantânea (Eq. 4) que é aquela correspondente ao momento exato de passagem do satélite convertido para hora e leva em consideração, além da LE, a densidade da água e o calor latente de vaporização. A ET_{inst} é necessária para o cálculo da fração evapotranspirativa (ET_rF) que, juntamente com dados meteorológicos, determina-se a ET_a .

$$ET_{inst} = 3600 \frac{LE}{\lambda \rho_w} \quad (4)$$

onde ET_{inst} é a evapotranspiração instantânea (mm.h^{-1}); 3600 é a conversão de segundos, intervalo de tempo da passagem do satélite, para hora; LE é a energia latente consumida na ET (W m^{-2}); ρ_w é densidade da água ($\sim 1000 \text{ kg m}^{-3}$); e λ é o calor latente de vaporização (J kg^{-1}).

O λ representa o calor absorvido quando um kg de água é evaporada e foi calculado pela Eq. (5).

$$\lambda = [2.501 - 0.00236(T_s - 273.15)] \times 10^6 \quad (5)$$

onde T_s é a temperatura de superfície ($^{\circ}\text{K}$) determinada pela banda 10 do sensor TIRS (Tabela 1).

A ET_rF se baseia na razão entre a evapotranspiração instantânea (ET_{inst}), em cada *pixel*, pela evapotranspiração de referência horária para alfafa (ET_r), calculada pela estação meteorológica que se encontra dentro de cada cena orbital (Tabela 2). Assim, a ET_rF é determinada a partir da Eq. (6).

$$ET_rF = \frac{ET_{inst}}{ET_r} \quad (6)$$

onde ET_{inst} é a evapotranspiração instantânea (mm.h^{-1}) e ET_r é evapotranspiração de referência (mm.h^{-1}) padronizada para alfafa a 0.5 m de altura no momento da passagem do satélite.

Por fim, estabeleceu-se a evapotranspiração atual na cana-de-açúcar. Para isso, calculou-se a evapotranspiração de referência acumulada em um período de 24 h (ET_{r-24}) no dia que cada imagem foi adquirida. A ET_{r-24} , em cada data, foi calculada utilizando a equação Penman-Monteith padronizado pela ASCE (*American Society of Civil Engineers*) (ALLEN *et al.*, 2005) e os dados meteorológicos foram adquiridos nas estações disponíveis na Tabela 2. Finalmente, com os dados da ET_{r-24} e do ET_rF , determinou-se a ET_a pela Eq. (7).

$$ET_a = ET_rF \times ET_{r-24} \quad (7)$$

Todo o cálculo até chegar na ET_a descrito nesse tópico foi realizado utilizando o pacote *Water*, desenvolvido por Olmedo *et al.* (2018) para ambiente de linguagem R (R CORE TEAM, 2020).

2.3 Seleção das variáveis explicativas

A predição da variável resposta baseou-se em dois dos principais parâmetros diretamente relacionados ao processo evapotranspirativo: a planta e o clima. Para o parâmetro

planta, foram utilizados dados espectrais provenientes do sensor MSI embarcado nos satélites Sentinel. Já para o clima, foram utilizadas quatro fontes distintas, direta e indiretamente relacionadas: 1) dados meteorológicos do produto Nasa power (CHANDLER *et al.*, 2013); 2) dados climatológicos do produto WorldClim (FICK; HIJMANS, 2017); 3) dados de latitude local de cada *pixel*, elevação, dia do ano e o mês; e, por fim, 4) dados astronômicos. Portanto, para a adequada modelagem e identificação do padrão na variável a ser modelada, diferentes fontes e tipos de dados têm grande relevância.

2.3.1 MultiSpectral Instrument (MSI)

Os dados do sensor MSI foram coletados, em formato de imagem, exatamente no mesmo dia de passagem dos satélites Landsat-8 e 9. Essa harmonização visa adquirir informações espectrais da lavoura de cana-de-açúcar em condições atmosféricas e fisiológicas da cultura similares para os sensores embarcados nos satélites Landsats, evitando, desta forma, que diferentes mudanças atmosféricas e/ou na lavoura produzam resultados diferentes nos dados espectrais.

Tabela 3. Características das imagens do MSI e a área onde elas foram adquiridas.

Bioma	Área	Nº de imagens	Title	Ano
Caatinga	Caatinga	11	T23LPD	2018-2022
	Cerrado-1	6	T23KLB	2017-2022
Cerrado	Cerrado-2	9	T22KFE	2017-2022
	Cerrado-3	4	T22KFC	2018-2022
Mata Atlântica	Mata Atlântica-1	6	T22KHC	2018-2022
	Mata Atlântica-2	7	T22KEB	2018-2028
	Mata Atlântica-3	10	T22KEE	2017-2022

A aquisição desses dados ocorreu na plataforma Copernicus (<https://scihub.copernicus.eu/dhus/#/home>) em nível 2A, ou seja, corrigidos para radiometria e influências atmosféricas. O MSI possui nove bandas espectrais com resolução espacial de 20 m e quatro bandas com resolução de 10 m. No entanto, para este estudo, foram usadas apenas as bandas de 20 m, pois apresentaram melhor desempenho na modelagem da fração evapotranspirativa (DOS SANTOS *et al.*, 2022). As nove bandas espectrais abrangem os comprimentos de onda da região do visível (0.490 - 0.665 μm), borda do vermelho (0.705 – 0.783 μm), infravermelho médio (0.865 μm) e infravermelho de ondas curtas (1.610 - 2.190 μm). Os comprimentos de onda apresentam comportamento de reflectância diferentes de acordo com mudanças que ocorram na estrutura ou fisiologia da planta, desta forma são importantes variáveis que conseguem captar a sensibilidade dos vegetais quando estão submetidos a

determinado condições ambientais. Devido a isso, as nove bandas foram selecionadas como preditoras para treinamento dos algoritmos de *machine learning*.

Como as bandas possuem resolução espacial de 20 m, foi necessário compatibilizá-las espacialmente com os dados dos sensores no Landsat-8 e 9, que são de 30 m. Para isso, foi realizado o procedimento de agregação do *pixel* com fator de escala igual a 1.5 e função média. Posteriormente, aplicou-se o método *bilinear* no *resample*, utilizando como base as imagens dos sensores a bordo nos Landsats. Tanto na agregação quanto no *resample*, foi utilizado o pacote raster (HIJMANS *et al.*, 2020) desenvolvido para ambiente de linguagem R. Ao final do processo, as imagens do sensor MSI estavam compatibilizadas temporalmente (mesmo dia de aquisição) e espacialmente (mesma resolução espacial) com a ET_a calculada a partir dos sensores OLI e TIRS.

Com o objetivo de aumentar o número de preditoras entre as variáveis relacionadas com o parâmetro planta, foram aplicadas duas metodologias sobre as nove bandas espectrais que resultaram em produtos entre as bandas. A primeira foi a *normalized ratio procedure between bands* (NRPB), que gerou novas preditoras a partir da normalização entre as nove bandas. A segunda foi a *simple ratio procedure between bands* (SRPB), que produziu novas preditoras a partir da razão simples entre as bandas. A NRPB foi realizada utilizando a função `band_ratio` do pacote `labgeo` (FERNANDES FILHO 2019) para a linguagem R. Já para a SRPB, foi necessária alteração no código fonte da `band_ratio` do pacote `labgeo`. Ao final desses procedimentos, foram produzidas mais 72 variáveis, sendo 36 da NRPB e 36 da SRPB.

2.3.2 Nasa Power

As variáveis consideradas do produto Nasa Power foram radiação solar incidente (R_s , $\text{MJ m}^{-2} \text{ dia}^{-1}$), temperatura máxima do ar (T_{mx} , °C), temperatura mínima do ar (T_n , °C), temperatura média (T_{med} , °C), umidade relativa média (RH, %) e velocidade do vento médio (V_v , m s^{-1}). Os dados provenientes do Nasa Power têm como desvantagem o atraso de dois dias, ou seja, não é possível obter esses dados no dia corrente, o que dificultaria a aplicação dos modelos. Pensando nisso, esses dados foram adquiridos 2, 3, 4 e 5 dias antes da passagem dos satélites, obtendo assim 4 dados de R_s , 4 de T_{mx} , 4 de T_n , 4 de T_{med} , 4 de RH e 4 de V_v . Além disso, foram calculadas as médias e desvios padrão de cada variável, produzindo novas variáveis explicativas.

Os dados foram adquiridos em formato *data frame*, utilizando o pacote `nasapower` (SPARKS, 2018) em linguagem R, a partir das coordenadas do *pixel* do raster de ET_a , onde as máscaras para cana-de-açúcar (Figura 1) estão inseridas. Posteriormente, as variáveis foram

convertidas em raster para que todas as fontes de dados estivessem em um único formato. Optou-se também por calcular, utilizando essas variáveis meteorológicas do Nasa Power, a evapotranspiração de referência para grama (ET_o) e para alfafa (ET_r), extraindo a média e o desvio padrão. Com isso, ao final do processo, tinha-se 32 variáveis meteorológicas, mais 8 médias e 8 desvios padrão de cada uma delas, totalizando 48 variáveis.

2.3.3 Bioclimáticas

Foram selecionadas todas as 19 bioclimáticas disponíveis pela WorldClim (Tabela 4) e sua aquisição foi realizada diretamente na plataforma da WorldClim (<https://www.worldclim.org/data/worldclim21.html>). Esse tipo de dado é baseado em séries históricas mensais de temperatura e precipitação que, por meio de estatísticas, geram as bioclimáticas. Sua escolha se deu por ser importante na caracterização do clima, a qual tem impacto no processo evapotranspirativo. Por exemplo, uma região mais seca (bioma Caatinga) tem maior potencial evapotranspirativo do que uma região mais úmida (bioma Mata Atlântica).

Tabela 4. Bioclimáticas da base de dados globais do WorldClim.

Código	nomenclatura
bio1	Temperatura Média Anual
bio2	Intervalo Diurno Médio
bio 3	Isotermalidade
bio 4	Sazonalidade da Temperatura
bio 5	Temperatura máxima do mês mais quente
bio 6	Temperatura mínima do mês mais frio
bio 7	Faixa Anual de Temperatura
bio 8	Temperatura média do trimestre mais úmido
bio 9	Temperatura média do trimestre mais seco
bio 10	Temperatura Média do Trimestre Mais Quente
bio 11	Temperatura média do trimestre mais frio
bio 12	Precipitação Anual
bio 13	Precipitação do mês mais úmido
bio 14	Precipitação do Mês Mais Seco
bio 15	Sazonalidade de Precipitação
bio 16	Precipitação do trimestre mais úmido
bio 17	Precipitação do trimestre mais seco
bio 18	Precipitação do Quarto Mais Quente
bio 19	Precipitação do trimestre mais frio

As bioclimáticas foram adquiridas em formato de imagem tipo raster, assim como os dados do MSI, OLI e TIRS, porém com resolução espacial de 30 segundos, o que equivale a aproximadamente 925 m. Devido a essa baixa resolução espacial, foi necessário aplicar o processo inverso daquele aplicado no MSI, ou seja, a desagregação de 30 segundos para a resolução espacial de 30 m. Para isso, utilizou-se o mesmo pacote raster com a função

disaggregate, aplicando o fator de escala igual a 30.86 e a função de média. Isso permitiu aproximar os valores a 30 m. Em seguida, aplicou-se a função *resample* utilizando como base o raster de ET_a , resultando, ao final, em informações de bioclimáticas compatíveis com a resolução espacial dos demais dados.

2.3.4 Latitude, elevação, mês, dia do ano e astronômicas

A variável latitude foi extraída de cada *pixel* dos dados raster em formato decimal. Para a elevação, foi necessária a aquisição, na plataforma USGS, do modelo digital de elevação (MDE) em resolução de 30 m dos dados *Shuttle Radar Topography Mission* (SRTM), compatível com a resolução dos demais dados, em formato raster. Os dados do mês e dia do ano foram extraídos a partir da data de cada imagem, separando o mês e alocando-o a um *data frame* e convertendo a data em dia do ano para também alocá-lo.

Foram produzidas, por fim, cinco variáveis astronômicas, sendo a declinação solar (em graus decimais), o ângulo horário do nascer do sol (em radianos), o fotoperíodo (em horas), a distância Terra-Sol (em radianos) e a radiação global extraterrestre (em $MJ\ m^{-2}\ d^{-1}$). As equações para cada uma delas estão disponíveis em Brock (1981).

2.4 Preparo dos dados para treinamento e teste dos algoritmos

Para treinar e testar os algoritmos de *machine learning* utilizando dados em formato raster, foi necessário extrair seus valores para o formato tipo *data frame*, que é composto por linhas e colunas. Em cada coluna, tem-se a variável explicativa e, em cada linha, os valores correspondentes a essas variáveis.

2.4.1 Extração e seleção dos dados

A extração dos valores raster para armazená-los em um formato *data frame* foi realizada em dois grupos de imagens: o grupo 1 continha as imagens destinadas ao treinamento e o grupo 2 continha as imagens para o teste. O grupo 2 continha uma data selecionada aleatoriamente de cada área da Tabela 3, ou seja, 3 datas para o bioma Mata Atlântica, 3 para o Cerrado e 1 para a Caatinga, enquanto as imagens do grupo 1 continham o restante delas.

Para o grupo 1, extraiu-se aleatoriamente 175000 dados para cada bioma, totalizando 525000 dados. Ao finalizar a extração, obteve-se um *data frame* contendo 9 colunas com dados espectrais, 36 colunas com NRPB, 36 colunas com SRPB, 48 colunas com dados do Nasa Power, 19 colunas com as bioclimáticas, mais 4 colunas contendo dia do ano, mês, latitude e elevação, 5 colunas com as astronômicas e, por fim, 1 com a ET_a . Optou-se por extrair esse comprimento de dados devido ter sido utilizado por Dos Santos *et al.* (2022) na predição da

fração evapotranspirativa, com resultados satisfatórios. Já para o grupo 2, o procedimento de obtenção dos dados ocorreu de modo semelhante, com a diferença de que aqui foram selecionados 75000 dados, mas produzindo o mesmo número de colunas com suas respectivas variáveis. Assim, dois arquivos em formato *data frame* foram produzidos, sendo um para o treino e outro para o teste.

2.4.2 Redução de preditoras autocorrelacionadas e de baixa importância

Observando as boas práticas de programação e buscando simplificar a análise, foram aplicadas duas metodologias para reduzir as variáveis preditoras. A primeira foi a remoção de variáveis com alta correlação, visando reduzir o excesso de variáveis e melhorar a capacidade de processamento dos dados. Variáveis explicativas com correlação acima de 96% entre si foram removidas para diminuir a redundância de informações. Em seguida, foi realizada a remoção de variáveis de menor importância utilizando o algoritmo *Recursive Feature Elimination* (RFE) com a função *Random Forest* do pacote *caret* (KUHNS *et al.*, 2020) em linguagem R. O algoritmo constrói um modelo base utilizando todas as preditoras que restaram após a remoção da autocorrelação, calcula o peso que cada preditora tem sobre esse modelo e identifica as preditoras que não trazem grandes ganhos ao modelo, removendo-as.

2.4.3 Treinamento e teste

O grupo de dados destinado ao treino após o processo de filtragem foi inserido em 13 algoritmos de regressão, sendo eles: *linear regression* (LM), *cubist*, *extreme gradient boosting* - método linear (XgbLinear), *extreme gradient boosting* - método tree (XgbTree), *principal components regression* (pca), *recursive partitioning* (rpart), *bayesian regularized neural networks* (brnn), *partial least squares* (pls), *ridge regression* (ridge), *generalized linear model* (glm), *lasso*, *stochastic gradient boosting* (gbm) e *k-nearest neighbor* (knn). Esses modelos foram escolhidos por apresentarem elevado potencial preditivo e serem mais rápidos no treinamento, o qual foi realizado com auxílio do pacote *caret* (KUHNS *et al.*, 2020) em linguagem R.

Para avaliar o desempenho de cada algoritmo após o treino, utilizou-se o segundo grupo de conjunto de dados para realizar a predição e aplicação de estatísticas entre os valores observados e os valores preditos. As análises estatísticas aplicadas foram o coeficiente de determinação (R^2), Eq. (8), raiz do erro quadrático médio (RMSE), Eq. (9), raiz do erro quadrático médio relativo (RRMSE), Eq. (10), erro médio absoluto (MAE), Eq. (11), erro médio absoluto relativo (MAPE), Eq. (12), e erro médio de viés (MBE), Eq. (13).

$$R^2 = \frac{[\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})]^2}{[\sum_{i=1}^n (P_i - \bar{P})^2][\sum_{i=1}^n (O_i - \bar{O})^2]} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (9)$$

$$RRMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i)^2}} \times 100 \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (11)$$

$$RMAE = \frac{\frac{1}{n} \sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n (O_i)} \times 100 \quad (12)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (13)$$

onde P_i é o valor previsto pelo modelo; O_i é o valor observado; \bar{O} é o valor médio observado e n é o número de pares de dado.

2.5 Valores SHAP (explicação do aditivo Shapley)

Os modelos de *machine learning* são conhecidos por serem caixas pretas, ou seja, é possível compreender suas entradas e saídas, mas muitas vezes é difícil entender o que ocorre em seu interior, o que dificulta sua interpretabilidade. Para solucionar esse problema, Lundberg e Lee (2017) desenvolveram uma metodologia baseada na teoria dos jogos que permite compreender o comportamento de cada variável dentro dos modelos de *machine learning*. Essa metodologia é conhecida como SHAP (*Shapley Additive exPlanations*), que é um método aditivo de atribuição de características. Em outras palavras, a saída do modelo é uma adição linear das variáveis de entrada (FENG *et al.*, 2021).

Para aplicar a metodologia SHAP em linguagem R, foi necessário utilizar dois pacotes: o fastshap (GREENWELL, 2021) e o shapviz (MAYER, 2023). O primeiro pacote calcula o valor SHAP de cada variável explicativa, e para isso utilizou-se 30% dos conjuntos de dados do teste devido ao seu processo ser lento e moroso. Posteriormente, utilizou-se o segundo pacote para gerar gráficos dos valores SHAP e dos valores absolutos médios para cada variável.

2.6 Aplicação

Com o objetivo de verificar a integração das plataformas Landsat e Sentinel na estimativa da ET_a na cana-de-açúcar ao longo do ciclo, realizou-se a aplicação em uma área de aproximadamente 400 hectares localizada na parte noroeste da região Mata Atlântica-3 (conforme apresentado na Figura 1). Para isso, foram adquiridas todas as imagens úteis disponíveis ao longo do ciclo da safra 2021/2022 dos sensores OLI e TIR embarcados nos

satélites Landsat-8 e Landsat-9, e do MSI embarcado nos satélites Sentinel-2A e 2B, além de outras variáveis explicativas. No total, foram adquiridas 8 imagens do Landsat-9, 7 do Landsat-8 e 28 dos dois Sentinels (2A e 2B), totalizando 43 imagens ao longo do ciclo. As imagens dos Landsat-8 e 9 foram destinadas ao modelo METRIC, enquanto as imagens dos Sentinel, juntamente com as demais variáveis explicativas, foram destinadas ao modelo de *machine learning* com melhor performance. É importante ressaltar que apenas imagens sem nuvens foram utilizadas.

3 Resultados e Discussão

3.1 Seleção das variáveis e desempenho dos modelos

Ao gerar a matriz de correlação nas 157 variáveis e eliminar aquelas com correlação superior a 0.96 restaram, então, 110. Na Figura 4 observa-se as 110 preditoras que foram submetidas ao *recursive feature elimination* (RFE).

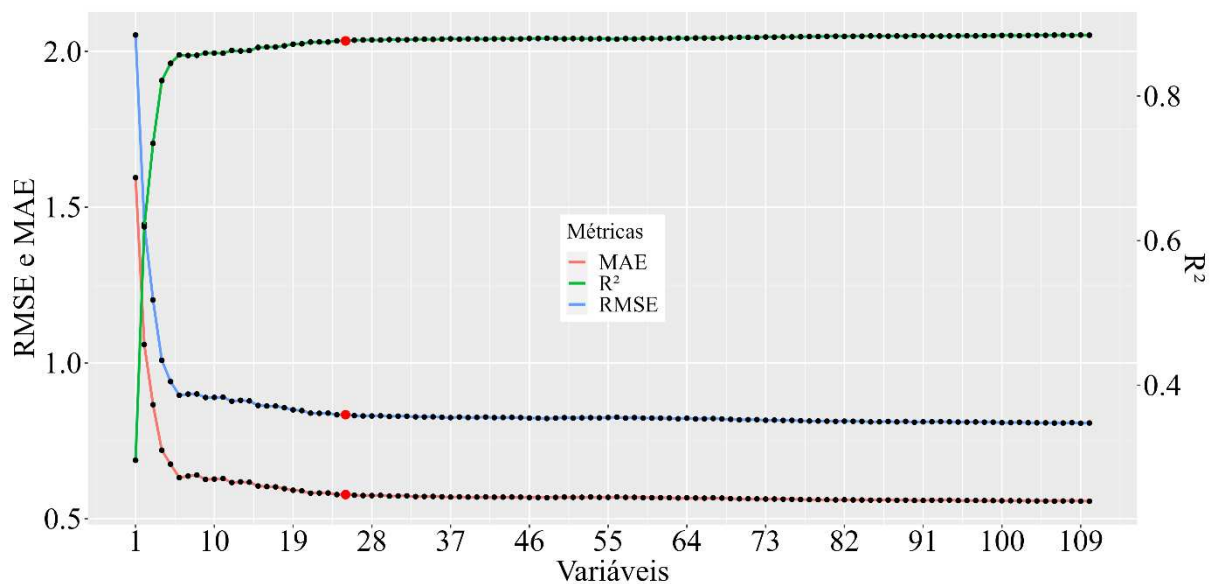


Figura 4. Seleção dos recursos preditivos para ET_a com base no Random Forest

Nota-se, na Figura 4, que as métricas RMSE e MAE apresentam altos valores com um número ínfimo de preditoras, enquanto o R^2 apresenta baixos valores. No entanto, à medida que o número de *features* é aumentado, essas métricas melhoram, reduzindo os valores do RMSE e MAE e aumentando o valor do R^2 até alcançar seu melhor desempenho, após o qual permanecem próximas com mínimas oscilações. Quando adicionar mais variáveis não produz ou produz muito pouca melhoria nas métricas, pode-se selecionar o número ideal de preditoras para prosseguir com o treinamento dos algoritmos de *machine learning*. Observa-se na Figura 4 que o número ideal selecionado foram 25 variáveis (conforme Tabela 5), um valor

substancialmente inferior às 157 produzidas inicialmente. Essa redução tem duas grandes vantagens, sendo a primeira a otimização computacional e a outra a simplificação da aplicação. Em função dessas vantagens, a técnica de remoção de variáveis com pouca ou nenhuma importância para os modelos é amplamente aplicada na ciência de dados em diversas áreas, principalmente em sensoriamento remoto e evapotranspiração (DIAS *et al.*, 2021; DOS SANTOS *et al.*, 2022b; FILGUEIRAS *et al.*, 2019b; IMPOLLONIA *et al.*, 2022; STAPLETON; EICHELMANN; ROANTREE, 2022).

Tabela 5. Variáveis selecionadas a partir do RFE para treinar os modelos

Nº	preditoras	Nº	Preditoras	Nº	preditoras
1	b3	11	b4/b3	21	média_tmédia
2	b4	12	b7/b6	22	dda
3	b8	13	b8/b6	23	bio02
4	b11	14	b11/b6	24	bio04
5	b12	15	b11/b7	25	bio15
6	$(b8-b6)/(b8+b6)$	16	b11/b8		
7	$(b11-b6)/(b11+b6)$	17	b12/b2		
8	$(b12-b2)/(b12+b2)$	18	b12/b3		
9	$(b12-b3)/(b12+b3)$	19	b12/b11		
10	$(b12+b11)/(b12+b11)$	20	ângulo_horário_nascer_sol		

média_tmédia: média da temperatura média; dda: dia do ano.

A Tabela 5 apresenta uma variedade de variáveis com origens distintas, provenientes de sensores orbitais, astronômicos, climáticos e informações de posição em relação ao ano. As variáveis dos sensores orbitais possuem uma dinâmica espacial e temporal, permitindo uma descrição precisa das condições na lavoura de cana-de-açúcar, tanto em termos de respostas espectrais (bandas individuais) quanto normalizadas e em razão simples. As variáveis astronômicas também são dinâmicas no espaço e tempo, levando em consideração a latitude e o dia do ano, caracterizando a quantidade de energia que pode chegar em uma determinada área ao longo do ano. As variáveis climáticas são dinâmicas no espaço e tempo, como a média da temperatura média (média_tmédia), e estáticas no tempo, mas dinâmicas no espaço (bio02, bio04, bio15). Por fim, as variáveis de posição em relação ao ano são dinâmicas apenas no tempo. É perceptível que há grande variação nos tipos e origens das variáveis selecionadas, mas todas contribuem para a estimativa da evapotranspiração. Esse conjunto de dados foi utilizado para treinar e testar os modelos de *machine learning*.

Na Figura 5, são apresentadas as métricas estatísticas utilizadas para avaliar o desempenho de cada algoritmo de *machine learning* após o treino e teste. Os resultados dos testes de cada modelo estão organizados em um ranking para facilitar a compreensão dos melhores e piores resultados.

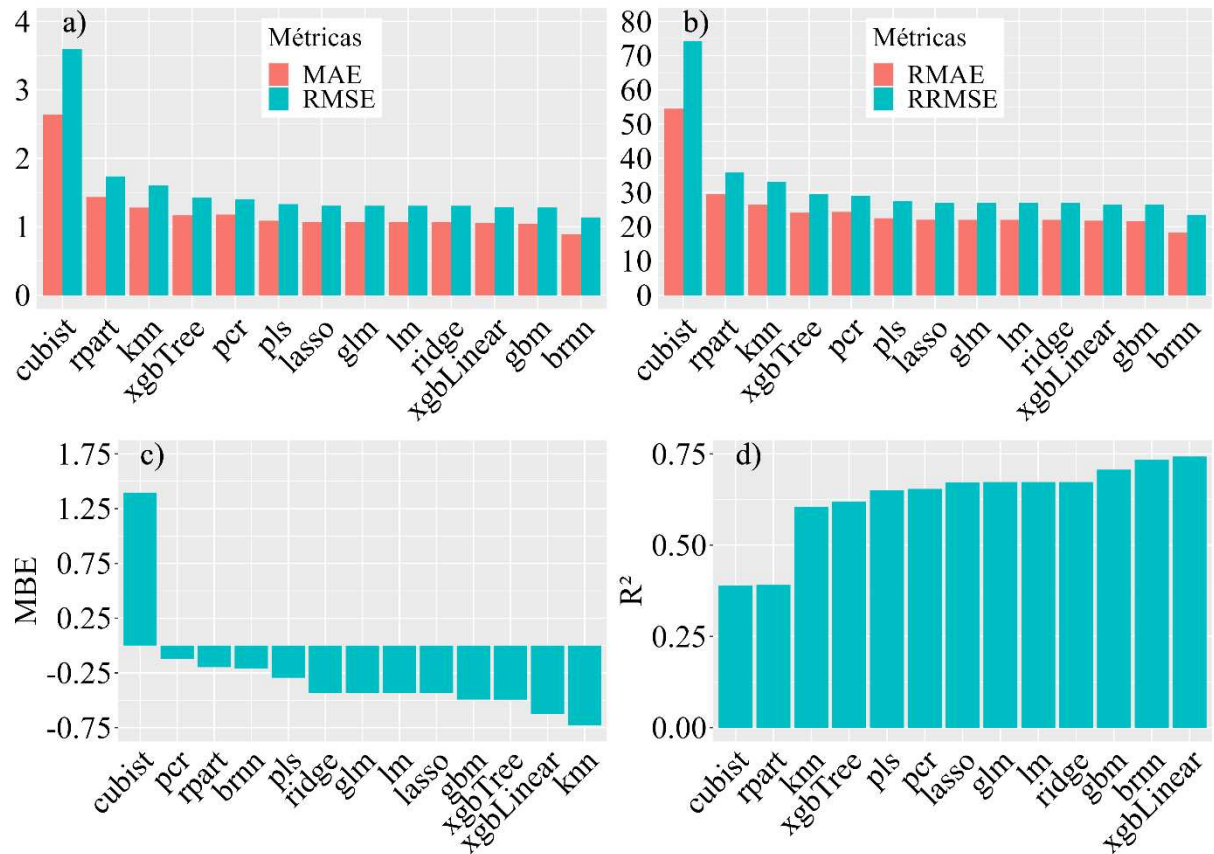


Figura 5. Desempenho estatístico do conjunto de teste para os 13 modelos de *machine learning* utilizando no estudo.

Percebe-se claramente, analisando a Figura 5, que o cubist teve o pior desempenho em todas as métricas entre os 13 algoritmos, seguido do rpart e knn, onde eles obtiveram valores de RMSE acima de 1.5 mm e RRMSE acima de 30%. Valores acima de 30% de RRMSE indicam que as previsões da evapotranspiração atual são imprecisas, tornando inadequado o uso desses modelos devido ao alto erro. Por outro lado, os outros 10 modelos apresentaram RMSE inferior a 1.5 mm e RRMSE inferior a 30%, tendo desempenho razoável quando seus valores estão em um intervalo de 20 a 30%. Os menores valores dessas métricas foram identificados em três modelos: brnn, gbm e XgbLinear, sendo, respectivamente, 23.48, 26.46 e 26.52%. Ao analisar a capacidade dos modelos em subestimar ou superestimar a evapotranspiração através do MBE (Figura 5c), constata-se que apenas o cubist superestimou atingindo, em média, 1.25 mm, enquanto todos os outros subestimaram em um valor médio máximo de 0.75 mm. Para o

R^2 (Figura 5d), dois modelos se destacaram como superiores: o XgbLinear com valor igual a 0.74 e o brnn com valor igual a 0.73. Estimar a evapotranspiração não é uma tarefa trivial, pois envolve diversas variáveis que a modelam, além das complexidades das áreas de coleta. Os melhores resultados encontrados neste estudo foram R^2 de 0.74 e 0.73, mas essa performance não é restrita a este estudo. Pesquisas realizadas por Ke *et al.* (2017) para a área agrícola encontraram R^2 de 0.72, enquanto Bunting *et al.* (2014), utilizando dados MODIS e regressão linear múltipla, obtiveram R^2 de 0.74 e, por fim, Nagler *et al.* (2013), também utilizando MODIS na predição da ET_a , encontraram R^2 igual a 0.74.

O brnn e o XgbLinear foram os dois modelos selecionados para avaliação em cada localidade de coleta de dados, permitindo uma avaliação micro após a avaliação macro inicial. Esses modelos se destacaram em relação aos outros 11 algoritmos avaliados. O brnn é um modelo de rede neural que incorpora o teorema de Bayes juntamente com um esquema de regularização para fornecer estimativas de incerteza quantificáveis e reduzir o *overfitting*, melhorando a capacidade de generalização (HUANG *et al.*, 2023). Este algoritmo foi aplicado com sucesso em parâmetros biofísicos de índice de área foliar, umidade da bainha foliar, teor de clorofila e concentração de nitrogênio foliar em cana-de-açúcar usando dados espectrais do Sentinel-2 (HAJEB *et al.*, 2023). Por outro lado, o XgbLinear é uma implementação de modelo linear dentro da estrutura do XGBoost que combina modelos lineares generalizados e árvores de decisão (CHEN; GUESTRIN, 2016). Ele realiza o pré-processamento do conjunto de treinamento, ajusta modelos lineares e utiliza árvores de decisão para corrigir resíduos, com a saída final sendo a soma dos modelos lineares e previsões das árvores de decisão. De acordo com Dos Santos *et al.* (2022), o XgbLinear apresentou boa performance na predição da fração evapotranspirativa em lavoura de cana-de-açúcar irrigada.

3.2 Desempenho do brnn e XgbLinear nas áreas de estudo

3.2.1 Desempenho geral

A dispersão dos dados entre os valores observados e estimados podem ser visualizados na Figura 6.

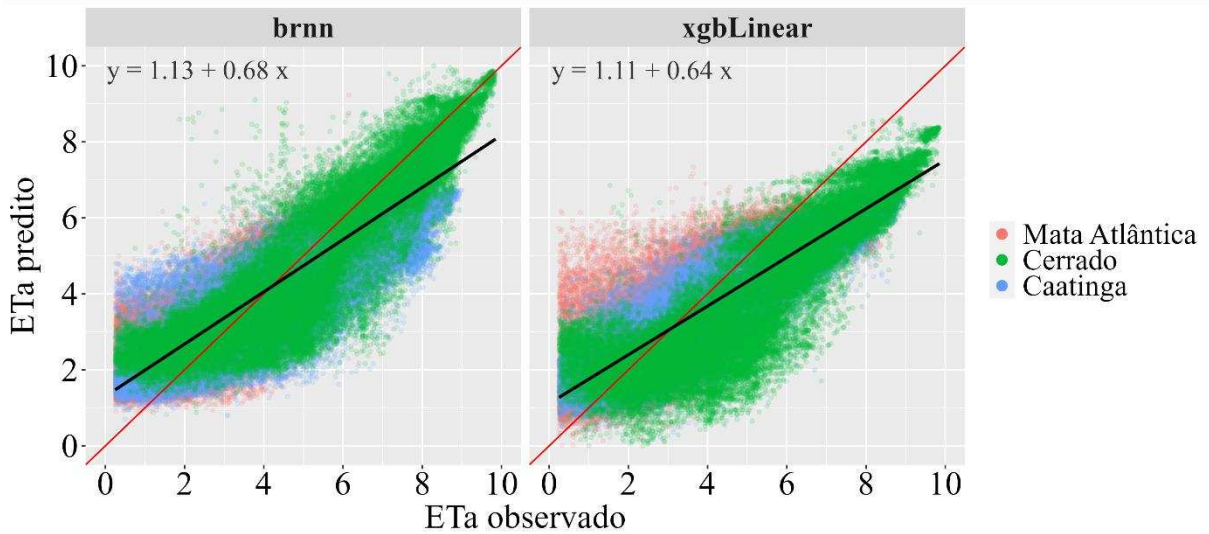


Figura 6. Dispersão entre valores predito e observado para os algoritmos brnn e XgbLinear no conjunto geral de dados destinado ao teste.

Ao analisar cuidadosamente a Figura 6, observa-se que os valores máximos observados atingem cerca de 10 mm e o mínimo próximo de 1 mm. O modelo brnn mostra ter capacidade de explicar a variável resposta tanto em baixos valores, próximos a 1, quanto em altos valores, próximos a 10 mm. Para esse modelo, a reta de tendência apresenta baixa superestimativa em valores inferiores a 1 mm e subestimativa para ET_a superior a 4.0 mm, tendo um maior número de amostras subestimando os valores observados, o que é coerente com o apresentado na Figura 5c. Ao analisar o modelo XgbLinear, fica evidente que ele não foi capaz de prever valores próximos a 10 mm, pois a predição máxima está em torno de 8.0 mm, ou seja, uma subestimativa maior do que a subestimativa do brnn. Na Figura 5c, isso fica evidente, pois o brnn subestimou, em média, 0.21 mm e o XgbLinear, em média, 0.62 mm, fato também notado pelas proximidades entre a reta de tendência (reta preta) e a reta 1:1 (reta vermelha) em ambos os algoritmos. Outro fato que chama a atenção na Figura 6 é a alta dispersão dos dados quando a ET_a observada é menor que 4 mm para ambos os algoritmos, principalmente nas amostras coletadas nas áreas de cana-de-açúcar localizadas nos biomas Mata Atlântica e Cerrado. Essa alta dispersão em valores baixos pode estar relacionada com o menor índice de área foliar no início do ciclo da cultura, onde há parte de solo exposto e/ou restos de palhada da última colheita, fazendo com que a área apresente maior heterogeneidade e, assim, os modelos não conseguem captar com boa performance a ET_a . Dos Santos *et al.* (2022) observaram, para cana-de-açúcar em região semiárida, que valores de fração evapotranspirativa, estimado por *machine learning*, abaixo de 0.6 possuem alta dispersão nos dados, de modo igual KE *et al.* (2017) também perceberam alta dispersão entre o predito e o observado da evapotranspiração em áreas agrícolas, com justificativa da heterogeneidade nessas áreas.

O comportamento das principais variáveis apresentadas na Tabela 5 para os algoritmos brnn e XgbLinear pode ser visualizado na Figura 7. Nela, estão os valores SHAP, que permitem conhecer a forma como os algoritmos entendem cada variável que explica a ET_a na cana-de-açúcar.

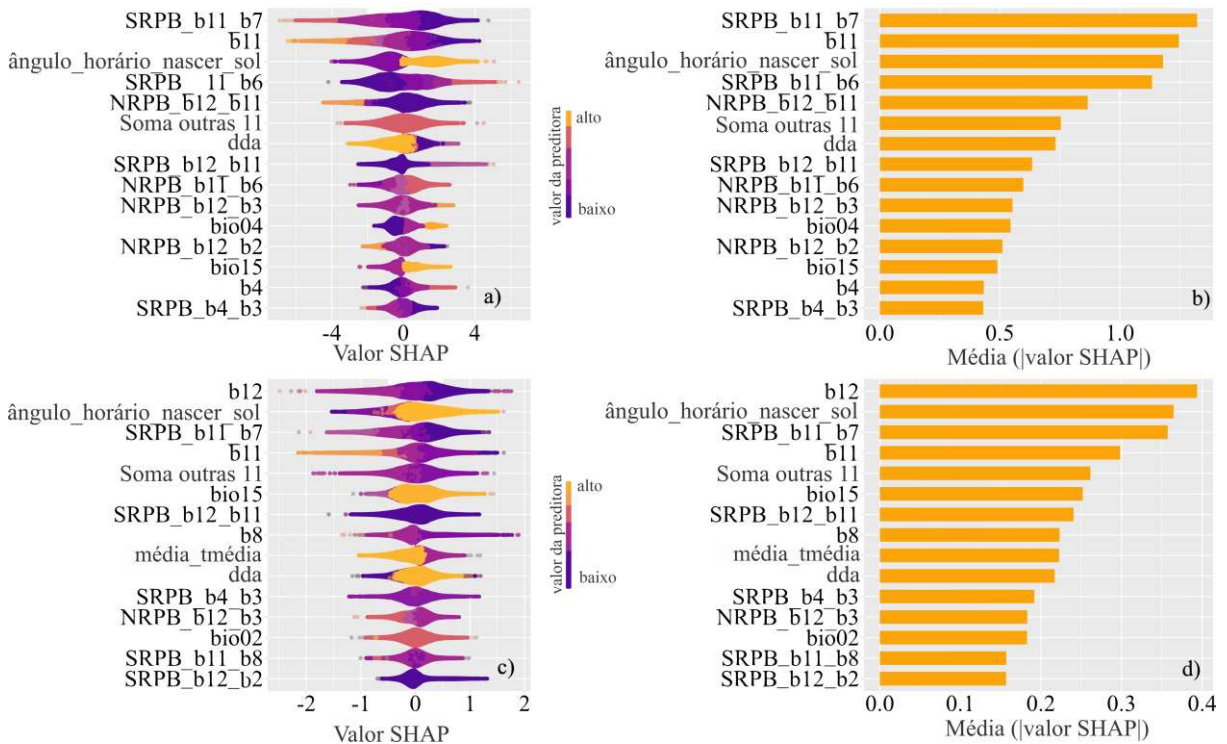


Figura 7. Interpretação global dos modelos brnn (a) e XgbLinear (c) com os valores SHAP e seus respectivos valores absolutos médios (b e d).

O entendimento geral dos valores SHAP da Figura 7 é que os valores à direita representam observações que aumentam o valor predito, enquanto os valores à esquerda representam observações que diminuem o valor predito. Ao analisar cada preditora para o brnn (Figura 7a e 7b), percebe-se que a SRPB_b11_b7 e b11 têm uma grande influência na evapotranspiração e são semelhantes entre si. Assim, quando a reflectância do comprimento de onda 1610 nm (b11) apresentar baixos valores e a razão simples entre essa reflectância com a reflectância do comprimento de onda 783 nm (b7) também apresentar baixos valores, o brnn entende que terá um efeito positivo na evapotranspiração. À medida que seus valores aumentam, o efeito se torna negativo. O comprimento de onda 1610 nm está na faixa do infravermelho de ondas curtas (SWIR) e uma de suas características é a absorção pela água, portanto, quando a folha está bem hidratada, refletirá menos essa faixa de comprimento de onda, e quando estiver menos hidratada, ou seja, sob estresse hídrico, sua reflectância aumentará. Chen *et al.* (2012) analisaram a reflectância hiperespectral em diferentes estresses hídricos na cana-de-açúcar e constataram que a reflectância aumenta do estresse leve ao estresse severo,

com destaque para o comprimento 1610 nm, que na condição leve a reflectância foi de 0.10 e na severa de 0.20, ou seja, a reflectância dobrou em condições de estresse severo. Já Yang *et al.* (2022) em seus estudos identificaram que o comprimento de onda de 1610 nm foi sensível a mudanças de temperatura foliar sob estresse no algodoeiro, onde, à medida que se aumentava a temperatura foliar, também aumentava a reflectância desse comprimento de onda. Assim, quando ocorre maiores reflectâncias, a evapotranspiração é reduzida, e essa variável impacta negativamente na resposta preditiva no modelo brnn, como observado na Figura 7a. O comprimento de onda de 783 nm é considerado borda do vermelho, mas ele está mais próximo da faixa do infravermelho próximo. Por causa disso, é possível que seu comportamento se assemelhe ao infravermelho próximo, que está relacionado com a estrutura celular. Assim, quanto melhor a estrutura celular, maior será a reflectância desse comprimento de onda (Chen *et al.*, 2012). Em situações de estresse ou conforto hídrico, a estrutura celular é alterada para adaptação a essa nova condição (ILYAS *et al.*, 2021), e, dessa forma, a reflectância do infravermelho próximo será maior ou menor. Quando a célula está bem hidratada, logo com boa estrutura, a reflectância será maior, e o contrário é verdadeiro. Com isso, é compreensível que a razão entre as bandas b11 e b7 seja menor quando a planta estiver em conforto hídrico e maior quando estiver sob condições de estresse.

O ângulo horário do nascer do sol e a razão simples entre a reflectância da banda b11 com a reflectância do comprimento de onda de 740 nm (b6) têm influência completamente inversa em relação às variáveis discutidas anteriormente. Em outras palavras, quando essas variáveis apresentam alto valores, terá impacto positivo no *output* de evapotranspiração, e quando apresentarem menores valores, o impacto será negativo. O ângulo horário do nascer do sol representa a posição do sol em relação ao plano horizontal e é influenciado pela latitude e pela declinação solar (BUSHNELL, 1982). Ele tem relação com a quantidade de energia que chega até a superfície terrestre. No hemisfério sul, quanto maior o ângulo horário do nascer do sol, em radianos, maior será essa energia. Logo, mais energia estará disponível para o processo evapotranspirativo. A razão entre b11 e b6 segue o mesmo entendimento que a razão entre b11 e b7, com a diferença de que a b6 está tendo maior influência nessa equação, e ela é um forte indicativo de estresse hídrico. A b6 corresponde à região espectral da borda do vermelho, tida como importante região no que diz respeito ao conteúdo de água nas plantas. Sua reflectância tende a ser menor quando hidratada, ou seja, ocorre maior absorção desse comprimento de onda pelas folhas. Estudos têm utilizado a região da borda do vermelho para identificar o impacto do estresse hídrico, como mostram Sun *et al.* (2020) que observaram o potencial dessa região do comprimento de onda no estresse hídrico na cultura do arroz, assim como Zhang e Zhou (2019)

e Ren *et al.* (2022), que constataram a sensibilidade da borda do vermelho ao conteúdo de água na cultura do milho.

As variáveis SRPB_b11_b7, b11, ângulo horário do nascer do sol e SRPB_b11_b6 apresentam maior impacto no modelo, como pode ser observado na Figura 7b, onde a média dos valores absolutos do SHAP é maior para elas. As demais variáveis, mesmo tendo menor influência, são necessárias para o ajuste adequado do modelo. Assim, as variáveis que apresentam impacto negativo na predição da evapotranspiração quando estão em valores elevados são NRPB_b12_b11, dda, NRPB_b12_b2 e SPRB_b4_b3. Enquanto as variáveis SRPB_b12_b11, NRPB_b11_b6, NRPB_b12_b13, bio04, bio15 e b4 influenciam positivamente em valores elevados.

No modelo XgbLinear (Figura 7d e 7e), é possível observar que quatro variáveis se destacam em relação às demais, sendo que três delas (b12, SRPB_b11_b7 e b11) deslocam o valor predito para a esquerda quando seus valores são altos e para a direita quando são baixos, e uma delas (ângulo horário de nascer do sol) tem influência negativa na predição da evapotranspiração quando seus valores são baixos e positiva em altos valores. A única diferença em relação ao brnn é a presença da b12 entre essas quatro variáveis, que corresponde ao comprimento de onda de 2190 nm. No entanto, sua compreensão é a mesma da b11, pois é um infravermelho de ondas curtas absorvido pelo conteúdo de água. Desta forma, em condições de menor capacidade transpirativa das plantas, esse comprimento de onda será mais refletido e, ao contrário, mais absorvido, o que significa que a reflectância desse comprimento de onda apresenta baixos valores quando há restrição hídrica, e valores mais altos quando as plantas estão hidratadas. Portanto, o XgbLinear entende que baixos valores de reflectância de b12 terão um efeito positivo na predição da evapotranspiração, enquanto altos valores terão um impacto negativo.

As outras variáveis apresentaram menor impacto na predição. As bandas b8, média_tmédia e NRPB_b12_b3 geram influência negativa na evapotranspiração em altos valores, enquanto as bio15, dda e bio2 possuem comportamento contrário. As variáveis SRPB_b12_b11, SRPB_b4_b3, SRPB_b11_b8 e SRPB_b12_b2 chamam atenção por apresentarem uma influência menos clara na predição da evapotranspiração. Em outras palavras, tanto altos quanto baixos valores dessas variáveis geram um mesmo comportamento, o que pode aumentar o erro na estimativa da variável resposta. É importante destacar que essas quatro variáveis são derivadas da razão simples entre bandas, o que requer mais atenção ao utilizá-las no modelo XgbLinear.

3.2.2 Desempenho por biomas e áreas de coletas

Na Figura 8 é possível observar a dispersão dos dados observados e preditos para cada bioma onde concentra as maiores áreas cultivadas com cana-de-açúcar do país.

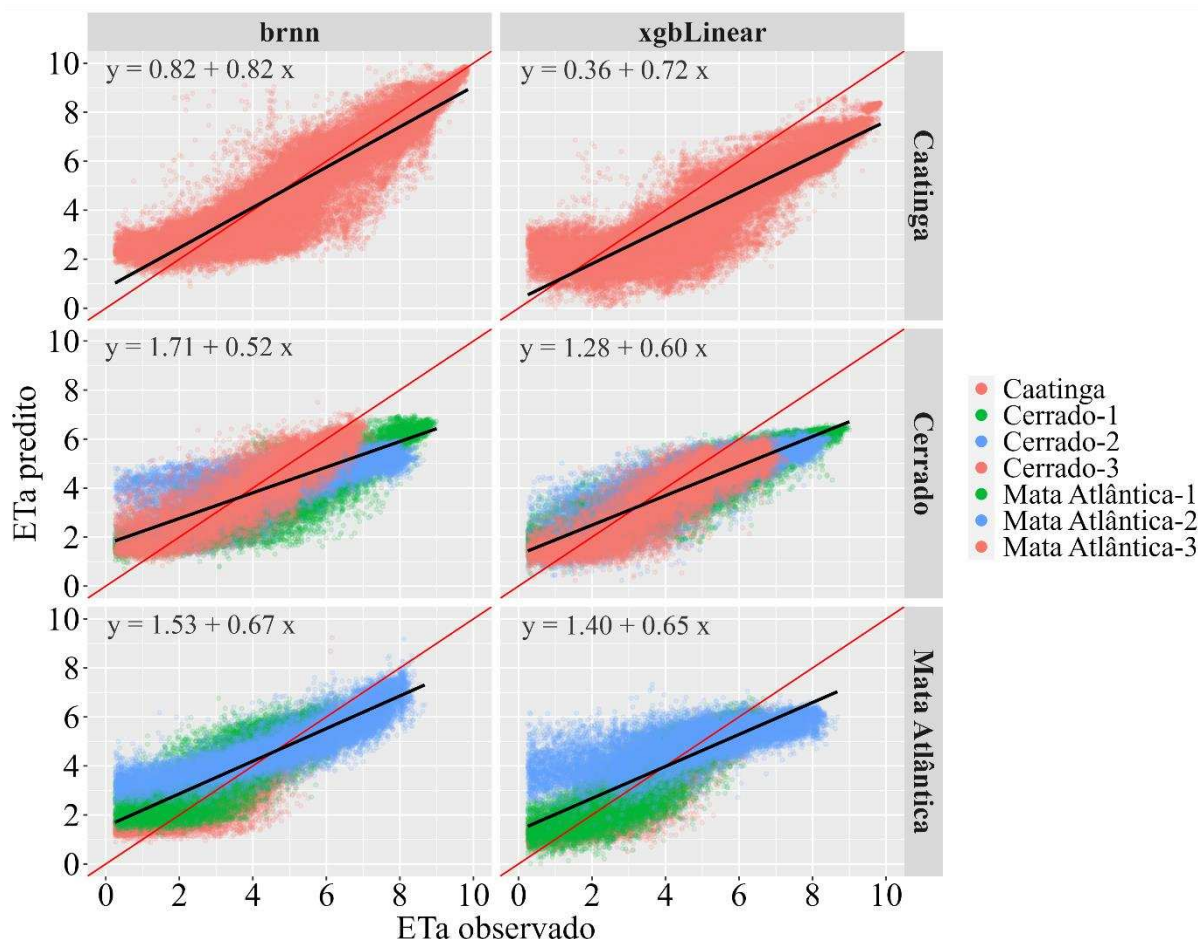


Figura 8. Dispersão entre valores predito e observado para os algoritmos brnn e XgbLinear para o conjunto dados em cada bioma.

É perceptível que os altos valores de evapotranspiração estão concentrados no bioma Caatinga, onde os valores observados atingiram cerca de 10 mm. O segundo bioma que apresentou maiores valores de evapotranspiração foi o Cerrado e, por último, a Mata Atlântica. O bioma Caatinga, em comparação aos outros dois, é caracterizado por altas insolações, temperaturas elevadas e baixa umidade (COUTINHO, 2016), o que gera grande quantidade de energia disponível para o processo evapotranspirativo, tornando as áreas agrícolas, principalmente as irrigadas, com maior capacidade de evapotranspiração. Já o Cerrado também é propício para elevadas evapotranspirações, no entanto, suas condições climáticas são menos severas do que as apresentadas no bioma Caatinga (COUTINHO, 2016). Por fim, a Mata Atlântica é o bioma, entre os dois outros discutidos, que apresenta condições mais equilibradas

entre radiação, temperatura e umidade relativa do ar (COUTINHO, 2016), resultando em menos energia disponível para o processo evapotranspirativo em comparação aos demais.

Analisando as retas de tendência da Figura 8 e a Tabela 6, fica claro que o brnn superestimou apenas nas áreas da Mata Atlântica, enquanto para a Caatinga e o Cerrado ele subestimou, sendo neste último com maior erro. Já o XgbLinear subestimou em todos os biomas, destacando-se a Caatinga com valor médio de -1.11 mm.

Tabela 6. Estatística do conjunto de dados destinado ao teste para brnn e XgbLinear nos três biomas

Bioma	\bar{x} obs mm	\bar{x} pred mm	dif %	R^2	MAE mm	RMAE %	RMSE mm	RRMSE %	MBE mm
brnn									
Mata Atlântica	4.26	4.37	2.58	0.76	0.76	17.83	0.98	22.70	0.10
Caatinga	5.40	5.27	2.41	0.82	0.74	14.10	0.99	18.48	-0.14
Cerrado	4.85	4.25	12.37	0.66	1.14	23.44	1.40	28.71	-0.60
XgbLinear									
Mata Atlântica	4.26	4.15	2.58	0.70	0.79	18.70	1.05	24.71	-0.11
Caatinga	5.40	4.29	20.56	0.80	1.33	24.65	1.53	28.35	-1.11
Cerrado	4.85	4.20	13.40	0.80	1.03	21.26	1.21	25.12	-0.64

Ainda na Tabela 6, observa-se que o brnn obteve desempenho superior para as áreas de cultivo nos biomas Mata Atlântica e Caatinga, com R^2 iguais a 0.76 e 0.82, respectivamente, e RMSE de 0.98 e 0.99, nessa ordem. A diferença entre a média dos valores observados e preditos foi de 2.58% para as áreas no bioma Mata Atlântica e 2.41% para a Caatinga. O XgbLinear foi melhor para as lavouras no Cerrado, onde o R^2 foi de 0.80 e o brnn, para esse bioma, de 0.60, enquanto o RMSE foi de 1.21 e 1.40, respectivamente. No entanto, a diferença da média entre os valores observados e preditos foi melhor para o brnn com 12.37%, mas próxima dos 13.40% do XgbLinear. Assim, observa-se que cada modelo performou de maneira diferente para cada localidade, e isso está relacionado com suas arquiteturas. A característica linear e resíduos corrigidos com árvore de decisão do XgbLinear foi favorável para melhor desempenho no Cerrado, equilibrando o erro entre todos os conjuntos de dados, enquanto a regularização bayesiana possibilitou melhor ajuste nas preditoras para favorecer na Mata Atlântica e Caatinga. Esse fato pode ser observado pelo R^2 que para o brnn nos biomas Mata Atlântica, Caatinga e Cerrado foram 0.76, 0.82 e 0.66, nessa ordem, enquanto para o XgbLinear foi de 0.70, 0.80 e 0.80, respectivamente. Ou seja, houve redução na Mata Atlântica e na Caatinga e aumento no Cerrado, mantendo as métricas mais equilibradas, algo que não aconteceu com o brnn.

Ao analisar a Figura 9, fica evidente que a área de coleta 3 (Mata Atlântica-3) foi responsável pelo maior erro de predição geral nesse bioma, pois se observa maiores super e

subestimativas para os modelos, ao passo que nas demais áreas existem maior proximidade entre a reta de tendência e a reta 1:1.

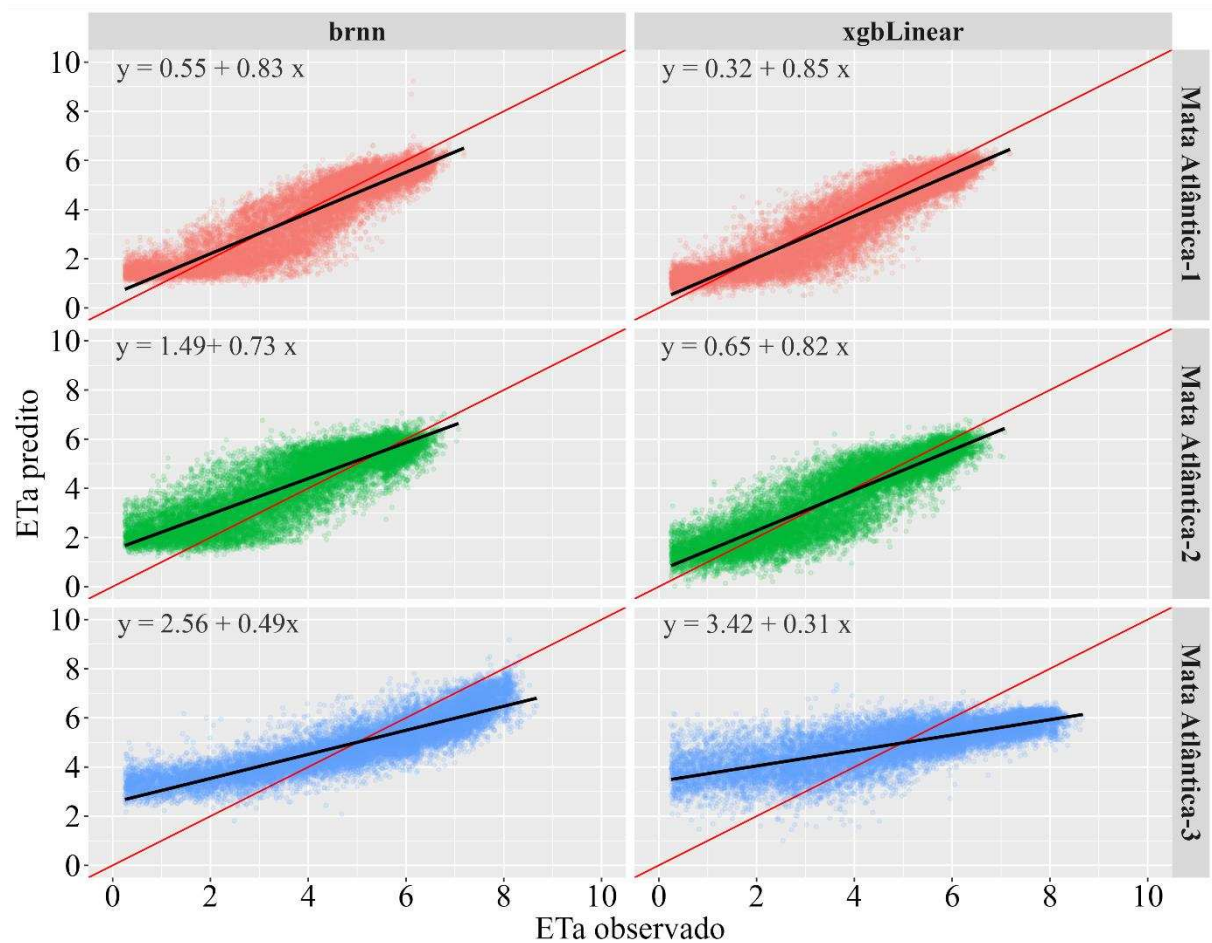


Figura 9. Dispersão entre valores predito e observado para os algoritmos brnn e XgbLinear para as áreas no bioma Mata Atlântica.

Na Tabela 7 estão disponíveis os dados estáticos para as três áreas de coleta desse bioma, onde se nota que o XgbLinear foi melhor nas áreas 1 e 2, porém na área 3 seu resultado foi inferior, tendo R^2 de 0.58 e RMSE de 1.51, enquanto o brnn obteve R^2 de 0.78 e RMSE de 1.18 mm para a mesma área. Contrastando esses resultados com os resultados da Tabela 6 para a Mata Atlântica, em que a predição geral desse bioma com o XgbLinear foi inferior, estatisticamente, ao brnn, fica evidente que essa inferioridade foi proporcionada pela área de coleta 3, em que o modelo não apresentou boa performance, e que isso influenciou negativamente no desempenho geral do modelo para os cultivos nesse bioma.

Tabela 7. Estatística do conjunto de dados destinado ao teste para brnn e XgbLinear nas áreas do bioma Mata Atlântica

Mata Atlântica	\bar{x} obs mm	\bar{x} pred mm	dif %	R^2	MAE mm	RMAE %	RMSE mm	RRMSE %	MBE mm
brnn									
Area-1	4.02	3.88	3.48	0.83	0.58	14.58	0.72	17.96	-0.14
Area-2	3.74	4.21	12.57	0.75	0.74	19.93	0.94	25.29	-0.47
Area-3	5.03	5.02	0.20	0.78	0.95	18.88	1.18	23.43	-0.01
XgbLinear									
Area-1	4.02	3.77	6.22	0.85	0.58	14.43	0.72	17.82	-0.30
Area-2	3.74	3.70	1.07	0.80	0.57	15.18	0.73	19.40	-0.03
Area-3	5.03	4.99	0.80	0.58	1.24	24.75	1.51	30.09	-0.03

Ao analisar a Figura 10, pode-se observar que os modelos brnn e XgbLinear apresentaram subestimativas para as áreas 1 e 2 do bioma Cerrado, com valores observados próximos a 8.5 mm e valores preditos abaixo de 7.5 mm. Já para a área 3, os valores observados foram próximos a 6.5 mm e ambos os modelos estimaram valores semelhantes. No entanto, o modelo brnn apresentou uma superestimativa nessa área, enquanto o XgbLinear apresentou uma subestimativa em todas as áreas. Portanto, é possível concluir que os modelos tiveram desempenhos diferentes para cada área no bioma Cerrado.

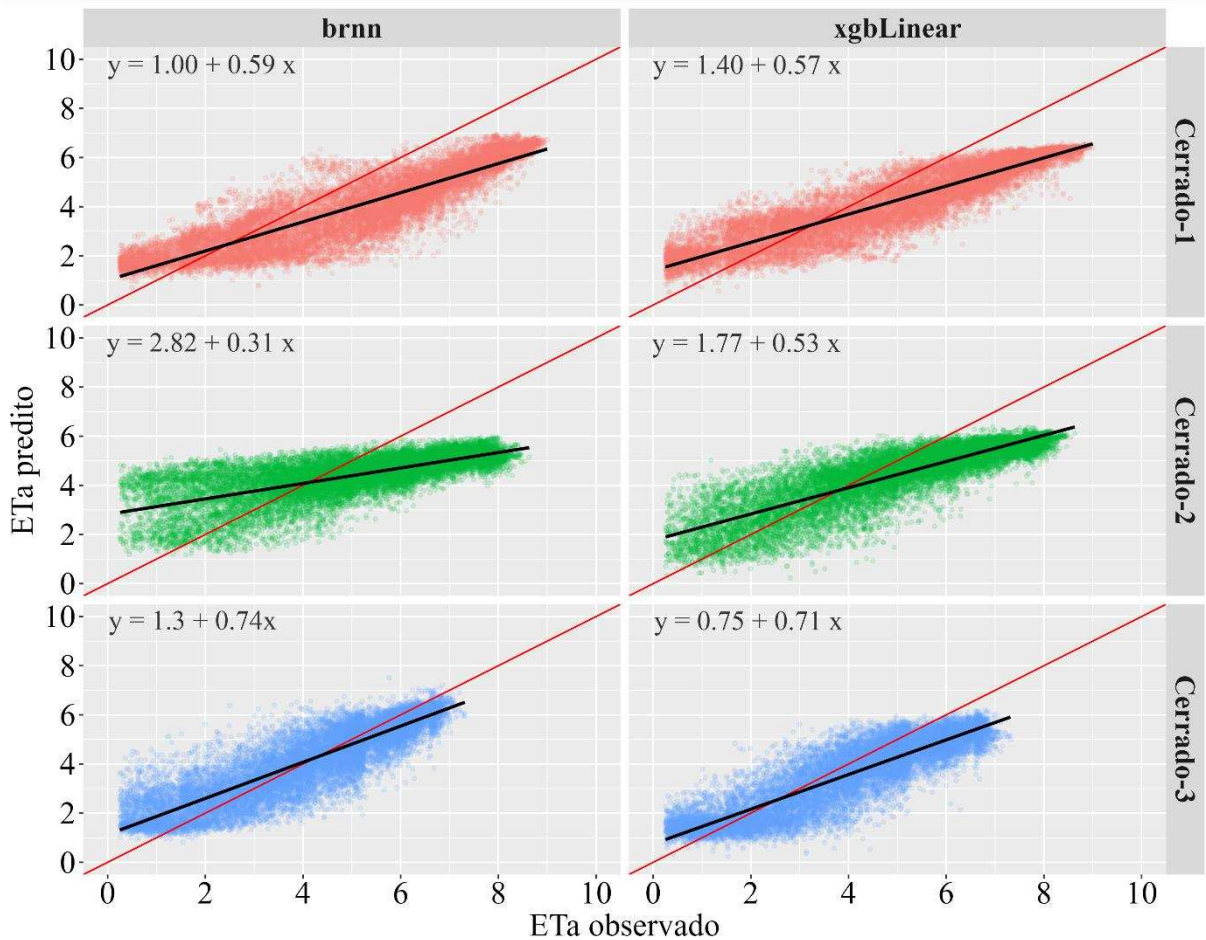


Figura 10. Dispersão entre valores predito e observado para os algoritmos brnn e XgbLinear para as áreas no bioma Cerrado.

Observando a Tabela 8, complementar a Figura 10, os melhores desempenhos dos modelos no Cerrado o erro foi superior que nos melhores desempenho nas áreas da Mata Atlântica. Para as áreas 1 e 2, a melhor performance foi com o XgbLinear com R^2 de 0.85 e 0.74, respectivamente e RMSE de 1.39 e 1.27 mm, nessa ordem, enquanto para área 3 o melhor ajuste foi com o brnn com R^2 igual a 0.79 e RMSE 0.82 mm. Com isso, verifica-se que o melhor desempenho aqui não é superior aos melhores para áreas da Mata Atlântica, ou seja, a predição geral para área da Mata Atlântica os modelos performou-se melhor que para o Cerrado. Outro ponto que chama atenção, tanto nos dados da Figura 9 quando da Figura 10 é que quando os valores observados ultrapassam 7 mm o erro na estimativa aumentou, pois os modelos apresentam boa capacidade preditiva até valores próximos a 6.5. Esse fato pode ocorrer devido à concentração dos dados observados próximos a esse valor, o que pode ter penalizado mais os modelos, e à medida que a evapotranspiração se aproxima dos valores extremos, a capacidade de aprendizagem dos modelos pode reduzir, o que aumenta o erro na predição.

Tabela 8. Estatística do conjunto de dados destinado ao teste para brnn e XgbLinear nas áreas do bioma Cerrado.

Cerrado	\bar{x} obs mm	\bar{x} pred mm	dif %	R ²	MAE mm	RMAE %	RMSE mm	RRMSE %	MBE mm
brnn									
Area-1	5.12	4.05	21.00%	0.81	1.40	27.31	1.58	30.83	-1.07
Area-2	5.18	4.45	14.10%	0.52	1.38	26.70	1.62	31.41	-0.72
Area-3	4.26	4.26	0.07%	0.79	0.63	14.86	0.82	19.43	0.03
XgbLinear									
Area-1	5.12	4.34	15.32%	0.85	1.22	23.86	1.39	27.08	-0.78
Area-2	5.18	4.53	12.49%	0.74	1.06	20.53	1.27	24.48	-0.64
Area-3	4.26	3.76	11.86%	0.80	0.81	19.01	0.96	22.64	-0.50

3.3 Aplicação

Na Figura 11 estão disponíveis os mapas da evapotranspiração obtidos através do modelo brnn e do METRIC. O mapa gerado pelo modelo brnn utiliza variáveis preditoras da Tabela 5, os quais estão inseridos os dados dos satélites Sentinels que possui resolução espacial de 20 m. Já o mapa gerado pelo METRIC utiliza dados dos satélites Landsat-8 e 9, com resolução espacial de 30 m.

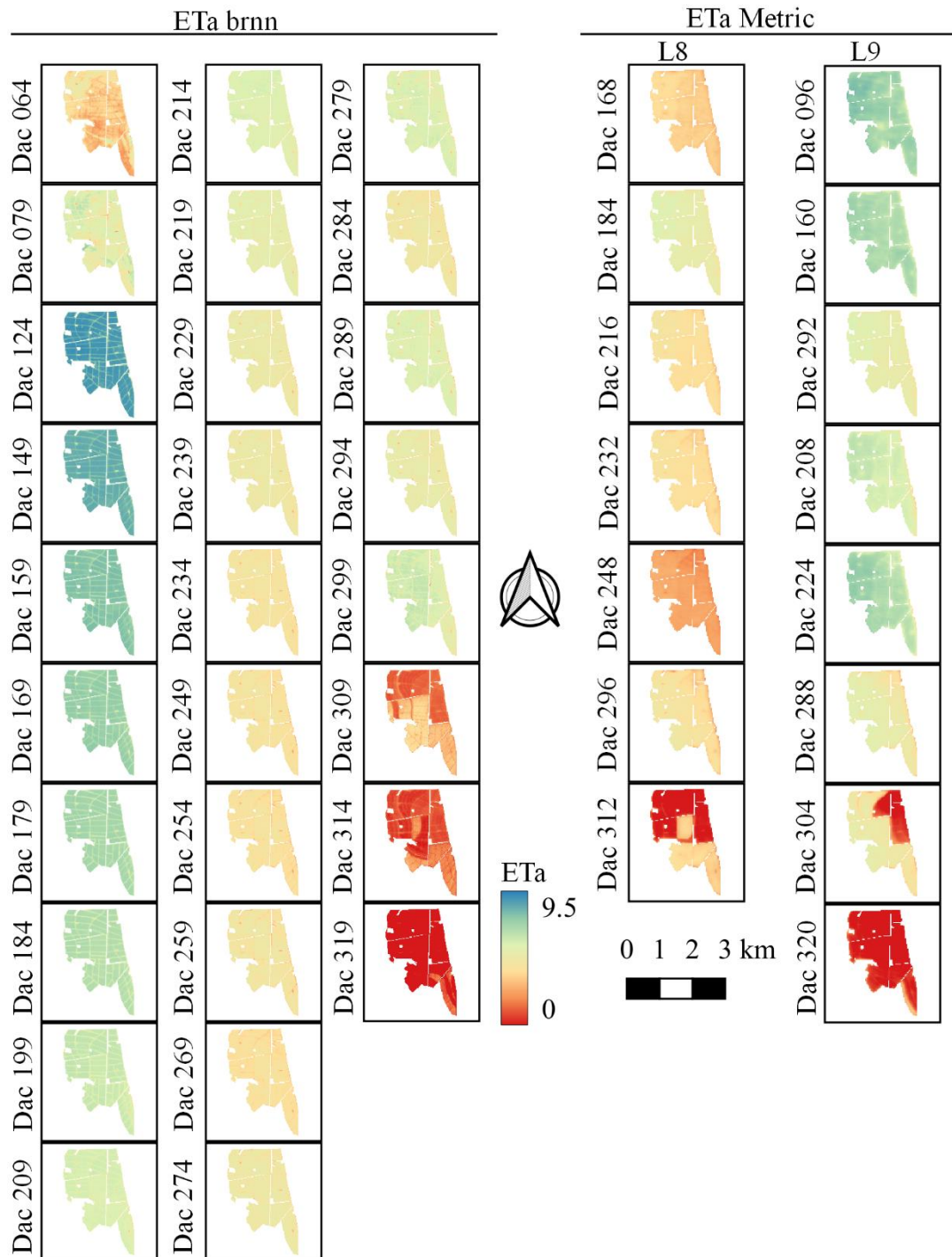


Figura 11. Evapotranspiração atual da cana-de-açúcar com modelo brnn utilizando preditoras e METRIC com Landsat-8 e 9.

Na Figura 11, pode-se observar que a evapotranspiração, quando analisada separadamente, apresenta um número reduzido de datas. A evapotranspiração processada com brnn possui um maior número de datas após a colheita da cana-de-açúcar (Dac), pois o conjunto sensor/satélite dos Sentinels disponibiliza dados a cada 5 dias, o que permitiu adquirir 28 imagens livres de nuvem ao longo da safra, que juntamente com outros dados, foram convertidos em 28 mapas de evapotranspiração. Os Landsats foram responsáveis por

disponibilizar 15 imagens úteis, o que permitiu produzir apenas um pouco mais da metade dos mapas em relação ao Sentinel e demais preditoras. Separadamente, os dados de evapotranspiração são limitados em relação ao número de datas que podem ser adquiridas, mas ao integrar as informações, ou seja, unificar dados da evapotranspiração brnn e METRIC providos pelos Landsats, esse número aumenta, chegando a um total de 43 mapas de evapotranspiração computados ao longo da safra. Outro importante ponto na Figura 11 é a possibilidade de identificar maiores detalhes espaciais com a evapotranspiração brnn devido à sua resolução espacial de 20 m, enquanto o METRIC, com resolução espacial de 30 m, apresenta menos detalhes. O aumento do número de informações espaciais em áreas agrícolas é fundamental para avanços importantes na agricultura de precisão, como destacado por Johansen *et al.* (2022) e Nguyen *et al.* (2022), principalmente na irrigação de precisão, que ainda é incipiente.

Na Figura 12 pode ser analisada a média da evapotranspiração atual em cada data após a colheita, tanto separadamente (Figura 12a) quanto unificadas (Figura 12b). Com isso, percebe-se que a unificação dos dados permite uma compreensão mais precisa da evapotranspiração ao longo do ciclo da cultura, resultando em uma curva mais consistente.

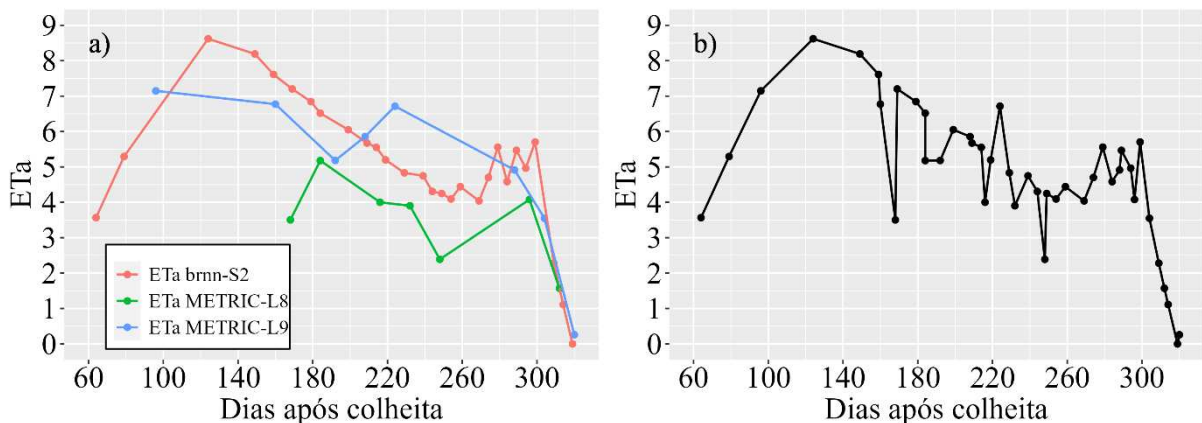


Figura 12. Evapotranspiração atual ao longo do ciclo da cana-de-açúcar por diferentes plataformas (a) e integração das plataformas (b).

Ainda na Figura 12, é possível visualizar que a evapotranspiração tem suas maiores médias até 180 dias após o corte (Dac), e as menores concentra-se por volta de 240 a 260 Dac. Apesar de a área se encontrar no bioma Mata Atlântica, ela é cercada por áreas do bioma Cerrado (conforme mostrado na Figura 1) e, portanto, pode sofrer influência climática desse bioma. Assim, as variáveis climáticas tendem a ser mais amenas entre os meses de abril a setembro, o que reduz a energia disponível para a evapotranspiração e, por consequência, diminui a evapotranspiração atual da cultura. A data de início do último corte foi observada em 21 de setembro de 2021, com isso, por volta dos 180 dias, estão os meses de março. Enquanto

240 a 260 correspondem aos meses de maio e junho, respectivamente, períodos de menor demanda evapotranspirativa da atmosfera. Observa-se que, após 280 Dac, há uma retomada do crescimento da evapotranspiração, mas logo após os 300 Dac ocorre um acelerado declínio da evapotranspiração provocado pela colheita que se iniciou, cortando e destruindo as partes vegetais responsáveis pelo processo transpiratório, ocasionando a redução da evapotranspiração até atingir valor zero quando a colheita é finalizada (Figura 11 – Dac 320) e a área passa a ter apenas biomassa seca.

4 Conclusão

Os modelos brnn e XgbLinear apresentaram desempenho estatisticamente superior aos outros 11 modelos avaliados. O brnn apresentou um desempenho geral com R^2 , RMSE, RRMSE e MBE de 0.73, 1.10, 23.48 e 0.24, respectivamente, enquanto o modelo XgbLinear obteve valores de 0.74, 1.25, 26.52 e 0.60 para estas métricas, respectivamente. Embora os resultados sejam satisfatórios, é importante destacar que a interpretação do comportamento desses modelos também é um desafio devido às suas características de black-box. No entanto, a utilização de metodologias como SHAP neste estudo permitiu compreender a influência que cada variável exerce nesses modelos para explicar a evapotranspiração atual.

Ao analisar especificamente cada localidade de coleta dos dados, a superioridade do brnn se manteve nas áreas cultivadas nos biomas Mata Atlântica e Caatinga, enquanto o XgbLinear prevaleceu no Cerrado. No entanto, dentro das três áreas coletadas do bioma Mata Atlântica, o XgbLinear apresentou melhor desempenho nas áreas 1 e 2, enquanto o brnn teve melhor desempenho na área 3. Já nas três áreas do Cerrado, a superioridade estatística foi observada apenas no XgbLinear.

2.3 Artigo 3

Evapotranspiração da cana-de-açúcar baseado em *machine learning* com abordagens de sensoriamento remoto utilizando o sensor SAR e dados Nasa Power

Resumo

A ET_a (evapotranspiração atual) é amplamente usada na agricultura irrigada. Métodos tradicionais usam parâmetros do solo, planta e atmosfera, mas dependem de valores tabelados e estações meteorológicas. Diante disso, este trabalho teve como objetivo treinar algoritmos de *machine learning* para estimar diretamente a ET_a , sem necessidade de estação meteorológica, utilizando sensoriamento remoto ativo, que não é afetado pela presença de nuvens. Para isso, foram coletados dados de áreas cultivadas com cana-de-açúcar nos biomas Mata Atlântica, Caatinga e Cerrado, utilizando os sensores OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*) embarcados nos satélites Landsat-8 e 9. Esses dados foram utilizados para calcular a ET_a por meio do modelo METRIC (*Mapping Evapotranspiration at High Resolution with Internalized Calibration*), que foi utilizado como variável de resposta. Além disso, foram coletados dados do sensor SAR (*Synthetic Aperture Radar*) embarcado no satélite Sentinel-1, bem como dados de solo, meteorológicos, climatológicos, elevação, latitude, dia do ano e informações astronômicas, que foram utilizados como variáveis explicativas. A fim de evitar a inclusão de variáveis desnecessárias, foi aplicada a remoção de autocorrelação acima de 95% e o método RFE (*recursive feature elimination*). Após essa seleção, restaram 31 variáveis que foram utilizadas em 13 modelos de *machine learning*. Entre esses modelos, o brnn e o XgbTree se destacaram, apresentando um R^2 de 0.45 e 0.46, respectivamente, e um RMSE de 1.5 e 1.6, nessa ordem. Para os dados coletados na Mata Atlântica, o brnn apresentou melhor desempenho estatístico, enquanto para os coletados na Caatinga e Cerrado, o XgbTree foi superior. Portanto, apesar dos valores estatísticos relativamente baixos, é possível estimar a ET_a em áreas com presença de nuvens utilizando os algoritmos de *machine learning* desenvolvidos neste estudo e sem necessidade de estação meteorológica.

Palavras-Chave: Agricultura irrigada. Inteligência artificial. Satélites.

1 Introdução

O sensoriamento remoto é amplamente aplicado em diversos estudos na agricultura, especialmente nas pesquisas relacionadas à evapotranspiração. Desde o lançamento dos primeiros satélites, foram iniciados estudos para determinar a evapotranspiração de áreas agrícolas como uma alternativa aos métodos tradicionais já empregados (BAUSCH, 1995; CHRISTOPHER M. U. NEALE; WALTER C. BAUSCH; DALE F. HEERMANN, 1990; D'URSO; MENENTI, 1995). Portanto, houve muitos avanços, mas ainda são insuficientes para tornar a metodologia aplicada como substituição da tradicional.

O conceito clássico de evapotranspiração corresponde à perda de água através das plantas por meio de estruturas vegetais denominadas estômatos, conhecida conceitualmente como transpiração, e através das superfícies de corpos hídricos, solo e até mesmo da condensação sobre as superfícies vegetais, conhecida como evaporação (ALLEN *et al.*, 1998). Na agricultura, esse conceito é aplicado, mas é na agricultura irrigada que tem maior destaque, pois representa a principal forma de calcular as necessidades de irrigação, ou seja, a partir da evapotranspiração, determina-se a umidade do solo e a quantidade de água que deve ser reposta. No entanto, ainda são utilizados métodos de estimativa tradicionais baseados em estações meteorológicas para quantificar a evapotranspiração de referência para grama (ET_o) ou alfafa (ET_r) e valores tabelados de coeficiente de cultivo, K_c , (ALLEN *et al.*, 1998). Ao multiplicar ET_o ou ET_r pelo K_c , encontra-se a evapotranspiração da cultura, e ao multiplicar novamente pelo coeficiente de umidade do solo, K_s , encontra-se a evapotranspiração atual que está ocorrendo no dia em questão.

Com a perspectiva de estimar a evapotranspiração em áreas agrícolas de forma mais precisa e eficiente, diversos estudos têm sido conduzidos para desenvolver modelos que aproveitem o sensoriamento remoto passivo, aquele que dependem da luz solar para medir a reflectância em diferentes comprimentos de onda, incluindo o visível, infravermelho próximo, de ondas curtas e médio, além da emitância no infravermelho térmico. Esses modelos dispensam a necessidade de utilizar valores pré-definidos de coeficientes de cultura e de umidade do solo. Em vez disso, eles se baseiam na captura de dados da superfície terrestre, aproveitando a luz solar para obter informações valiosas. Os principais modelos com essa capacidade são: TSEB - *Two-Source Energy Balance* (NORMA; KUSTAS; HUMES, 1995), ALEXI - *Atmosphere-Land Exchange Inverse* (ANDERSON, 1997), SEBAL - *Surface Energy Balance Algorithm for Land* (BASTIAANSSEN, 1999), S-SEBI - *Simplified Surface Energy Balance Index* (ROERINK; SU; MENENTI, 2000), METRIC - *Mapping Evapotranspiration at high Resolution with Internalized Calibration* (ALLEN; TASUMI; TREZZA, 2007) e

SAFER - *Simple Algorithm for Evapotranspiration Retrieving* (TEIXEIRA, 2010), sendo o SEBAL e o METRIC os mais amplamente aplicados. Esses modelos superam a necessidade dos coeficientes de cultura e da umidade do solo. No entanto, em sua maioria, requerem sensores térmicos para computar os *pixels* quente e frios, além da dependência de dados provenientes de estações meteorológicas em campo, pois tais dados são utilizados para a estabilidade do complexo modelo aerodinâmico que faz parte da maioria das metodologias para estimar a evapotranspiração via sensoriamento remoto passivo. Além disso, os dados das estações meteorológicas são necessários para determinar a evapotranspiração de referência, seja da grama (ET_o) ou da alfafa (ET_r), que é multiplicada pelo resultado gerado dos modelos via sensoriamento remoto e assim determinar a evapotranspiração atual (ET_a).

Com base no exposto acima, além da dependência da estação meteorológica em campo, os modelos mencionados também compartilham a dependência de sensores passivos e a necessidade de incluir o comprimento de onda do infravermelho térmico, o que limita ainda mais sua aplicação. Os sensores passivos são limitados quando há nebulosidade sobre a área de interesse, pois dependem de uma fonte de luz solar que incide sobre a superfície e é refletida e/ou emitida até os sensores que captam os dados (JENSEN, 2009). Portanto, esses sensores operam nas faixas de comprimento de onda do visível ao infravermelho, que não conseguem atravessar as camadas atmosféricas quando há bloqueio causado por nuvens. Assim, na presença de nuvens, a reflectância ou a emissão não são captadas por esses tipos de sensores, tornando impossível determinar a evapotranspiração por meio desses modelos. Além disso, o comprimento de onda termal é limitado, pois está presente em poucos sensores, o que inviabiliza sua aplicação com maior resolução temporal. Isso implica em dificuldades na obtenção de dados precisos e frequentes sobre a evapotranspiração ao longo do tempo e em áreas específicas. Logo, a utilização desses sensores termais pode ser restrita quando se busca uma análise mais detalhada e abrangente da evapotranspiração.

Em relação à necessidade da banda termal Dos Santos *et al.* (2022) e Costa *et al.* (2023) propuseram uma solução utilizando a constelação Sentinel-2 e técnicas de *machine learning* para quantificar a fração evapotranspirativa, que ao ser multiplicado pela evapotranspiração de referência chega-se na evapotranspiração atual. Como vantagem, os autores mostram independência do sensor termal, no entanto, essa solução apresentada ainda depende de estações meteorológicas, o que pode dificultar sua aplicação. Além disso, o Sentinel-2 possui um sensor passivo, o que significa que sua aplicação será inviabilizada na presença de nebulosidade. Quanto à problemática das nuvens, tem sido tentado solucioná-la por meio do uso de veículos aéreos não tripulados com sensores embarcados (FENG *et al.*, 2022; QIN *et al.*, 2022; SANTOS

et al., 2020). No entanto, essas pequenas aeronaves possuem limitações quanto à aplicação em extensas áreas, a exemplo das áreas cultivada com cana-de-açúcar.

Em função das limitações dos sensores ópticos e da cobertura de área por aeronaves não tripuladas, os satélites que possuem sensores SAR (*Synthetic Aperture Radar*) podem ser uma alternativa viável. Além de abrangerem extensas áreas, os sensores SAR não possuem banda termal e não são limitados pela presença de nuvens. Os sensores SAR são ativos, ou seja, emitem seu próprio sinal de radar que interage com os alvos na superfície terrestre e retorna ao sensor. Esses sensores operam na banda C (C-band) com comprimento de onda entre 3.75 cm e 7.5 cm, o que permite uma menor interação com as partículas presentes nas nuvens e facilita a penetração do sinal. Alguns estudos estão sendo desenvolvidos para utilizar dados provenientes desses sensores na estimativa da evapotranspiração (KAPLAN *et al.*, 2021, 2023; MUSYIMI *et al.*, 2022; OUAADI *et al.*, 2023). Esses estudos estão direcionados para a quantificação do coeficiente de cultivo, representando um avanço, pois eliminaria a dependência de coeficientes tabelados utilizados no manejo da irrigação. Contudo, ainda é necessário o uso de estações meteorológicas, uma vez que o coeficiente de cultivo representa uma parte do processo de evapotranspiração atual. Por conseguinte, é evidente a necessidade de aprimorar as técnicas que permitam estimar diretamente a evapotranspiração atual mesmo na presença de nuvens, sem depender de estações meteorológicas físicas.

A estimativa da evapotranspiração atual é um processo complexo, pois envolve múltiplas informações do sistema solo-planta-atmosfera, cada uma com suas peculiaridades e dinâmica. No entanto, é possível utilizar diversas fontes de dados e informações em algoritmos de inteligência artificial, como os modelos de *machine learning*. Esses modelos permitem que o sistema computacional aprenda de forma intuitiva a partir de exemplos e experiências previamente estabelecidos (BATTA, 2018). Assim, com múltiplas fontes de dados, os modelos podem identificar os padrões presentes nesses dados, atribuindo-lhes pesos e melhorando a consistência das estimativas.

Com base nos argumentos apresentados sobre a capacidade dos algoritmos de inteligência artificial e na necessidade de estimar diretamente a evapotranspiração atual na presença de nuvens por meio do sensoriamento remoto, este estudo tem como objetivo treinar algoritmos de *machine learning* para estimar a evapotranspiração atual utilizando dados de radar e múltiplas fontes de informação na cultura da cana-de-açúcar quando há presença de nuvens.

2 Materiais e Métodos

2.1 Área de coleta dos dados

Para realizar este estudo, foram coletados dados nos três principais biomas brasileiros com a maior área cultivada de cana-de-açúcar: a Mata Atlântica em primeiro lugar, seguida do Cerrado e da Caatinga (Figura 1). A coleta de dados abrangeu diferentes latitudes no Brasil, permitindo obter dados mais abrangente e representativos dos cultivos nessas localidades.

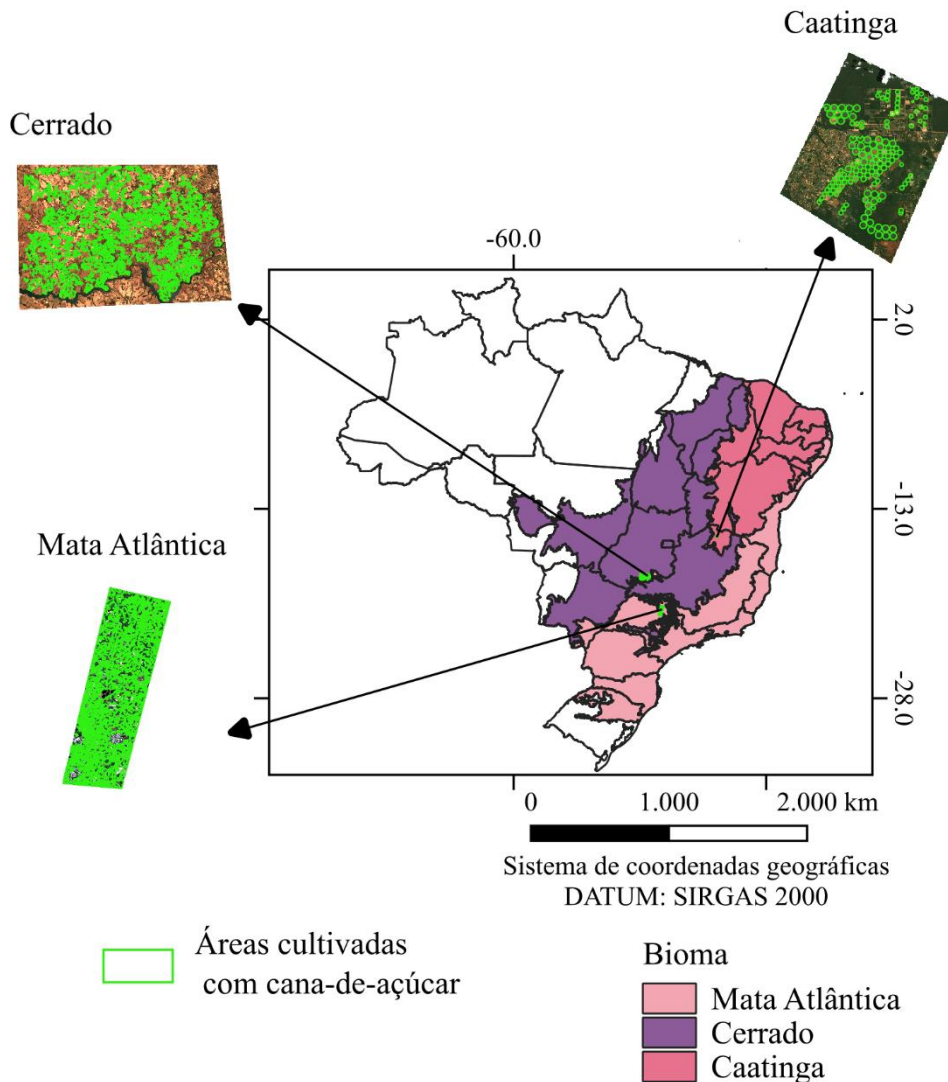


Figura 1. Localização geográfica das áreas de coleta dos dados para modelagem da evapotranspiração atual na cana-de-açúcar

As três áreas selecionadas por estarem em biomas diferentes apresentam características climáticas distintas (Figura 2), o qual favorece no processo de modelagem pois é possível obter maior variabilidade nos dados.

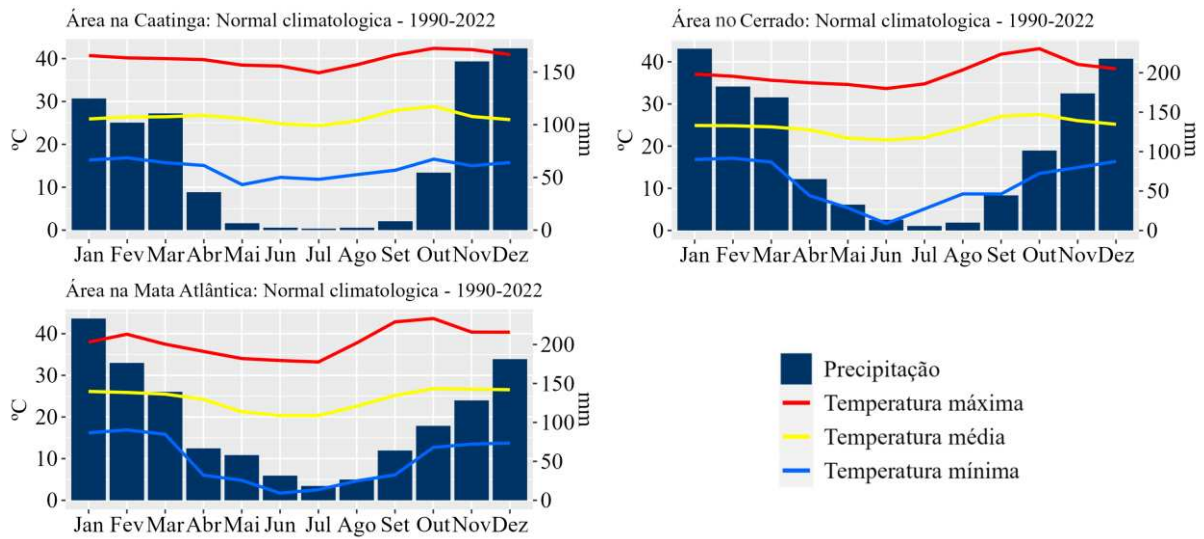


Figura 2. Climatologia com base nos produtos Nasa Power das três áreas selecionadas para o estudo.

Os dados coletados geraram a variável resposta e as variáveis explicativas, que foram incorporadas em modelos de *machine learning*, treinados e testados para avaliar seu desempenho. Na Figura 3, é possível visualizar de forma simplificada os procedimentos metodológicos de aquisição dessas variáveis e o subsequente treinamento dos algoritmos.

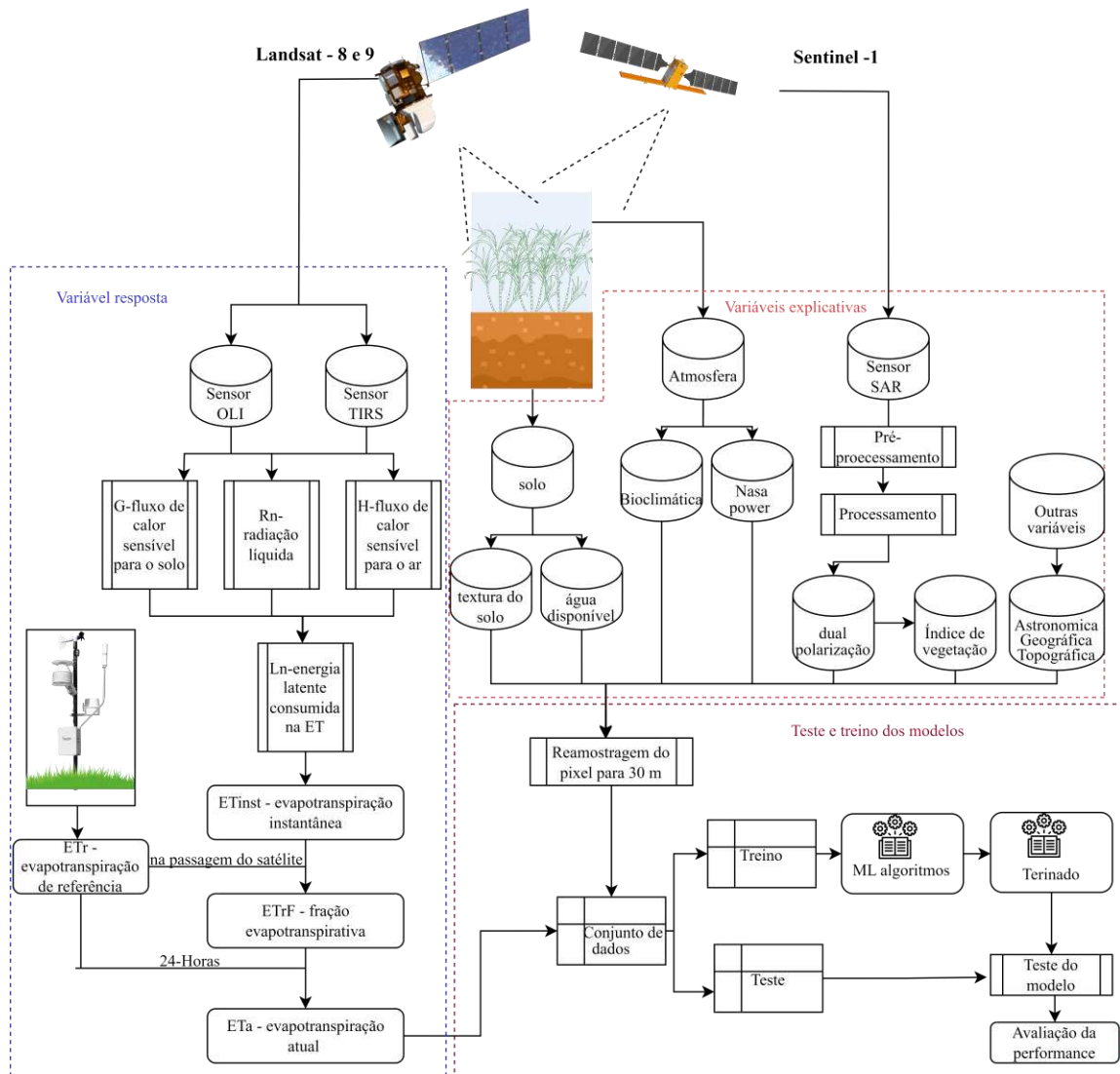


Figura 3. Simplificação metodológica da aquisição da variável resposta e variáveis explicativa ao treinamento dos modelos de *machine learning*.

2.2 Variável resposta - Evapotranspiração atual

A evapotranspiração atual na cana-de-açúcar foi calculado a partir do modelo METRIC (ALLEN; TASUMI; TREZZA, 2007), o qual exige dados espectrais de reflectância e emitância da superfície. Para tal, adquire-se os dados do sensor OLI (*Operational Land Imager*), capta a reflectância, e o TIRS (*Thermal Infrared Sensor*), capta a emitância, que estão embarcados nos satélites Landsat-8 e 9. Assim, foram adquiridos, na plataforma USGS (*United States Geological Survey*) (<https://earthexplorer.usgs.gov/>), 53 imagens desses satélites na coleção-2 em nível-2 entre os anos de 2017 a 2022 (Tabela 1).

Tabela 1. Número de imagens adquiridas para cada bioma de coleta dos dados.

Bioma	Nº de imagens	Path/Row	Ano
Caatinga	16	218/071	2017-2021
Cerrado	18	222/073	2017-2021
Mata Atlântica	19	220/075	2017-2022

2.2.1 Determinação da ET_a

A ET_a foi determinado com base no modelo METRIC (ALLEN; TASUMI; TREZZA, 2007) e os cálculos para alcançar o seu valor dependem das energias de radiação líquida (R_n), fluxo sensível do calor transferido para o solo (G) e fluxo sensível do calor convectado para o ar (H), os quais envolvidos é possível extrair a LE que é a energia latente consumida no processo evapotranspirativo da superfície, assim, a LE, em $W m^{-2}$, pode ser determinada pela Eq.1.

$$LE=R_n-G-H \quad (1)$$

A R_n é a energia radiante da superfície que é particionada em H , G e LE, e é determinada pela Eq. (2)

$$R_n=R_{s\downarrow}-\alpha R_{s\downarrow}+R_{L\downarrow}-R_{L\uparrow}-(1-\varepsilon_0)R_{L\downarrow} \quad (2)$$

onde $R_{s\downarrow}$ é a entrada da radiação de ondas curtas ($W m^{-2}$); α é o albedo de superfície (adim.), determinado pelas bandas 2, 3, 4, 5, 6 e 7 do Landsat-8; $R_{L\downarrow}$ e $R_{L\uparrow}$ é a entrada e saída de ondas longas ($W m^{-2}$), respectivamente; e ε_0 é a emissividade térmica da superfície.

O calor transferido para o solo foi determinado utilizando o modelo de Tasumi (2003) que envolve a R_n , índice de área foliar (LAI), estimado pela metodologia aplicada por Allen *et al.* (2011), e temperatura de superfície (T_s). Este parâmetro necessita de uma condicional, onde se a LAI for igual ou superior a 0.5, então faze-o uso da Eq. (2a), senão, faz-se o uso da Eq. (2b).

$$\frac{G}{R_n}=0.05 + 0.18e^{-0.521 LAI} \quad \text{para } LAI \geq 0.5 \quad (2a)$$

$$\frac{G}{R_n} = \frac{1.80(T_s-273.15)}{R_n+0.084} \quad \text{para } LAI < 0.5 \quad (2b)$$

Por fim, determinou-se o fluxo sensível do calor convectado para o ar através da Eq. (3). Para estimá-lo foi necessário a função aerodinâmica que considera a ar em duas alturas, geralmente, a 0.1 m e 2.0 m do solo, que, por sua vez, é calculado usando a velocidade do vento extrapolada de alguma altura acima da superfície do solo, normalmente 100 para 200 m, e um esquema de correção de estabilidade interativa com base nas funções Monin-Obhukov.

$$H=\rho_{\text{air}} C_p \frac{dT}{r_{ah}} \quad (3)$$

onde ρ_{air} é a densidade do ar ($kg m^{-3}$); C_p é o calor específico do ar a pressão constante ($J kg^{-1} K^{-1}$); dT é a diferença de temperatura entre duas alturas (z_1 e z_2) em uma área próxima à superfície; e r_{ah} é a resistência aerodinâmica ($s m^{-1}$) entre essas duas alturas.

Após a LE devidamente calculada, estabeleceu-se a evapotranspiração instantânea pela Eq. 4, e correspondente a evapotranspiração que ocorre no momento exato da passagem do satélite convertido para hora. A ET_{inst} é necessária para o cálculo da fração evapotranspirativa (ET_{rF}) que, juntamente com dados meteorológicos, determina-se a ET_a .

$$ET_{inst} = 3600 \frac{LE}{\lambda \rho_w} \quad (4)$$

onde ET_{inst} é a evapotranspiração instantânea ($mm.h^{-1}$); 3600 é a conversão de segundos, intervalo de tempo da passagem do satélite, para hora; LE é a energia latente consumida na ET ($W m^{-2}$); ρ_w é densidade da água ($\sim 1000 kg m^{-3}$); e λ é o calor latente de vaporização ($J kg^{-1}$).

O λ representa o calor absorvido quando um kg de água é evaporada e foi calculado pela Eq. (5).

$$\lambda = [2.501 - 0.00236(T_s - 273.15)] \times 10^6 \quad (5)$$

onde T_s é a temperatura de superfície ($^{\circ}K$) determinada pela banda 10 do sensor TIRS (Tabela 1).

A ET_{rF} se baseia na razão entre a evapotranspiração instantânea (ET_{inst}), em cada *pixel*, pela evapotranspiração, horária, de referência para alfafa (ET_r), calculada pela estação meteorológica que se encontra dentro de cada cena orbital. Assim, a ET_{rF} é determinada a partir da Eq. (6).

$$ET_{rF} = \frac{ET_{inst}}{ET_r} \quad (6)$$

onde ET_{inst} é a evapotranspiração instantânea ($mm.h^{-1}$) e ET_r é evapotranspiração de referência ($mm.h^{-1}$) padronizada para alfafa a 0.5 m de altura no momento da passagem do satélite.

Com intuito de chegar no valor da ET_a da cana-de-açúcar, calculou-se a evapotranspiração de referência acumulada em um período de 24 h (ET_{r-24}) no dia que cada imagem foi adquirida. A ET_{r-24} , em cada data, foi calculada utilizando a equação Penman-Monteith padronizado pela ASCE (*American Society of Civil Engineers*) (ALLEN *et al.*, 2005) e os dados meteorológicos. Finalmente, com os dados da ET_{r-24} e do ET_{rF} , determinou-se a ET_a pela Eq. (7).

$$ET_a = ET_{rF} \times ET_{r-24} \quad (7)$$

Todo o cálculo até chegar na ET_a descrito nesse tópico foi realizado utilizando o pacote *Water*, desenvolvido por Olmedo *et al.* (2018) para ambiente de linguagem R (R CORE TEAM, 2020).

2.3 Seleção das variáveis explicativas

Devido à complexidade da estimativa da evapotranspiração atual na cana-de-açúcar a partir do sensoriamento por radar e na ausência de estação meteorológica em campo, foi necessário utilizar covariáveis que estão relacionadas com o processo evapotranspirativo. Assim, além dos dados do Sentinel-1, que possui o sensor SAR, utilizou-se dados de solo, meteorologia via Nasa Power, elevação, e bioclimáticas via WorldClim. Com isso, percebe-se que as variáveis explicativas selecionadas neste estudo fazem parte da relação solo-planta-atmosfera.

2.3.1 Planta

No que diz respeito a parte da planta, no sistema, utilizou-se dados de sensoriamento remoto por radar do satélite Sentinel-1, devido sua característica de imageamento independente das condições climáticas.

2.3.1.1 Aquisição e processamento dos dados do Sentinel-1

O Sentinel-1 é uma constelação com dois satélites, o Sentinel-1A e 1B, responsável por monitorar a superfície terrestre a cada seis dias, porém, para o Brasil atualmente, devido falha de operacional no 1B, seu monitoramento é com frequência de 12 dias em uma resolução espacial, após os pré-processamentos, de 10 metros. Esses dois satélites possuem embarcado o Radar de Abertura Sintética (SAR) na banda-C que corresponde a frequência de 4 a 8 GHz e comprimento de onda entre 3.8 a 7.5 cm.

Os dados do Sentienel-1 foram adquiridos na plataforma *Alaska Satellite Facility* (<https://search.asf.alaska.edu/>) no mesmo dia em que ocorreu a passagem dos satélites Landsat-8 e 9 (Tabela 1), objetivando adquirir dados sem nenhuma alteração das condições da cultura em campo. Esses dados foram adquiridos no nível-1 produzidas como GRD (*Ground Range Detected*) no modo de faixa IW (*Interferometric Wide*) que fornece imagens de polarização dupla (VV e VH) com *pixels*, aproximadamente, de 20 m x 20 m. Posterior sua aquisição, foi necessário realizar os devidos pré-processamentos para tornar as imagens úteis a serem utilizadas nas etapas de processamento e extração da informação para alimentar os modelos de *machine learning*. Desta forma, um fluxo de trabalho em sete etapas foi realizado para reduzir as fontes de erros que pode consistir nos dados. As sete etapas, sequencialmente, foram: 1)

Aplicação da órbita; 2) Remoção de ruído térmico; 3) Remoção de ruído de borda; 4) Calibração radiométrica; 5) Filtro para remoção do speckle; 6) Correção doopler de alcance do terreno; e 7) Conversão para decibéis (dB). Todas essas etapas foram realizadas com códigos para linguagem Python (SANTOS; BHOGAPURAPU, 2022).

A etapa de aplicação da órbita é necessária, pois os vetores do estado da órbita contidos nos metadados no momento da passagem do satélite possuem baixa precisão. No entanto, a precisão melhora após dias ou semanas da passagem do satélite e da geração do produto. A atualização com os vetores de órbita mais precisos é realizada por meio de download automático ao executar o código em Python e adquirir os metadados do produto com informações mais precisas. A remoção de ruído térmico para arquivos do tipo GRD é realizada utilizando LUT (*look-up table*), fornecida pelo produto do Sentinel-1, que contém informações sobre o ruído presente nos dados. Os valores na tabela são fornecidos em potenciais lineares, que, ao serem aplicados, geram um perfil de ruído de calibração para remover o ruído térmico. Outra importante correção realizada é a remoção de ruído de borda, que corrige o horário de início da amostragem e remove variações indesejáveis na intensidade do sinal na imagem. Após essas três etapas, procede-se à calibração radiométrica, que associa os valores produzidos na imagem, em número digital, com o retroespalhamento da superfície, para o qual a equação (8) é aplicada.

$$\sigma_i^0 = \frac{(DN_i^2 + b)}{A_i} \quad (8)$$

em que σ_i^0 corresponde ao retroespalhamento, DN_i é a intensidade do sinal em número digital; b é um deslocamento constante; e A_i é a calibração sigma para cada *pixel* (i).

Com os dados do Sentinel-1 calibrados radiometricamente, foi necessário aplicar um filtro para reduzir o ruído *speckle*, o que garante uma imagem de melhor qualidade. O *speckle* é um tipo de ruído alta frequência muito comum em dados SAR e pode gerar dois problemas: degradação quantitativa da imagem e dificuldade de interpretação visual. Para removê-lo, foi aplicado o filtro Lee, que é o mais utilizado para esse tipo de ruído em dados de radar. Ele é baseado no critério do erro quadrático médio mínimo (GAGNON; JOUAN, 1997), removendo os ruídos e mantendo a nitidez das imagens.

Devido às distorções geométricas geradas nas imagens SAR causadas pela visada oblíqua do sensor, foi preciso realizar sua correção. A correção foi feita a partir do operador implementado no código Python de correção do Doppler de alcance do terreno, que implementa o método de ortorretificação de amplitude Doppler a partir do modelo digital de elevação de 30 metros, que é adquirido automaticamente ao executar o código. Por fim, os coeficientes de

retroespalhamento, gerados na equação 8 e devidamente corrigidos, são convertidos para escalonamento em decibéis (dB) e esse processo é realizado aplicando a Eq. (9).

$$\sigma_i^0(\text{db})=10\log_{10}\sigma_i^0 \quad (9)$$

Ao fim de todas as etapas, tinha-se duas polarizações, a VH e a VV, representadas pelos coeficientes de retroespalhamento σ^0_{VH} e σ^0_{VV} , respectivamente, devidamente corrigidas, com ruídos reduzidos e em decibéis. Em função do número reduzido de variáveis predictoras advinda do sensor embarcado no Sentinel-1, foi necessário gerar novas variáveis a partir dos coeficientes de retroespalhamento. Desta forma, foram gerados mais seis índices a partir dos valores de retroespalhamento, os quais estão disponíveis na Tabela 2 da Eq. 10 a 15.

Tabela 2. Variáveis predictoras geradas a partir do retroespalhamento

Índice	Equação	Referência
Pol	$\frac{\sigma^0_{VH}-\sigma^0_{VV}}{\sigma^0_{VH+VV}}$ (10)	(HIRD <i>et al.</i> , 2017)
CR	$\frac{\sigma^0_{VV}}{\sigma^0_{VH}}$ (11)	(FRISON <i>et al.</i> , 2018)
DPSVI	$\frac{\sigma^{VH(i)} \left[(\sigma^{VV}_{(max)}\sigma^{VH(i)}-\sigma^{VV(i)}\sigma^{VH(i)+})+\sigma^2_{VH(i)} \right]}{\sqrt{2} \sigma^{VV(i)} + (\sigma^{VV}_{(max)}\sigma^{VV(i)}-\sigma^2_{VV(i)}+\sigma^{VH(i)}\sigma^{VV(i)})}$ (12)	(PERIASAMY, 2018)
RVIm	$\frac{4\sigma^0_{VV}}{\sigma^0_{VH}+\sigma^0_{VH}}$ (13)	(NASIRZADEHDIZAJI <i>et al.</i> , 2019)
DPSVIm	$\frac{\sigma^0_{VV^2}+\sigma^0_{VV}\sigma^0_{VH}}{\sqrt{2}}$ (14)	(DOS SANTOS; DA SILVA; DO AMARAL, 2021)
DpRVic	$\frac{q(q+3)}{(q+1)^2}$ (15)	(BHO GAPURAPU <i>et al.</i> , 2022)

em que q corresponde a razão de intensidade de retroespalhamento.

Após o processo de pré-processamento e geração dos índices, foram obtidos oito variáveis predictoras para os modelos, que incluem as duas polarizações e seis índices. No entanto, a resolução espacial de cada *pixel* é de 10 m, o que é incompatível com o Landsat 8 e 9, que possuem *pixels* de 30 m. Para compatibilizar os dados do Sentinel-1 com os dados do Landsat, foi aplicado o método de agregação utilizando a função *aggregate* do pacote raster (HIJMANS *et al.*, 2020) na linguagem R, com um fator de escala de 3, resultando em *pixels* com dimensões de 30 m x 30 m. Para evitar quaisquer incompatibilidades, foi utilizada a função *resample* com o método bilinear, também do pacote raster, utilizando como base a imagem raster da ET_a. Agora, todos os dados do provenientes do sensor SAR, embarcado no Sentine-1, estavam compatíveis com os dados advindos dos sensores OLI e TIR, abordo nos Landsat's.

2.3.2 Solo

Para o componente solo do sistema, foram adquiridos dados de silte, areia, argila e água disponível na plataforma GeoInfo (<http://geoinfo.cnps.embrapa.br/>), que hospeda a base de dados espaciais da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). Foram obtidos três tipos de dados para cada textura de solo e disponibilidade de água, representando a profundidade da camada do solo. Assim, foram coletados dados das camadas de 0-0.05 m, 0.05-0.15 m e 0.15-0.30 m. Esses produtos são oriundos de mapeamento digital de solos a partir de covariáveis geoespaciais em modelos de *machine learning*.

Como os dados foram estratificados em três camadas, optou-se por calcular a média dos valores para gerar outro arquivo com a média das camadas de 0-0.30 m. Os dados foram adquiridos em formato raster com resolução espacial de 90 m, e foi necessário aplicar a metodologia inversa ao que se aplicou para os dados do SAR. Desta forma, aplicou-se a metodologia de desagregação dos *pixels* a partir da função *disaggregate*, também do pacote raster, com fator de escala igual 3, com isso os *pixels* foram redimensionados de 90 metros para 30 metros. Para evitar incompatibilidade foi utilizado a função *resample* com o método bilinear com a mesma imagem raster da ET_a.

2.3.3 Atmosfera

Para representar a atmosfera, foram utilizados dados relacionados diretamente com as condições climatológicas e meteorológicas de cada localidade, bem como dados relacionados indiretamente. Para os dados diretamente relacionados, optou-se por duas fontes: a primeira são os produtos de bioclimáticas do WorldClim, que se baseiam em séries históricas de dados climáticos e caracterizam o bioclima de uma localidade; e a segunda são dados meteorológicos do produto Nasa Power, que estão disponíveis diariamente e caracterizam as condições meteorológicas diárias, as quais interferem no potencial evapotranspirativo da atmosfera.

2.3.3.1 Nasa Power

Nasa power é um conjunto de dados e uma ferramenta online fornecidos pela NASA (*National Aeronautics and Space Administration*) que fornece informações sobre radiação solar, temperatura do ar, umidade relativa, velocidade do vento e outras variáveis meteorológicas em todo o mundo. Esses dados foram adquiridos por meio do pacote *nasapower* para o ambiente R, que leva o nome do próprio produto (SPARKS, 2018). A obtenção desses dados por meio do pacote ocorre em formato de *data frame*, que consiste em dados do tipo linha e coluna, com resolução espacial de 0.5° x 0.5°, utilizando as coordenadas de latitude e longitude como referência para a localidade. Para evitar a reamostragem de *pixels* desses dados, foram utilizadas

as latitudes e longitudes dos *pixels* contidos nos dados raster, que são do tipo matriz, da ET_a . Com isso, os dados do nasapower foram adquiridos em formato de *data frame* correspondente a cada par de coordenadas do raster de 30 m. Em seguida, esses dados foram convertidos em raster por meio do pacote raster (HIJMANS *et al.*, 2020) em linguagem R.

As variáveis obtidas foram aquelas utilizadas para calcular a ET_r . Dessa forma, adquiriu-se a radiação solar incidente (R_s , $MJ\ m^{-2}\ dia^{-1}$), a temperatura máxima do ar (T_{mx} , °C), a temperatura mínima do ar (T_n , °C), a temperatura média (T_{med} , °C), a umidade relativa média (RH, %) e a velocidade média do vento (V_v , $m\ s^{-1}$) a 2 m de altura. No entanto, essas variáveis não foram obtidas na mesma data de passagem do satélite, pois os dados do nasapower apresentam um atraso de dois dias. Para contornar essa problemática, as variáveis foram adquiridas 2, 3, 4 e 5 dias antes da passagem do satélite. Além dessas datas individuais, foi calculada a média e o desvio padrão desses quatro dias, com o intuito de melhor representar as condições meteorológicas do dia da passagem do satélite. Decidiu-se, também, calcular a evapotranspiração de referência para grama (ALLEN *et al.*, 1998) e para alfafa (ALLEN *et al.*, 2005) com as quatro datas e suas respectivas médias e desvio padrão.

2.3.3.2 Bioclimáticas

As bioclimáticas foram adquiridas na plataforma da WorldClim (<https://www.worldclim.org/data/worldclim21.html>) em formato raster com resolução espacial de, aproximadamente, 925 m, ou seja, baixa resolução espacial, sendo totalmente incompatível para ser utilizado, em sua forma original, com dados de 30 m. Diante disso, aplicou-se a mesma metodologia daquela aplicada às covariáveis de solos com a diferença que, aqui, aplicou-se o fator de escala igual a 30.86. Todas as bioclimáticas obtidas podem ser visualizadas na Tabela 3.

Tabela 3. Bioclimáticas da base de dados globais do WorldClim.

Código	nomenclatura
bio1	Temperatura Média Anual
bio2	Intervalo Diurno Médio
bio 3	Isotermalidade
bio 4	Sazonalidade da Temperatura
bio 5	Temperatura máxima do mês mais quente
bio 6	Temperatura mínima do mês mais frio
bio 7	Faixa Anual de Temperatura
bio 8	Temperatura média do trimestre mais úmido
bio 9	Temperatura média do trimestre mais seco
bio 10	Temperatura Média do Trimestre Mais Quente
bio 11	Temperatura média do trimestre mais frio
bio 12	Precipitação Anual
bio 13	Precipitação do mês mais úmido
bio 14	Precipitação do Mês Mais Seco
bio 15	Sazonalidade de Precipitação
bio 16	Precipitação do trimestre mais úmido
bio 17	Precipitação do trimestre mais seco
bio 18	Precipitação do Quarto Mais Quente
bio 19	Precipitação do trimestre mais frio

2.3.3.3 Outras variáveis relacionadas com atmosfera

Além das variáveis diretamente relacionadas, também optou pelas variáveis que interfere nas condições climáticas, sendo a latitude, elevação, dia do ano, mês do ano e as astronômicas. As latitudes foram facilmente extraídas dos próprios arquivos raster em cada *pixel* de 30 m, já a elevação foi necessária adquirir via plataforma USGS, dos dados SRTM (*Shuttle Radar Topography Mission*), em formato raster, com dimensão de 30 m x 30 m, ou seja, já compatível com os dados. O dia do ano e o mês foram extraídas das datas referente a cada dia de passagem do satélite. Por fim, produziu-se as variáveis astronômicas com base na latitude e dia ano, gerando as variáveis de declinação solar (em graus decimais), o ângulo horário do nascer do sol (em radianos), o fotoperíodo (em horas), a distância Terra-Sol (em radianos) e a radiação global extraterrestre (em MJ m⁻² d⁻¹) seguindo a metodologia proposto por Brock (1981).

2.4 Treinamento dos modelos

Após todos os dados devidamente organizados e reamostrados para as mesmas dimensões, separou-se eles em dois conjuntos, um destinado ao treino do modelo e o outro destinado ao teste. A separação ocorreu selecionando de modo aleatório uma data dentro de cada bioma, ou seja, um conjunto de dados contendo 3 datas distintas. Essas três datas foram destinadas à avaliação da performance dos modelos, enquanto as demais foram utilizadas para treiná-los. A separação desse modo garante maior confiabilidade nos resultados do desempenho.

Se os dados fossem extraídos juntos e depois separados em treinamento e teste, poderia acontecer de dados espacialmente próximos e de mesma data serem incluídos em diferentes conjuntos, o que levaria a um desempenho aparentemente bom do modelo devido à semelhança dos dados, mas não necessariamente por sua capacidade preditiva real.

Após a devida separação dos dados, foram extraídos conjuntos de pontos dos arquivos raster, resultando em 175.000 dados para cada bioma, totalizando 525.000 dados. Esse processo foi realizado utilizando a função *extract* do pacote raster (HIJMANS *et al.*, 2020) na linguagem computacional R. Essa função permite obter os valores dos *pixels* juntamente com suas coordenadas geográficas e armazená-los em um *data frame*, onde as colunas representam as variáveis explicativas e a variável resposta.

Observa-se que o número de variáveis inicialmente selecionadas como preditoras (explicativas) é elevado, o que pode levar à presença de autocorrelação entre elas, assim como a inclusão de variáveis com pouco impacto no modelo. Manter todas essas variáveis resultaria em um gasto computacional desnecessário. Diante disso, utilizou-se a função *findCorrelation* do pacote caret (KUHN *et al.*, 2020), também na linguagem R, com um limite de correlação (*cutoff*) igual a 0.95, removendo assim as variáveis com correlação acima desse valor, o qual pode evitar que ocorra, após os treinos dos modelos, *overfitting*. Esse processo reduziu significativamente o número de variáveis a serem utilizadas no método de seleção subsequente, chamado RFE (*Recursive Feature Elimination*) com a utilização da função *Random Forest*.

O RFE é amplamente aplicado em ciências de dados para selecionar as características mais relevantes de um conjunto de dados para os algoritmos de *machine learning*. Ele atua de forma iterativa, eliminando sucessivamente as características menos importantes com base em critérios de relevância, até que reste um conjunto final com as variáveis preditoras mais significativas para o treinamento (DUARTE *et al.*, 2022; HAN *et al.*, 2019; IMPOLLONIA *et al.*, 2022).

Após todo o preparo do conjunto de dados, os modelos foram treinados. Utilizou-se os algoritmos de *linear regression* (LM), *cubist*, *extreme gradient boosting* - método linear (XgbLinear), *extreme gradient boosting* - método tree (XgbTree), *principal components regression* (pca), *recursive partitioning* (rpart), *bayesian regularized neural networks* (brnn), *partial least squares* (pls), *ridge regression* (ridge), *generalized linear model* (glm), *lasso*, *stochastic gradient boosting* (gbm) e *k-nearest neighbor* (knn) A escolha desses modelos foi baseada em sua boa capacidade preditiva e na diversidade de suas arquiteturas, permitindo explorar diferentes abordagens e identificar o modelo com melhor desempenho.

2.5 Performance e exploração das variáveis

Para avaliar a performance dos 13 modelos após o treinamento, adquiriu-se dos conjuntos de dados destinado a esse intuito 75.000 dados para cada bioma, assim gerou-se no total 225.000. Com esses dados, realizou-se a predição da evapotranspiração atual na cana-de-açúcar e em seguida fez-se uso dos dados observados e preditos em métricas estatísticas para avaliar a capacidade preditivas. As métricas utilizadas foram: o R^2 (coeficiente de determinação), Eq. (15), RMSE (raiz do erro quadrático médio), Eq. (16), RRMSE (raiz do erro quadrático médio relativo), Eq. (17), MAE (erro médio absoluto), Eq. (18), MAPE (erro médio absoluto relativo), Eq. (19), e MBE (erro médio de viés), Eq. (20).

$$R^2 = \frac{[\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})]^2}{[\sum_{i=1}^n (P_i - \bar{P})^2][\sum_{i=1}^n (O_i - \bar{O})^2]} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i)^2}} \times 100 \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (18)$$

$$RMAE = \frac{\frac{1}{n} \sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n (O_i)} \times 100 \quad (19)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (20)$$

onde P_i é o valor previsto pelo modelo; O_i é o valor observado; \bar{O} é o valor médio observado e n é o número de pares de dado.

As métricas estatísticas indicam a performance de cada modelo, podendo ser escolhido aquele com melhor capacidade preditiva. No entanto, para conhecer como os modelos de *machine learning* compreende cada variável explicativa aplicou-se a metodologia SHAP (*Shapley Additive exPlanations*), que é baseada na teoria dos jogos e foi adaptada para explorar as variáveis explicativa dentro dos algoritmos de *machine learning* (LUNDBERG; LEE, 2017). Essa técnica é fundamental na interpretação e explicação, pois ela fornece uma abordagem sistemática e consistente para medir a importância relativa das características ou variáveis explicativas em um modelo. A sua aplicação se deu em dois modelos de *machine learning* que apresentaram melhores performance em relação aos demais. Para isso, foi necessário utilizar dos pacotes fastshap (GREENWELL, 2021) e o shapviz (MAYER, 2023), o primeiro gera as

explicações em cada variável preditora, enquanto o segundo gera a visualização gráfica a partir da explicações gerada no primeiro.

2.6 Aplicação

Para aplicação, foram adquiridas três imagens do SAR que correspondiam ao mesmo dia de passagem do Sentinel-2, o qual possui um sensor óptico que permite a geração de imagens de cor verdadeira, essas imagens foram utilizadas para verificar a presença de nuvens. A área em questão está localizada nas coordenadas centrais -18.55 sul de latitude e -50.45 oeste de longitude, abrangendo aproximadamente 400 hectares de cultivo de cana-de-açúcar na safra 2021/2022, em sistemas com irrigação por gotejamento subsuperficial. Além disso, a área possui uma estação meteorológica, cujos dados foram analisados no dia de passagem dos satélites. Com base nos dados do sensor embarcado no Sentinel-1 e outras variáveis, aplicou-se o modelo de *machine learning* com melhor desempenho. É importante mencionar que essa área não foi utilizada no treinamento ou teste dos modelos, pois não havia dados disponíveis para o mesmo dia de passagem do Sentinel-1 e dos satélites Landsat 8 e 9. Também, por esse motivo, optou-se por utilizar o Sentinel-2 para obter a imagem em cor verdadeira.

3 Resultados e Discussão

3.1 Desempenho da seleção das variáveis e dos algoritmos treinados

Na Figura 4 é possível observar as métricas estatísticas gerada durante o processo da seleção das variáveis via RFE (*recursive feature elimination*) após remover autocorrelação acima de 95%.

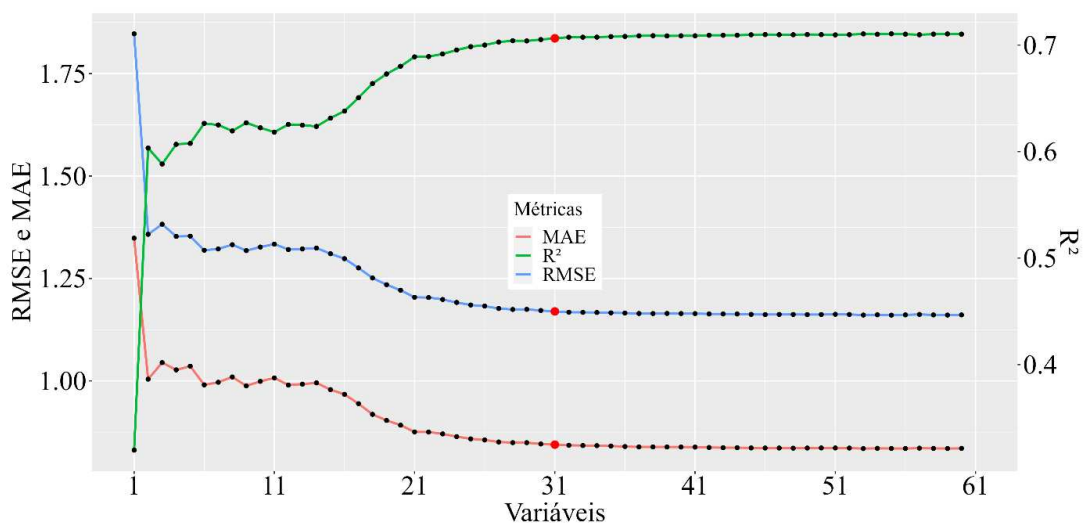


Figura 4. Número de variáveis selecionadas para o treino por meio do RFE.

A aplicação do RFE no conjunto de dados das 60 variáveis, que restaram após a remoção da autocorrelação, selecionou 31 variáveis, as quais sugerem que elas são mais relevantes para prever a evapotranspiração atual na cana-de-açúcar. Para essa seleção, o RFE, aplica um processo retroativo, aloca importâncias às variáveis disponíveis e, por fim, remove as menos relevantes no processo (DUARTE *et al.*, 2022).

Ao analisar a Tabela 4 em conjunto com Figura 4, percebe-se que até 14ª variável há presença das preditoras provenientes do sensor SAR, astronômicas, topográfica e das bioclimáticas, as quais não foram o suficiente para atingir os melhores resultados das métricas. No entanto, ao ser adicionadas as variáveis de solo juntamente com as variáveis meteorológicas advindas do NASA Power as métricas alcançam seus melhores valores, sendo próximo de 0.7 para R^2 e inferior a 1.25 para o MAE e RMSE.

Tabela 4. Variáveis explicativas selecionadas no processo de RFE

Número	Variáveis	Número	Variáveis
1	ângulo_horário_nascer_sol	17	argila_5_15
2	Sigma0_VH	18	bio08
3	DPSVI	19	dai_5_Rs
4	bio02	20	dai_4_Tn
5	bio04	21	argila_15_30
6	DPSVIm	22	ad_5_15
7	bio18	23	silte_0_5
8	Sigma0_VV	24	dda
9	bio13	25	dai_4_Rs
10	mde	26	areia_5_15
11	DPRVIC	27	dai_2_Tméd
12	CR	28	ad_0_30
13	RVIm	29	ad_15_30
14	bio07	30	dai_5_RH
15	Pol	31	areia_0_5
16	bio03		

mde= modelo digital de elevação; dai=dia ante da imagem; ad=água disponível

Com pode ser visualizado na Tabela 4 as variáveis selecionadas compõem o sistema solo-planta-atmosfera, sendo que para o solo têm-se oito variáveis, para o componente planta mais oito e 12, entre climáticas e meteorológicas, para a atmosfera. Além dessas, também estão presentes o ângulo horário ao nascer do sol (ângulo_horário_nascer_sol), elevação (mde) e dia do ano (dda) que estão relacionadas com a dinâmica da radiação solar, distribuição espacial do clima e sazonalidade climática, respectivamente. Assim, essa diversidade de variáveis previamente selecionadas garante melhor capacidade preditiva para os modelos, principalmente por se utilizar dados de radar, uma vez que suas leituras diretas são limitadas, pois há apenas um comprimento de onda que após processada gera duas polarizações. Diferente, por exemplo, dos

sensores ópticos como sentinel-2 que é possível obter maior número de leituras diretas e utilizá-las como preditoras (COSTA *et al.*, 2023; DOS SANTOS *et al.*, 2022b).

A performance de cada algoritmos após o treino com as 31 variáveis utilizadas para prever a evapotranspiração atual da cana-de-açúcar podem ser visualizadas na Figura 5. Nela, pode ser observado, de modo ranqueado, a evolução do desempenho de cada modelo.

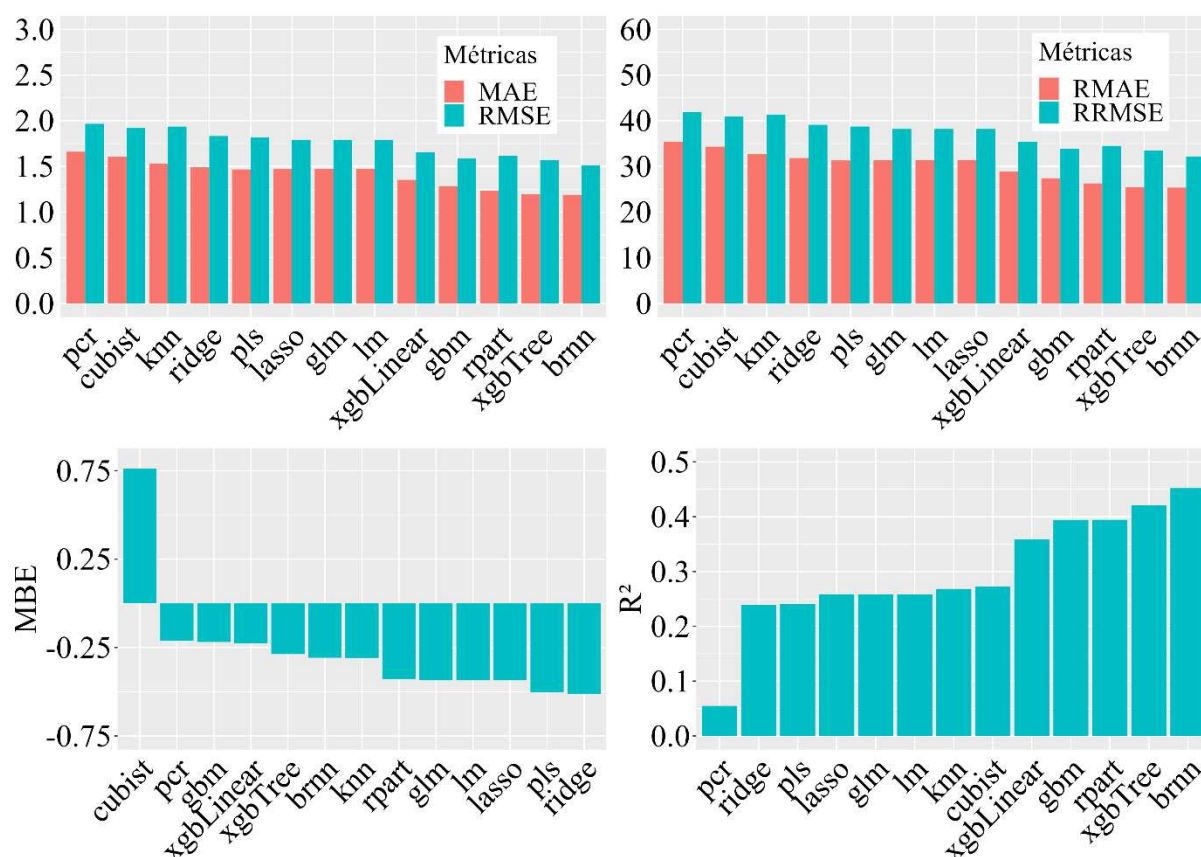


Figura 5. Performance dos 13 algoritmos de *machine learning* destinado ao treino e teste da ET_a.

Nota-se, ao analisar a Figura 5, que o pcr, cubist e knn tiveram os piores desempenhos, onde os valores de RRMSE foram superiores a 40%, já o ridge, pls, lasso, glm e lm apresentaram performance intermediária, com valores de RRMSE abaixo de 40%, mas próximo a ele. Por fim, o XgbLinear, gbm, rpart, XgbTree e brnn foram superiores a todos os demais com seus valores de RRMSE próximo a 35%, porém ainda acima de 30%, o que indica, para o conjunto de dados geral, baixa capacidade preditiva. Ainda, entre os cinco modelos que se destacam frente aos demais, dois deles, o XgbTree e o brnn, sobressaíram, sendo eles os melhores entre os 13 treinados utilizados. Ao observar o MBE, na Figura 5, constata-se que o cubist foi o único modelo a superestimar os valores observados, enquanto os demais subestimaram. O R² seguiu-se o mesmo comportamento das demais métricas para os desempenhos dos modelos, onde há modelos com R² abaixo de 0.1, superior a 0.2 e inferior a

0.3 e modelos superior a 0.3, mas inferior a 0.5, mais uma vez, destacando o XgbTree e o brnn. Observar-se, então, que com o conjunto de dados gerais a predição não apresenta boa correlação com os dados observados, indicando desta forma, que os modelos apresentam baixa capacidade preditiva. Esse fato pode ser explicado devido à complexidade que é estimar a evapotranspiração sendo, ainda, um desafio científico sua adequada quantificação (CASTELLVÍ; SNYDER, 2010; LOOS; GAYLER; PRIESACK, 2007; OCHEGE *et al.*, 2019). Essa complexidade se torna ainda maior quando o objetivo é o proposto neste estudo, ou seja, estimar a evapotranspiração atual sem utilizar estações meteorológica em campo e fazendo uso dos dados de radar que é limitado em função do número de informação que ele pode adquirir da cultura.

Na Figura 6 pode ser analisado o comportamento dos dados entre o predito e observado para os dois melhores modelos. Constata-se que o brnn apresenta, de modo geral, linha de tendencia (linha preta) paralela a reta 1:1 (linha vermelha) onde os valores observados em baixo e em altos valores superestimam os preditos, por outro lado o XgbTree os preditos superestimam os valores observados quando é superior a 6 mm.

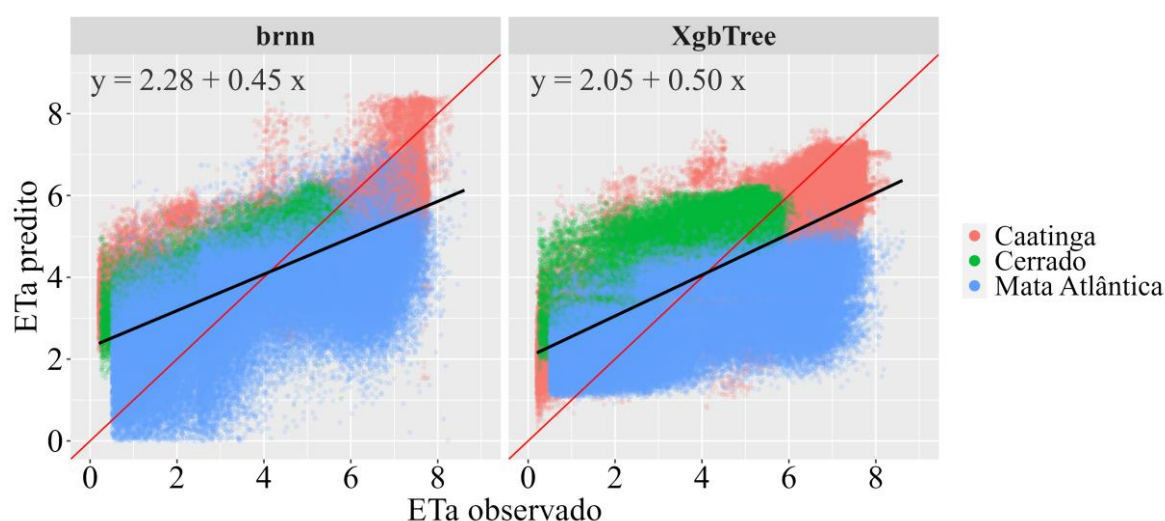


Figura 6. Predito vs observado da ET_a dos dois melhores modelos de *machine learning* para o conjunto de dado geral destinado ao teste.

Ainda, é possível notar, na Figura 6, que ambos os modelos nas três localidades de coleta a dispersão são elevados ocasionado, principalmente, pela baixa capacidade preditiva dos modelos, haja visto que as métricas estatísticas (Figura 5) não apresentou bons resultados, tanto pelo número reduzido de variáveis associada a planta como a dificuldade de estimar a evapotranspiração em áreas de cultivos como explica Ke *et al.* (2017).

A partir da Figura 7 pode ser observado, com maior detalhamento, o comportamento individual de cada conjunto de dados com os respectivos modelos de *machine learning*.

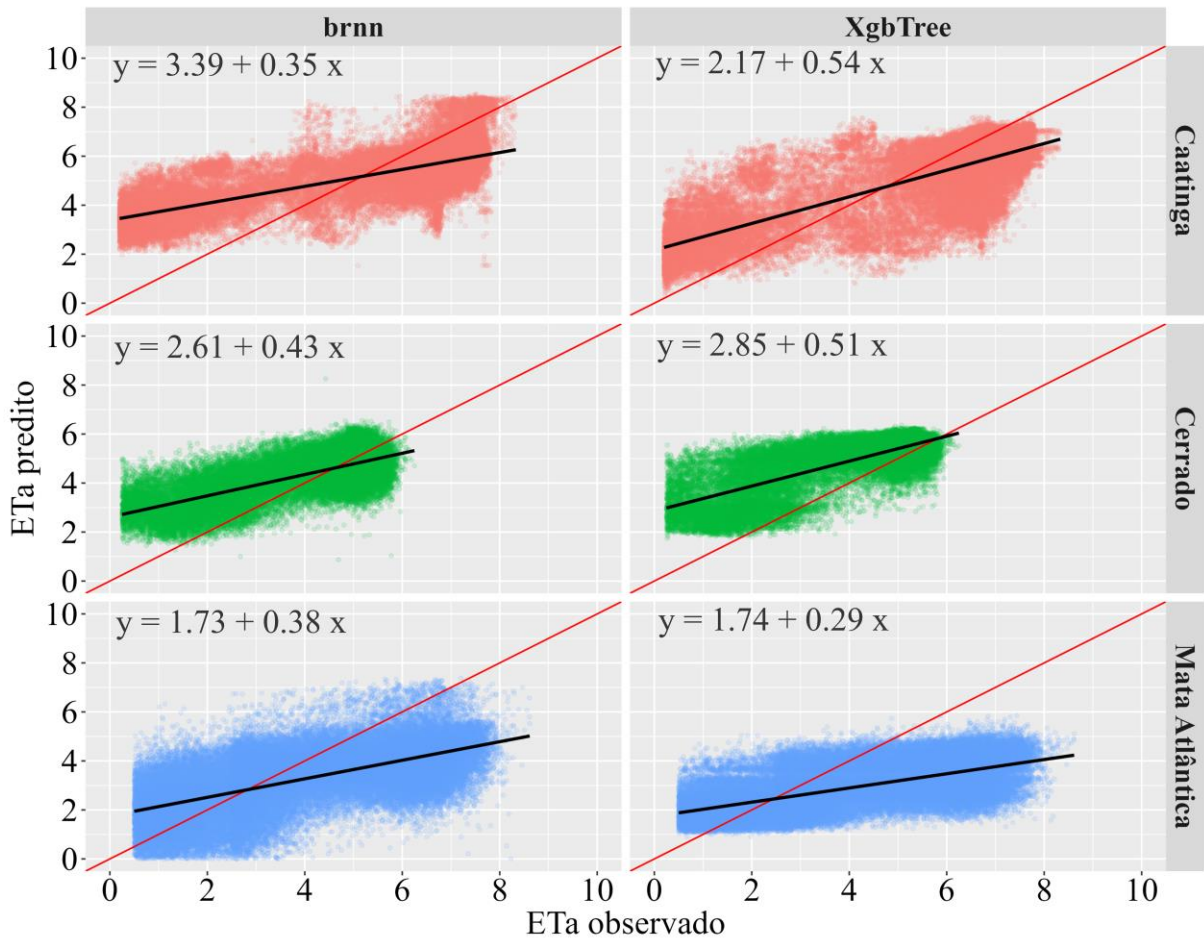


Figura 7. Predito vs observado da ET_a dos dois melhores modelos de *machine learning* em cada bioma destinado ao teste.

Na Figura 7, é evidente que para o modelo brnn, os dados coletados na área de cultivo Caatinga apresentam dois comportamentos distintos. Valores menores que 5.0 mm são superestimados pelo modelo, enquanto valores acima desse limiar são subestimados. Além disso, é observado um platô de saturação quando os valores observados estão próximos a 8.0 mm. Por outro lado, o modelo XgbTree, para os mesmos dados, apresenta um comportamento de linha de tendência mais próximo à reta 1:1, de forma mais sutil. Para a área de cultivo no Cerrado, a linha de tendência da predição do modelo brnn também superestima os valores observados em torno de 5.0 mm. No caso do modelo XgbTree, ele superestima os valores observados em todos os dados. Por fim, para os dados coletados na Mata Atlântica, tanto o modelo brnn quanto o XgbTree subestimam os valores observados. Ambos os modelos têm dificuldade em estimar valores acima de 5.0 mm. Enquanto os valores observados atingem

próximo a 8.0 mm, a predição dos modelos atinge no máximo 6.5 mm para o brnn e não ultrapassa 6.0 mm para o XgbTree.

A discrepância nos dados do bioma Mata Atlântica em relação aos demais pode estar relacionada à baixa variabilidade das variáveis explicativas, principalmente climatológicas e meteorológicas, nesse bioma. Ao contrário do Cerrado e da Caatinga, que apresentam alta variabilidade nessas variáveis. Portanto, os algoritmos possivelmente aprenderam com os dados de maior variabilidade nos biomas da Caatinga e do Cerrado, mas ficaram limitados a essa baixa variabilidade na Mata Atlântica, o que resultou em baixa capacidade de predição para valores acima de 6.5 mm nesse bioma.

Os resultados estatísticos apresentados na Tabela 5 permite melhor compreensão dos dados da Figura 7, onde é observado que o brnn foi superior em todos os biomas em relação ao XgbTree.

Tabela 5. Estatística do conjunto de dados destinado ao teste para os modelos brnn e XgbTree nos três biomas

Bioma	\bar{x} obs mm	\bar{x} pred mm	dif %	R^2	MAE mm	RMAE %	RMSE mm	RRMSE %	MBE mm
brnn									
Mata Atlântica	4.15	3.18	20.01	0.41	1.49	35.79	1.81	43.52	-0.83
Caatinga	5.51	5.30	4.00	0.53	1.40	25.10	1.64	29.78	-0.22
Cerrado	4.40	4.52	2.71	0.47	0.71	16.02	0.95	21.57	0.12
XgbTree									
Mata Atlântica	4.15	2.94	29.17	0.41	1.65	39.80	2.04	49.11	-1.21
Caatinga	5.51	5.15	6.29	0.63	1.14	20.61	1.42	25.83	-0.35
Cerrado	4.40	5.10	15.79	0.59	0.80	18.00	1.10	24.80	0.70

Com base na Tabela 5, fica evidente que os resultados para os dados coletados na área do bioma Mata Atlântica são insatisfatórios. Mesmo para o melhor modelo (brnn), a média prevista é aproximadamente 20% inferior à observada, e os valores de RMSE são altos, atingindo 1.81 mm. Isso resulta em um RRMSE de 43.52%, que ultrapassa o limite de 30% considerado como indicador de incapacidade de previsão do modelo. No entanto, apesar do erro elevado, há uma correlação moderada ($R^2 = 0.41$, $r = 0.64$), o que indica que o erro está principalmente associado à baixa capacidade preditiva do modelo para valores acima de 6.5 mm. Os dados da área da Caatinga apresentaram uma diferença média entre observado e predito muito próxima (4% de diferença) para o melhor modelo, o que representa um bom resultado. No entanto, ao observar as métricas estatísticas, nota-se que o R^2 foi de 0.53 ($r = 0.73$), mas o RMSE e o RRMSE foram de 1.64 e 29.78, respectivamente, o que está próximo de 30%. O

modelo brnn para o conjunto de dados do Cerrado teve o melhor desempenho, com uma diferença média de apenas 2.71% entre observado e previsto. Os erros RMSE e RRMSE foram de 0.95 e 21.57%, respectivamente, mas o R^2 foi de 0.47 ($r = 0.69$), inferior ao da Caatinga. Observa-se, com essa abordagem, que o desempenho do modelo varia de acordo com a localidade onde os dados foram coletados. Apesar da utilização de diversas variáveis, incluindo variáveis de posição geográfica, o modelo apresenta baixa capacidade de generalização, sendo, nesses casos mais complexo apropriado a construção de modelos independente para cada localidade ou região.

3.2 Shapley Additive explanations

A partir da Figura 8 pode ser observado a amplitude e a distribuição do impacto (8a) das 15 variáveis explicativas mais importantes na predição evapotranspiração atual (ET_a) da cana-de-açúcar, bem como o seu ranqueamento (8b) para o modelo brnn. Cada ponto nesta figura representa um valor Shapley para cada variável de entrada e é ordenada em valores baixos (azul) a alto (amarelo).

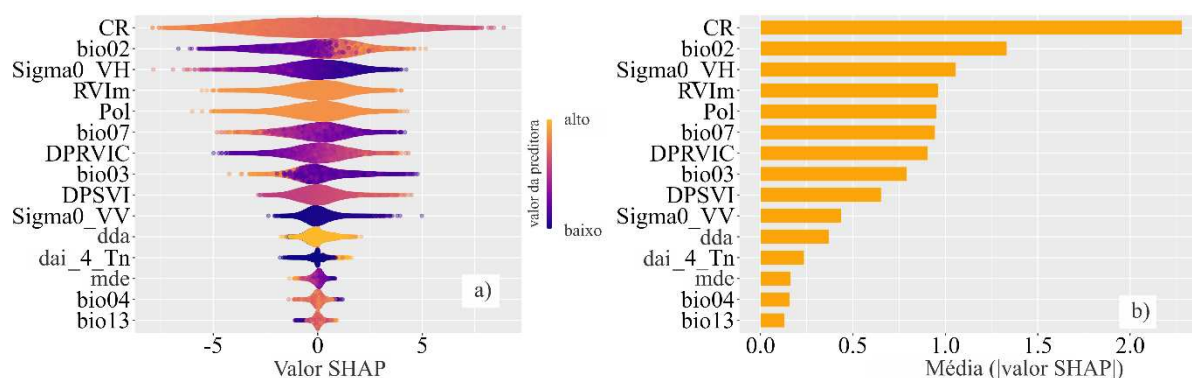


Figura 8. Características das 15 variáveis que mais contribuíram para o modelo brnn (8a) e ranqueamento das mais importantes em valores absolutos (8b).

Ao analisar a Figura 8a para o modelo brnn, observa-se três comportamentos diferentes nas variáveis. Algumas variáveis têm um impacto positivo na evapotranspiração quando seus valores são baixos, outras têm impacto positivo quando os valores são altos, e outras mantêm um efeito constante tanto para baixos quanto para altos valores. Verifica-se que, quanto maiores os valores das variáveis CR, Sigma0_VH, bio07, bio03, mde (modelo digital de elevação) e bio04, menor é o valor de Shapley e menor é seu impacto na ET_a da cana-de-açúcar.

A variável CR, que é derivada da razão entre os retroespalhamentos VH e VV, está relacionada à biomassa e ao índice de área foliar. O aumento da biomassa da cultura resulta em um aumento de CR, e essa variável tem sido utilizada como uma alternativa ao NDVI para acompanhar o crescimento e desenvolvimento de lavouras de milho, canola, cebola e trigo de

inverno (KHABBAZAN *et al.*, 2019; SHORACHI; KUMAR; STEELE-DUNNE, 2022). Alguns estudos também mostraram uma correlação positiva entre CR e conteúdo de água na planta, embora essa correlação seja baixa (KHABBAZAN *et al.*, 2019; VREUGDENHIL *et al.*, 2020, 2018). No entanto, os resultados obtidos no modelo brnn são opostos aos encontrados nesses estudos. O brnn entende que essa variável é contrária ao conteúdo de água, pois um aumento em seu valor resulta em uma redução na evapotranspiração. Porém, um estudo realizado por Den Besten *et al.* (2021) na cultura da cana-de-açúcar encontrou uma correlação negativa entre CR e a produtividade de TCH (Tonelada de Colmo por Hectare), ou seja, maiores produtividades e maior biomassa, para essa cultura, estão associadas a menores valores de CR. Esses autores mencionam que o acúmulo de sacarose dentro da planta reduz o teor de água livre, o que influencia significativamente a constante dielétrica da vegetação, reduzindo o retrospalhamento e, conseqüentemente, os valores de CR. Enquanto os sensores ópticos, que podem calcular o NDVI, capturam o dossel das plantas e têm seus principais valores relacionados à condição da superfície foliar, o radar de banda-C não se limita apenas à superfície, mas consegue penetrar na vegetação, gerando dados de toda a planta (DEN BESTEN *et al.*, 2021; VREUGDENHIL *et al.*, 2020). Dessa forma, os dados do radar captam informações tanto das folhas quanto dos colmos da cana-de-açúcar, sendo que nos colmos está armazenada a sacarose, ocupando o espaço da água livre. Por esse motivo, o modelo brnn ponderou que essa variável gera um impacto negativo na ET_a . O mesmo entendimento para a variável CR se aplica à variável Sigma0_VH, com estudos publicados mostrando uma correspondência entre seu crescimento e desenvolvimento e a cultura da cana-de-açúcar (SHORACHI; KUMAR; STEELE-DUNNE, 2022; VREUGDENHIL *et al.*, 2020). Mais uma vez, Den Besten *et al.* (2021) usam o mesmo argumento da constante dielétrica, que é reduzida com o aumento da sacarose e a diminuição da água livre nos colmos.

Pode ser observado, também na Figura 8a, que a variável bio07, bio03 e bio04, provenientes do WorldClim e fixas no tempo, têm influência na caracterização bioclimática da região. A variável bio07 corresponde à amplitude anual térmica e é um indicador do clima e regime térmico de uma determinada área. Valores elevados indicam verões quentes e invernos rigorosos, o que resulta em uma redução na ET_a . A variável bio03 é uma variável isotérmica, expressa em porcentagem, calculada pela razão entre a faixa diurna média e a amplitude anual térmica. Ela indica a variação térmica e distribuição de temperaturas ao longo do ano de um determinado local. Valores altos mostram temperaturas mais equilibradas, e esperava-se que essa variável tivesse um impacto positivo na ET_a . No entanto, o impacto negativo observado no modelo pode estar relacionado a interações com outras variáveis, que contribuiriam para o efeito

contrário ao esperado. A distribuição dos valores de Shapley para a variável bio04 mostra uma influência positiva na ET_a quando os valores são extremamente baixos, o que é compreensível, já que essa variável representa o desvio padrão da sazonalidade da temperatura. Uma menor sazonalidade tem um impacto positivo na evapotranspiração de um local, pois indica menores valores extremos. A variável mde está relacionada à elevação do local, e seus valores mais altos têm um impacto negativo na evapotranspiração. Isso ocorre porque, com o aumento da elevação, há uma redução na temperatura e pressão atmosférica, o que diminui o potencial evapotranspirativo da atmosfera.

As variáveis bio02, DPRVIC, dai_4_tn e bio03 apresentam comportamento oposto às variáveis mencionadas anteriormente, ou seja, em altos valores, elas têm um impacto positivo na evapotranspiração atual da cana-de-açúcar. A bio02 corresponde às variações médias das temperaturas diárias ao longo de um mês, e quanto maior seu valor, indica uma maior diferença entre as temperaturas máxima e mínima. Assim, um maior afastamento entre essas temperaturas favorece o aumento da evapotranspiração. De fato, essa é uma das bases para o cálculo da evapotranspiração de referência proposta por Hargreaves and Samani (1985), que mostra resultados maiores em ambientes com maior amplitude térmica. A variável DPRVIC é gerada a partir dos dados de retroespalhamento e tende a ter altos valores em copas de vegetação bem desenvolvidas, como apontado por Bhogapurapu *et al.* (2022). Portanto, sua relação com a evapotranspiração é positiva, uma vez que uma vegetação bem desenvolvida indica um maior potencial de transpiração em comparação com estágios fenológicos iniciais. A variável dai_4_tn, que representa a temperatura mínima 4 dias antes da passagem do satélite, apresenta maiores valores quando há temperaturas mínimas mais altas. Isso resultará em uma maior evapotranspiração, pois temperaturas mínimas muito baixas afetam os processos metabólicos das plantas e, conseqüentemente, reduzem sua capacidade transpirativa. Ademais, temperaturas mais baixas reduzem a demanda evapotranspirativa da atmosfera. A variável bio13 indica a precipitação do mês mais chuvoso. Quanto maior a precipitação, mais água estará disponível para o processo evapotranspirativo. Portanto, seus maiores valores influenciam positivamente na evapotranspiração.

Por fim, as variáveis RVIm, Pol, DPSVI, Sigma_VV e dda (dia do ano) se destacaram como variáveis importantes (Figura 8b). No entanto, o brnn não compreende completamente seus maiores ou menores valores. Isso pode ser explicado pelo fato de que essas variáveis podem ter efeitos diferentes em diferentes regiões do conjunto de dados. Por exemplo, uma variável pode estar fortemente correlacionada com a variável resposta em algumas partes do conjunto de dados, o que levaria a um efeito positivo. No entanto, em outras partes do conjunto

de dados, a correlação pode ser negativa, resultando em um efeito negativo. Como resultado, a contribuição da variável pode modificar em diferentes regiões do conjunto de dados, levando a um impacto positivo e negativo ao mesmo tempo. Isso pode explicar a dificuldade do modelo brnn em interpretar corretamente essas variáveis e seus efeitos na evapotranspiração da cana-de-açúcar.

Na Figura 9 é possível visualizar a distribuição dos valores Shapley (Figura 9a) e a importância das 15 mais importantes variáveis do modelo treinado XgbTree.

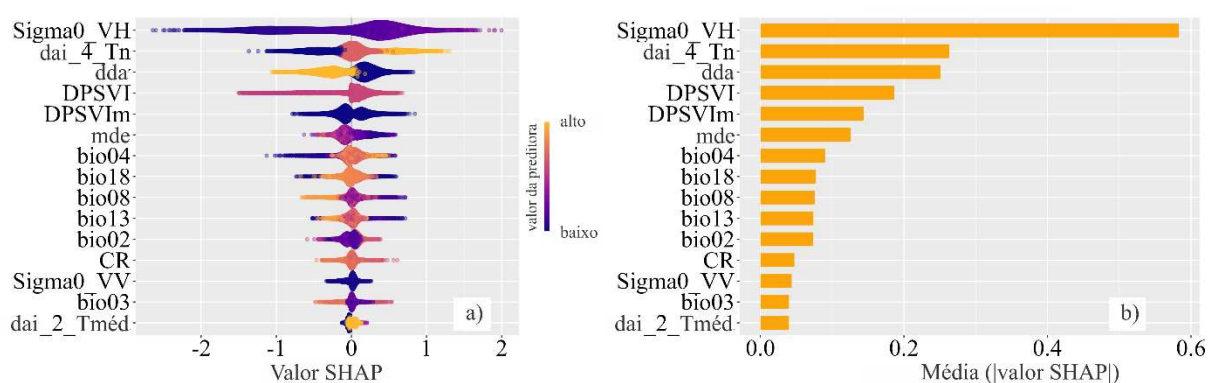


Figura 9. Características das 15 variáveis que mais contribuíram para o modelo XgbTree (9a) e ranqueamento das mais importantes em valores absolutos (9b).

O XgbTree apresenta três grupos de comportamento das variáveis: 1) em alto valores influenciando positivamente na evapotranspiração; 2) em alto valores passa a ter influencia negativamente; 3) que influenciam negativamente e positivamente tanto em alto quanto em baixos valores. Para o primeiro grupo têm-se as variáveis Sigma0_VH, dai_4_Tn e dai_2_Tméd, enquanto para o segundo são as dda, mde, e bio8 e para o terceiro bio02, bio03, bio04, bio18, bio13, DPSVI, DPSVIm, CR, Sigma_VV.

Ao analisar a variável Sigma0_VH, nota-se um comportamento totalmente oposto aos valores de Shapley para o modelo brnn. Nesse caso, valores maiores de Sigma0_VH têm um impacto positivo na predição da evapotranspiração. O mesmo comportamento é observado para a variável dai_4_Tn no modelo brnn. Por outro lado, a temperatura média dois dias antes da passagem do satélite (dai_2_Tméd) apresenta um comportamento oposto. Valores baixos dessa variável tendem a ter um impacto negativo na evapotranspiração, mas à medida que sua média aumenta, há um aumento positivo na evapotranspiração. Isso ocorre porque uma baixa temperatura média reduz a demanda evapotranspirométrica da atmosfera, enquanto temperaturas elevadas podem influenciar o fechamento estomático, reduzindo a transpiração. Valores intermediários são penalizados, pois propiciam evapotranspirações mais elevadas no algoritmo XgbTree. No caso da variável dda, quanto menor o valor, o modelo XgbTree entende

que a evapotranspiração será maior. Isso ocorre porque valores baixos de *dda* coincidem com períodos de início do ano, quando a demanda evapotranspirométrica atmosférica é mais alta no hemisfério sul, associada a períodos de intenso crescimento da cana-de-açúcar. Geralmente, entre outubro e setembro ocorre o plantio ou rebrota da cultura, estando em estágio inicial, portanto, altos valores de *dda* influenciam negativamente na evapotranspiração. As variáveis *Sigma0_VH*, *dai_4_Tn*, *dda* e *mde* foram as mais importantes para o modelo, como evidenciado pelos valores absolutos apresentados na Figura 9b. A variável *mde* segue o mesmo padrão encontrado no modelo *brnn*, onde valores baixos favorecem a evapotranspiração. A variável *bio08* (temperatura média do trimestre mais úmido) influencia a evapotranspiração de forma inversa, ou seja, em altos valores, ocorre uma menor evapotranspiração. Isso acontece porque a temperatura durante o trimestre mais úmido coincide com a época de verão nas áreas onde os dados foram coletados. Nessa época, as temperaturas são mais altas, o que pode afetar a condutância estomática devido ao aumento do déficit de pressão de vapor. As variáveis *bio02*, *bio03*, *bio04*, *bio08*, *bio13* e *bio18* são variáveis de caracterização climática e não apresentam um comportamento linear em relação à evapotranspiração. No caso das variáveis *bio04*, *bio13* e *bio18*, valores baixos têm um impacto negativo e, à medida que aumentam, passam a ter um efeito positivo até atingir um ponto de saturação, onde valores baixos voltam a ter um impacto negativo. O oposto ocorre com as variáveis *bio02* e *bio03*, onde valores altos têm um impacto negativo e, à medida que reduzem, há um aumento positivo na evapotranspiração até atingir um ponto de saturação. Após esse ponto, valores altos novamente passam a ter um impacto positivo. Esses comportamentos podem ser explicados pelas interações entre as variáveis quando são atribuídos pesos pelo algoritmo, o que pode resultar em mudanças no efeito da variável dependendo dos valores das outras variáveis presentes no modelo. Além disso, os pontos de saturação observados nessas variáveis, conforme mostrado na Figura 9a, também podem contribuir para esses padrões.

É possível observar nas variáveis *DPSVI*, *DPSVIm*, *CR* e *Sigma_VV*, oriundas de radar, que elas mantêm valores constantes ao longo da distribuição, seja alto ou baixo, mas com impacto variando na evapotranspiração. Apesar de apresentarem valores constantes, duas delas, *DPSVI* e *DPSVIm*, estão entre as cinco melhores preditoras, conforme mostrado na Figura 9b. Por outro lado, as variáveis *CR* e *Sigma0_VV* têm pouca influência na predição da evapotranspiração no modelo *XgbTree*.

Ao observar os comportamentos das variáveis para os modelos *brnn* (Figura 9a) e *XgbTree* (Figura 9a), é perceptível uma mudança no padrão de comportamento. Em outras palavras, uma variável pode apresentar altos valores e influenciar negativamente na

evapotranspiração em um modelo, enquanto no outro modelo ocorre o contrário, como aconteceu com a Sigma0_VH . Portanto, as diferenças nos valores de Shapley entre os modelos brnn e XgbTree podem ser resultado da combinação de fatores relacionados à estrutura e abordagens dos modelos na atribuição de pesos às variáveis. Além disso, uma variável pode influenciar a outra de maneira diferente em cada um dos modelos, o que também gera uma relação não linear entre as variáveis e a ET_a . Isso pode dificultar, em alguns casos, a interpretação das variáveis. Dessa forma, é importante considerar essas diferenças ao interpretar os resultados e não presumir uma relação direta de causalidade apenas com base nos valores de Shapley.

3.3 Aplicação

Após a aplicação, é evidente a potencialidade do sensoriamento remoto por radar e *machine learning* na predição da ET_a quando há presença de nuvens na área de interesse, como demonstrado na Figura 10.

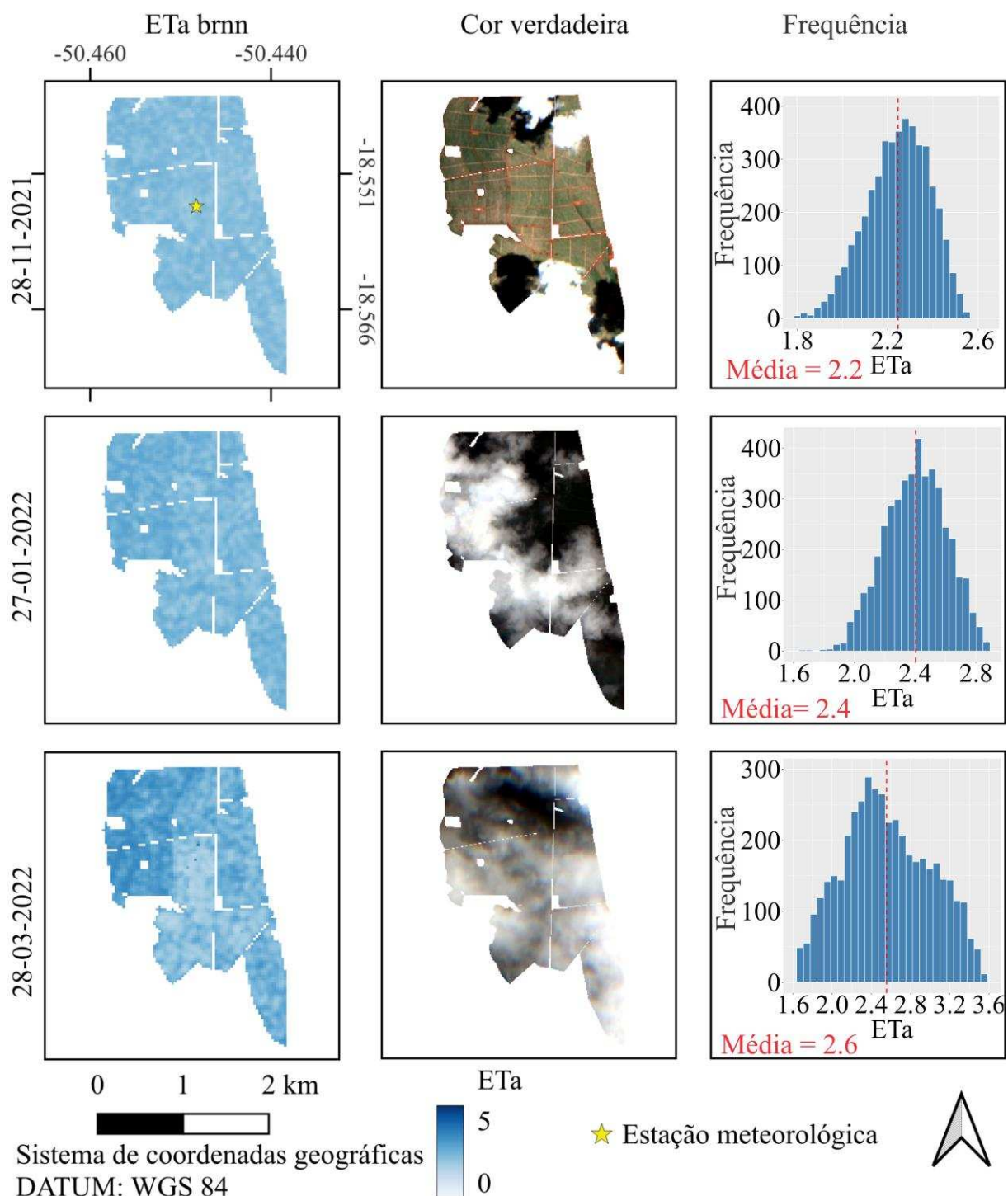


Figura 10. Evapotranspiração atual na cana-de-açúcar em área com presença de nuvens.

Na Figura 10, pode-se observar que, na data de 28-11-2021, há uma presença parcial de nuvens sobre a área, o que impossibilitaria a utilização na área total com sensoriamento óptico por satélite. No entanto, isso não impediu a estimativa da evapotranspiração por meio dos modelos de *machine learning* e dados de radar e covariáveis. Já para as duas datas subsequentes, seria completamente inviável o uso de sensores ópticos por satélite devido à cobertura quase total da área e ao sombreamento causado pelas nuvens. Porém, mesmo com essa limitação, foi

possível estimar a ET_a nessa área de interesse, uma vez que, como mencionado anteriormente, os dados de radar têm a capacidade de penetrar nas nuvens e interagir com os alvos abaixo delas, permitindo assim a estimativa da quantidade de água perdida através da ET_a .

Além disso, é possível observar que as três datas apresentam baixa ET_a , o que pode ser explicado por dois fatores. O primeiro fator se aplica à data de 28-11-2021, na qual as plantas de cana-de-açúcar estavam em fase inicial de crescimento, resultando em uma demanda transpiratória muito menor em comparação com plantas adultas. Apesar disso, a demanda evapotranspirativa da atmosfera estava elevada em relação às outras duas datas, conforme observado pelas variáveis, principalmente a DPV, na Tabela 6.

Tabela 6. Condições meteorológicas do dia que ocorreu passagem dos satélites Sentinel-1 e 2

Variáveis	28-11-2021	27-01-2022	28-03-2022
Rs (W/m ²)	314.00	252.00	129.00
T _{méd} (°C)	27.97	23.80	23.35
UR (%)	71.76	89.56	92.67
DPV (kPa)	1.34	0.40	0.28
Prec (mm)	0.40	31.60	13.80

Rs = Radiação solar incidente; T_{méd} = Temperatura média; UR = Umidade relativa; DPV = Déficit de pressão de vapor; Prec. = Precipitação pluviométrica.

O segundo motivo está relacionado às outras duas datas, em que ocorreram precipitações pluviométricas, modificando a atmosfera para condição amena, reduzindo drasticamente a demanda evapotranspirativa. Isso pode ser observado na Tabela 6, a partir da redução da radiação, temperatura e aumento da umidade relativa, o que proporcionou menor DPV e, conseqüentemente, menor capacidade da atmosfera em favorecer uma maior evapotranspiração atual da cultura.

4 Conclusão

Algoritmos de *machine learning* treinados com múltiplas fontes de dados, incluindo dados do sensor SAR, possibilitam a estimativa da ET_a em áreas com presença de nuvens e na ausência de dados de sensoriamento remoto óptico. Entretanto, apesar do potencial promissor, esses algoritmos apresentam baixa capacidade preditiva devido à complexidade da variável ET_a e às limitações dos dados provenientes dos radares com banda C.

Entre os algoritmos testados, o brnn e o XgbTree se destacaram em relação aos demais no conjunto de dados gerais. O brnn obteve um R^2 de 0.45 e um RMSE de 1.5, enquanto o XgbTree obteve um R^2 de 0.46 e um RMSE de 1.6. No entanto, para cada bioma seu comportamento foi diferente. Na Mata Atlântica, o brnn foi ligeiramente superior, com um R^2 de 0.41 e um RMSE de 1.8, em comparação ao XgbTree, que obteve um R^2 de 0.41 e um RMSE

de 2.0. Já na Caatinga, o XgbTree foi superior, com um R^2 de 0.63 e um RMSE de 1.42, em relação ao brnn, que obteve um R^2 de 0.53 e um RMSE de 1.64. No Cerrado, o XgbTree apresentou um R^2 de 0.59 e um RMSE de 1.10, sendo superior ao brnn, que obteve um R^2 de 0.47 e um RMSE de 0.95, mas teve um melhor RMSE.

O método *Shapley Additive Explanations* foi fundamental para compreender o comportamento das variáveis em relação à ET_a e identificar aquelas de maior importância. Com essa metodologia, ficou evidente que a CR, bio02 e Sigma0_VH foram as três variáveis mais relevantes para o brnn, enquanto a Sigma0_VH, dai_4_Tn e dda se destacaram como as mais importantes para o XgbTree.

3 Conclusões gerais

Os algoritmos de *machine learning* e os dados de sensoriamento remoto, tanto ativo quanto passivo, juntamente com outras fontes de informação, mostraram-se satisfatórios para estimar a evapotranspiração atual da cana-de-açúcar, oferecendo alternativas aos métodos tradicionalmente utilizados.

O uso dos dados do sensor MSI e técnicas de *machine learning* para prever apenas a fração evapotranspirativa apresenta resultados estatisticamente superiores em comparação com a estimativa direta da ET_a . Portanto, no primeiro capítulo, que teve como objetivo estimar a fração evapotranspirativa, o algoritmo XgbLinear foi superior, com R^2 de 0.91 e RMSE de 0.10.

No segundo capítulo, foi possível determinar diretamente a ET_a da cana-de-açúcar, eliminando a necessidade de estações meteorológicas em campo para calcular a evapotranspiração de referência e multiplicar pelos coeficientes de planta e solo. Dos 13 algoritmos testados, observou-se que o brnn e o XgbLinear apresentaram estatísticas superiores em relação aos demais, com R^2 de 0.73 e 0.74, respectivamente, e RMSE de 1.10 e 1.25, nessa ordem. Além disso, a utilização da técnica SHAP permitiu compreender a influência de cada variável explicativa na variável resposta dentro desses modelos, sendo uma metodologia que deve ser amplamente aplicada em todos os modelos de *machine learning* considerados como “*black-box*”.

O terceiro capítulo alcançou seus objetivos ao treinar, testar e aplicar algoritmos de *machine learning* a partir de dados provenientes de sensores ativos, juntamente com outras fontes de dados. Esse capítulo evidencia que é possível estimar a evapotranspiração por sensoriamento remoto orbital, mesmo na presença de nuvens e sem a necessidade de estações meteorológicas em campo. Nesse contexto, os algoritmos brnn ($R^2 = 0.45$ e RMSE = 1.5) e XgbTree ($R^2 = 0.46$ e RMSE = 1.6) demonstraram superioridade em relação aos demais, embora seus resultados tenham sido inferiores aos obtidos nos primeiros e segundos capítulos.

Outra observação importante que aborda os três capítulos é que à medida que aumentava a complexidade da estimativa, os modelos apresentavam uma queda no desempenho. No primeiro capítulo, que envolvia a estimativa mais simples, o melhor resultado de R^2 foi de 0.91. No segundo capítulo, com maior complexidade, o melhor R^2 foi de 0.74. Por fim, no terceiro capítulo, que possuía uma complexidade mais elevada devido às limitações das bandas espectrais, o melhor R^2 foi de 0.45. Assim, nota-se queda na qualidade da estimativa da evapotranspiração atual quando a necessidade eleva seu grau de dificuldade. Apesar desses resultados, de modo geral, a tese atingiu seu objetivo principal, que era estimar a ET_a da cana-de-açúcar utilizando técnicas de *machine learning* com dados de sensoriamento remoto, mesmo

sem a banda termal, sem a necessidade de instalar estações meteorológicas em campo e na presença de nuvens.

Apesar dos avanços encontrados, pode-se melhorar as estimativas da evapotranspiração atual para agricultura, principalmente irrigada, a partir de integração entre diversos modelos, combinado estimativas via clima, sensoriamento remoto passivo e ativo, e via solo como ponto de checagem e correção dos dados.

Referências

- ABOU ALI, A.; BOUCHAOU, L.; ER-RAKI, S.; HSSAÏSSOUNE, M.; BROUZIYNE, Y.; EZZAHAR, J.; KHABBA, S.; CHAKIR, A.; LABBACI, A.; CHEHBOUNI, A. Assessment of crop evapotranspiration and deep percolation in a commercial irrigated citrus orchard under semi-arid climate: Combined Eddy-Covariance measurement and soil water balance-based approach. **Agricultural Water Management**, v. 275, n. October 2022, p. 107997, jan. 2023. Disponível em: <<https://doi.org/10.1016/j.agwat.2022.107997>>.
- ADAB, H.; MORBIDELLI, R.; SALTALIPPI, C.; MORADIAN, M.; GHALHARI, G. A. F. Machine learning to estimate surface soil moisture from remote sensing data. **Water (Switzerland)**, v. 12, n. 11, p. 1–28, 2020.
- ALLEN, R. G., PEREIRA, L. S., RAES, D., SMITH, M. guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper Crop by. **Remote Sensing of Environment**, v. 300, n. No. 56, p. 173, 2006. Disponível em: <http://www.scsocourt.org/complexcivil/105CV049053/volume3/172618e_5xAGWax8.pdf>.
- ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. Crop evapotranspiration: Guidelines for computing crop water requirements. **FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations, Rome, Italy**, 1998. Disponível em: <<https://www.fao.org/3/x0490e/x0490e00.htm>>.
- ALLEN, R. G.; TASUMI, M.; TREZZA, R. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Applications. **Journal of Irrigation and Drainage Engineering**, v. 133, n. 4, p. 395–406, ago. 2007. Disponível em: <<http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9437%282007%29133%3A4%28395%29>>.
- ALLEN, R. G.; WALTER, I. A.; ELLIOTT, R.; HOWELL, T.; ITENFISU, D.; JENSEN, M. **The ASCE standardized reference evapotranspiration equation**. [s.l.: s.n.]. Disponível em: <<https://ascelibrary.org/doi/book/10.1061/9780784408056>>.
- ALLEN, R.; IRMAK, A.; TREZZA, R.; HENDRICKX, J. M. H.; BASTIAANSEN, W.; KJAERGAARD, J. Satellite-based ET estimation in agriculture using SEBAL and METRIC. **Hydrological Processes**, v. 25, n. 26, p. 4011–4027, 30 dez. 2011. Disponível em: <<http://doi.wiley.com/10.1002/hyp.8408>>.
- ALTHOFF, D.; DIAS, S. H. B.; FILGUEIRAS, R.; RODRIGUES, L. N. ETo-Brazil: A Daily Gridded Reference Evapotranspiration Data Set for Brazil (2000–2018). **Water Resources Research**, v. 56, n. 7, p. 1–16, 2020.
- ALTHOFF, D.; SANTOS, R. A. dos; BAZAME, H. C.; CUNHA, F. F. da; FILGUEIRAS, R. Improvement of Hargreaves–Samani Reference Evapotranspiration Estimates with Local Calibration. **Water**, v. 11, n. 11, p. 2272, 30 out. 2019. Disponível em: <<https://www.mdpi.com/2073-4441/11/11/2272>>.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES GONÇALVES, J. L.; SPAROVEK, G. Köppen’s climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.
- ANA. **Atlas irrigação: uso da água na agricultura irrigada**. 2. ed. Brasília: ANA, 2021.
- ANDERSON, M. A Two-Source Time-Integrated Model for Estimating Surface Fluxes Using Thermal Infrared Remote Sensing. **Remote Sensing of Environment**, v. 60, n. 2, p. 195–

216, maio 1997. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0034425796002155>>.

ANDERSON, M. C.; NORMAN, J. M.; MECIKALSKI, J. R.; TORN, R. D.; KUSTAS, W. P.; BASARA, J. B. A Multiscale Remote Sensing Model for Disaggregating Regional Fluxes to Micrometeorological Scales. **Journal of Hydrometeorology**, v. 5, n. 2, p. 343–363, abr. 2004. Disponível em: <[http://journals.ametsoc.org/doi/10.1175/1525-7541\(2004\)005%3C0343:AMRSMF%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1525-7541(2004)005%3C0343:AMRSMF%3E2.0.CO;2)>.

BASTIAANSEN, W. G. . SEBAL-based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. **Journal of Hydrology**, v. 229, n. 1–2, p. 87–100, mar. 1999. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022169499002024>>.

BASTIAANSEN, W. G. M.; PELGRUM, H.; WANG, J.; MA, Y.; MORENO, J. F.; ROERINK, G. J.; VAN DER WAL, T. A remote sensing surface energy balance algorithm for land (SEBAL), Part 1: Formulation. **Journal of Hydrology**, v. 212–213, n. JANUARY, p. 213–229, 1998. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0022169498002546>>.

BATTA, M. Machine Learning Algorithms - A Review. **International Journal of Science and Research (IJSR)**, v. 18, n. 8, p. 381–386, 2018.

BAUSCH, W. C. Remote sensing of crop coefficients for improving the irrigation scheduling of corn. **Agricultural Water Management**, v. 27, n. 1, p. 55–68, abr. 1995. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0378377495011253>>.

BAYRAM, S.; ÇITAKOĞLU, H. Modeling monthly reference evapotranspiration process in Turkey: application of machine learning methods. **Environmental Monitoring and Assessment**, v. 195, n. 1, p. 67, 3 jan. 2023. Disponível em: <<https://doi.org/10.1007/s10661-022-10662-z>>.

BELLIDO-JIMÉNEZ, J. A.; ESTÉVEZ, J.; GARCÍA-MARÍN, A. P. New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain. **Agricultural Water Management**, v. 245, n. September 2020, p. 106558, fev. 2021. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378377420321053>>.

BERNADO, S.; MANTOVANI, E. C. ; SILVA, D. D. da;; SOARES, A. A. **Manual de Irrigação**. Viçosa: Editora UFV, 2019. 545 p.

BERNI, J. A. J.; ZARCO-TEJADA, P. J.; SEPULCRE-CANTÓ, G.; FERERES, E.; VILLALOBOS, F. Mapping canopy conductance and CWSI in olive orchards using high resolution thermal remote sensing imagery. **Remote Sensing of Environment**, v. 113, n. 11, p. 2380–2388, 16 nov. 2009. Disponível em: <<http://dx.doi.org/10.1016/j.rse.2009.06.018>>.

BHOGAPURAPU, N.; DEY, S.; MANDAL, D.; BHATTACHARYA, A.; KARTHIKEYAN, L.; MCNAIRN, H.; RAO, Y. S. Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data. **Remote Sensing of Environment**, v. 271, n. January, p. 112900, mar. 2022. Disponível em: <<https://doi.org/10.1016/j.rse.2022.112900>>.

BROCK, T. D. Calculating solar radiation for ecological studies. **Ecological Modelling**, v. 14, n. 1–2, p. 1–19, nov. 1981. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0304380081900119>>.

BUNTING, D. P.; KURC, S. A.; GLENN, E. P.; NAGLER, P. L.; SCOTT, R. L. Insights for empirically modeling evapotranspiration influenced by riparian and upland vegetation in semiarid regions. **Journal of Arid Environments**, v. 111, p. 42–52, dez. 2014. Disponível em: <<http://dx.doi.org/10.1016/j.jaridenv.2014.06.007>>.

BUSHNELL, R. H. A solution for sunrise and sunset hour angles on a tilted surface without a singularity at zero azimuth. **Solar Energy**, v. 28, n. 4, p. 357, 1982. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0038092X82903115>>.

CARTER, C.; LIANG, S. Evaluation of ten machine learning methods for estimating terrestrial evapotranspiration from remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, v. 78, n. January, p. 86–92, jun. 2019. Disponível em: <<https://doi.org/10.1016/j.jag.2019.01.020>>.

CASTELLVÍ, F.; SNYDER, R. L. A comparison between latent heat fluxes over grass using a weighing lysimeter and surface renewal analysis. **Journal of Hydrology**, v. 381, n. 3–4, p. 213–220, fev. 2010. Disponível em: <<http://dx.doi.org/10.1016/j.jhydrol.2009.11.043>>.

CERVANTES, J.; GARCIA-LAMONT, F.; RODRÍGUEZ-MAZAHUA, L.; LOPEZ, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. **Neurocomputing**, v. 408, p. 189–215, set. 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925231220307153>>.

CHAI, R.; SUN, S.; CHEN, H.; ZHOU, S. Changes in reference evapotranspiration over China during 1960–2012: Attributions and relationships with atmospheric circulation. **Hydrological Processes**, v. 32, n. 19, p. 3032–3048, 2018.

CHANDEL, N. S.; RAJWADE, Y. A.; GOLHANI, K.; TIWARI, P. S.; DUBEY, K.; JAT, D. Canopy spectral reflectance for crop water stress assessment in wheat (<sc> *Triticum aestivum*, </sc> L.)*. **Irrigation and Drainage**, n. February, p. ird.2546, 8 nov. 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/ird.2546>>.

CHANDLER, W. S.; HOELL, J. M.; WESTBERG, D.; ZHANG, T.; STACKHOUSE JR, P. W. NASA Prediction of Worldwide Energy Resource High Resolution Meteorology Data For Sustainable Building Design. In: Proceedings of the Solar 2013 Conference (American Solar Energy Society), 2013, Baltimore, Maryland. [...]. Baltimore, Maryland: 2013.

CHEN, H.; WANG, P.; LI, J.; ZHANG, J.; ZHONG, L. Canopy Spectral Reflectance Feature and Leaf Water Potential of Sugarcane Inversion. **Physics Procedia**, v. 25, p. 595–600, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.phpro.2012.03.131>>.

CHEN, T.; GUESTRIN, C. XGBoost: A Scalable Tree Boosting System. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, v. 13-17- Augu, p. 785–794, 8 mar. 2016. Disponível em: <<http://arxiv.org/abs/1603.02754>>.

CHINTALA, S.; HARMYA, T. S.; KAMBHAMMETTU, B. V. N. P.; MOHARANA, S.; DUVVURI, S. Modelling high-resolution Evapotranspiration in fragmented croplands from the constellation of Sentinels. **Remote Sensing Applications: Society and Environment**, v. 26, n. January, p. 100704, abr. 2022. Disponível em: <<https://doi.org/10.1016/j.rsase.2022.100704>>.

CHRISTOPHER M. U. NEALE; WALTER C. BAUSCH; DALE F. HEERMANN. Development of Reflectance-Based Crop Coefficients for Corn. **Transactions of the ASAE**, v. 32, n. 6, p. 1891, 1990. Disponível em: <<http://elibrary.asabe.org/abstract.asp??JID=3&AID=31240&CID=t1989&v=32&i=6&T=1>>.

CONAB. **Boletim da safra de cana-de-açúcar**. Disponível em: <<https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>>.

COSTA, T. S.; FILGUEIRAS, R.; DOS SANTOS, R. A.; CUNHA, F. F. da. Actual evapotranspiration by machine learning and remote sensing without the thermal spectrum.

PLOS ONE, v. 18, n. 5, p. e0285535, 11 maio 2023. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0285535>>.

COUTINHO, L. M. **Biomass brasileiros**. [s.l: s.n.]121 p.

D'URSO, G.; MENENTI, M. <title>Mapping crop coefficients in irrigated areas from Landsat TM images</title>. (E. T. Engman, G. Guyot, C. M. Marino) In: Remote Sensing for Agriculture, Forestry, and Natural Resources, November 1995., 1995, [...]. 1995. v. 2585, p. 41–47.

DA SILVA, T. G. F.; DE MOURA, M. S. B.; ZOLNIER, S.; SOARES, J. M.; VIEIRA, V. J. d. S.; G. F. JÚNIOR, W. Water requirement and crop coefficient of irrigated sugarcane in a semi-arid region. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 16, n. 1, p. 64–71, 2012.

DEN BESTEN, N.; STEELE-DUNNE, S.; AOUIZERATS, B.; ZAJDBAND, A.; DE JEU, R.; VAN DER ZAAG, P. Observing Sucrose Accumulation With Sentinel-1 Backscatter. **Frontiers in Remote Sensing**, v. 2, n. December, p. 1–9, 20 dez. 2021. Disponível em: <<https://www.frontiersin.org/articles/10.3389/frsen.2021.778691/full>>.

DI NUNNO, F.; GRANATA, F. Future trends of reference evapotranspiration in Sicily based on CORDEX data and Machine Learning algorithms. **Agricultural Water Management**, v. 280, n. July 2022, p. 108232, abr. 2023. Disponível em: <<https://doi.org/10.1016/j.agwat.2023.108232>>.

DIAS, S. H. B.; FILGUEIRAS, R.; FERNANDES FILHO, E. I.; ARCANJO, G. S.; SILVA, G. H. da; MANTOVANI, E. C.; CUNHA, F. F. da. Reference evapotranspiration of Brazil modeled with machine learning techniques and remote sensing. **PLOS ONE**, v. 16, n. 2, p. e0245834, 9 fev. 2021. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0245834>>.

DINGRE, S. K.; GORANTIWAR, S. D. Determination of the water requirement and crop coefficient values of sugarcane by field water balance method in semiarid region. **Agricultural Water Management**, v. 232, n. February, p. 106042, abr. 2020. Disponível em: <<https://doi.org/10.1016/j.agwat.2020.106042>>.

DINGRE, S. K.; GORANTIWAR, S. D.; KADAM, S. A. Correlating the field water balance derived crop coefficient (Kc) and canopy reflectance-based NDVI for irrigated sugarcane. **Precision Agriculture**, v. 22, n. 4, p. 1134–1153, 20 ago. 2021. Disponível em: <<https://doi.org/10.1007/s11119-020-09774-8>>.

DONG, J.; ZHU, Y.; JIA, X.; SHAO, M.; HAN, X.; QIAO, J.; BAI, C.; TANG, X. Nation-scale reference evapotranspiration estimation by using deep learning and classical machine learning models in China. **Journal of Hydrology**, v. 604, n. November 2021, p. 127207, jan. 2022. Disponível em: <<https://doi.org/10.1016/j.jhydrol.2021.127207>>.

DOS SANTOS, E. P.; DA SILVA, D. D.; DO AMARAL, C. H. Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. **International Journal of Remote Sensing**, v. 42, n. 19, p. 7581–7609, 2 out. 2021. Disponível em: <<https://doi.org/10.1080/01431161.2021.1959955>>.

DOS SANTOS, E. P.; DA SILVA, D. D.; DO AMARAL, C. H.; FERNANDES-FILHO, E. I.; DIAS, R. L. S. A Machine Learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from Sentinel-1 and Landsat 8. **Computers and Electronics in Agriculture**, v. 194, n. January, p. 106753, mar. 2022a. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0168169922000709>>.

DOS SANTOS, R. A.; MANTOVANI, E. C.; FERNANDES-FILHO, E. I.;

FILGUEIRAS, R.; LOURENÇO, R. D. S.; BUFON, V. B.; NEALE, C. M. U. Modeling Actual Evapotranspiration with MSI-Sentinel Images and Machine Learning Algorithms. **Atmosphere**, v. 13, n. 9, p. 1518, 17 set. 2022b. Disponível em: <<https://www.mdpi.com/2073-4433/13/9/1518>>.

DUARTE, V. B. R.; VIOLA, M. R.; GIONGO, M.; ULIANA, E. M.; DE MELLO, C. R. Streamflow forecasting in Tocantins river basins using machine learning. **Water Supply**, v. 22, n. 7, p. 6230–6244, 1 jul. 2022. Disponível em: <<https://iwaponline.com/ws/article/22/7/6230/88092/Streamflow-forecasting-in-Tocantins-river-basins>>.

EKSTRØM, C. T. **MESS: Miscellaneous Esoteric Statistical Scripts**2020. Disponível em: <<https://cran.r-project.org/web/packages/MESS/MESS.pdf>>.

FENG, D.-C.; WANG, W.-J.; MANGALATHU, S.; TACIROGLU, E. Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. **Journal of Structural Engineering**, v. 147, n. 11, p. 1–13, 2021.

FENG, L.; LIU, Y.; ZHOU, Y.; YANG, S. A UAV-derived thermal infrared remote sensing three-temperature model and estimation of various vegetation evapotranspiration in urban micro-environments. **Urban Forestry & Urban Greening**, v. 69, n. January, p. 127495, mar. 2022. Disponível em: <<https://doi.org/10.1016/j.ufug.2022.127495>>.

FERNANDES FILHO, E. I. **labgeo: Collection of functions to fit models with emphasis in land use and soil mapping**2019. Disponível em: <<https://rdr.io/github/elpidiofilho/labgeo/>>.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017.

FILGUEIRAS, R.; MANTOVANI, E. C.; ALTHOFF, D.; FERNANDES FILHO, E. I.; CUNHA, F. F. da. Crop NDVI Monitoring Based on Sentinel 1. **Remote Sensing**, v. 11, n. 12, p. 1441, 18 jun. 2019a. Disponível em: <<https://www.mdpi.com/2072-4292/11/12/1441>>.

FILGUEIRAS, R.; MANTOVANI, E. C.; DIAS, S. H. B.; FERNANDES FILHO, E. I.; CUNHA, F. F. da. Optimizing the Monitoring of Natural Phenomena Through the Coupling of Orbital Multi-Sensors. **Geo UERJ**, n. 37, p. e37832, 2020.

FILGUEIRAS, R.; MANTOVANI, E. C.; DIAS, S. H. B.; FERNANDES FILHO, E. I.; CUNHA, F. F. da; NEALE, C. M. U. New approach to determining the surface temperature without thermal band of satellites. **European Journal of Agronomy**, v. 106, n. March 2019, p. 12–22, maio 2019b. Disponível em: <<https://doi.org/10.1016/j.eja.2019.03.001>>.

FRISON, P.-L.; FRUNEAU, B.; KMIHA, S.; SOUDANI, K.; DUFRÊNE, E.; TOAN, T. Le; KOLECK, T.; VILLARD, L.; MOUGIN, E.; RUDANT, J.-P. Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. **Remote Sensing**, v. 10, n. 12, p. 2049, 17 dez. 2018. Disponível em: <<http://www.mdpi.com/2072-4292/10/12/2049>>.

GAGNON, L.; JOUAN, A. Speckle filtering of SAR images: a comparative study between complex-wavelet-based and standard filters. (A. Aldroubi, A. F. Laine, M. A. Unser) In: *Wavelet Applications in Signal and Image Processing V*, October 1997., 1997, [...]. 1997. v. 3169, p. 80–91.

GOFFIN, B. D.; THAKUR, R.; DA CONCEIÇÃO CARLOS, S.; SRSIC, D.; WILLIAMS, C.; ROSS, K.; NEIRA-ROMÁN, F.; CORTÉS-MONROY, C. C.; LAKSHMI, V. Leveraging remotely-sensed vegetation indices to evaluate crop coefficients and actual

irrigation requirements in the water-stressed Maipo River Basin of Central Chile. **Sustainable Horizons**, v. 4, n. October, p. 100039, out. 2022. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2772737822000335>>.

GRANATA, F. Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. **Agricultural Water Management**, v. 217, n. March, p. 303–315, 2019. Disponível em: <<https://doi.org/10.1016/j.agwat.2019.03.015>>.

GREENWELL, B. **fastshap: Fast Approximate Shapley Values**2021. Disponível em: <<https://github.com/bgreenwell/fastshap>>.

HAJEB, M.; HAMZEH, S.; KAZEM ALAVIPANAH, S.; NEISSI, L.; VERRELST, J. Simultaneous retrieval of sugarcane variables from Sentinel-2 data using Bayesian regularized neural network. **International Journal of Applied Earth Observation and Geoinformation**, v. 116, n. January, p. 103168, fev. 2023. Disponível em: <<https://doi.org/10.1016/j.jag.2022.103168>>.

HAN, L.; YANG, G.; DAI, H.; XU, B.; YANG, H.; FENG, H.; LI, Z.; YANG, X. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. **Plant Methods**, v. 15, n. 1, p. 10, 4 dez. 2019. Disponível em: <<https://plantmethods.biomedcentral.com/articles/10.1186/s13007-019-0394-z>>.

HARGREAVES, G. H.; SAMANI, Z. A. Reference crop evapotranspiration from temperature. **Applied engineering in agriculture**, v. 1, n. 2, p. 96–99, 1985.

HIJMANS, R. J.; VAN ETTEN, J.; CHENG, J.; MATTIUZZI, M.; SUMNER, M.; GREENBERG, J. A.; LAMIGUEIRO, O. P.; BEVAN, A.; RACINE, E. B.; SHORTRIDGE, A.; OTHERS. **Raster: Geographic Data Analysis and Modeling**2020. Disponível em: <<https://cran.r-project.org/web/packages/raster/raster.pdf>>.

HIRD, J.; DELANCEY, E.; MCDERMID, G.; KARIYEVA, J. Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping. **Remote Sensing**, v. 9, n. 12, p. 1315, 14 dez. 2017. Disponível em: <<http://www.mdpi.com/2072-4292/9/12/1315>>.

HUANG, J.; PANG, Y.; LIU, Y.; YAN, H. Posterior Regularized Bayesian Neural Network incorporating soft and hard knowledge constraints. **Knowledge-Based Systems**, v. 259, p. 110043, jan. 2023. Disponível em: <<https://doi.org/10.1016/j.knosys.2022.110043>>.

ILYAS, M.; NISAR, M.; KHAN, N.; HAZRAT, A.; KHAN, A. H.; HAYAT, K.; FAHAD, S.; KHAN, A.; ULLAH, A. Drought Tolerance Strategies in Plants: A Mechanistic Approach. **Journal of Plant Growth Regulation**, v. 40, n. 3, p. 926–944, 22 jun. 2021. Disponível em: <<https://doi.org/10.1007/s00344-020-10174-5>>.

IMPOLLONIA, G.; CROCI, M.; MARTANI, E.; FERRARINI, A.; KAM, J.; TRINDADE, L. M.; CLIFTON-BROWN, J.; AMADUCCI, S. Moisture content estimation and senescence phenotyping of novel Miscanthus hybrids combining UAV-based remote sensing and machine learning. **GCB Bioenergy**, v. 14, n. 6, p. 639–656, 12 jun. 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/gcbb.12930>>.

IPEA. **Comércio exterior do agronegócio: janeiro de 2023**. Disponível em: <https://www.ipea.gov.br/cartadeconjuntura/wp-content/uploads/2023/02/230210_cc_58_nota_12_comercio_exterior_agro.pdf>.

JAAFAR, H.; MOURAD, R.; SCHULL, M. A global 30-m ET model (HSEB) using harmonized Landsat and Sentinel-2, MODIS and VIIRS: Comparison to ECOSTRESS ET and LST. **Remote Sensing of Environment**, v. 274, n. March 2021, p. 112995, jun. 2022.

Disponível em: <<https://doi.org/10.1016/j.rse.2022.112995>>.

JENSEN, J. R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. [s.l.] Parêntese, 2009. 598 p.

JOHANSEN, K.; ZILIANI, M. G.; HOUBORG, R.; FRANZ, T. E.; MCCABE, M. F. CubeSat constellations provide enhanced crop phenology and digital agricultural insights using daily leaf area index retrievals. **Scientific Reports**, v. 12, n. 1, p. 5244, 28 mar. 2022. Disponível em: <<https://doi.org/10.1038/s41598-022-09376-6>>.

KADKHODAZADEH, M.; VALIKHAN ANARAKI, M.; MORSHED-BOZORGDEL, A.; FARZIN, S. A New Methodology for Reference Evapotranspiration Prediction and Uncertainty Analysis under Climate Change Conditions Based on Machine Learning, Multi Criteria Decision Making and Monte Carlo Methods. **Sustainability**, v. 14, n. 5, p. 2601, 23 fev. 2022. Disponível em: <<https://www.mdpi.com/2071-1050/14/5/2601>>.

KAMBLE, B.; KILIC, A.; HUBBARD, K. Estimating crop coefficients using remote sensing-based vegetation index. **Remote Sensing**, v. 5, n. 4, p. 1588–1602, 2013.

KAPLAN, G.; FINE, L.; LUKYANOV, V.; MALACHY, N.; TANNY, J.; ROZENSTEIN, O. Using Sentinel-1 and Sentinel-2 imagery for estimating cotton crop coefficient, height, and Leaf Area Index. **Agricultural Water Management**, v. 276, n. July 2022, p. 108056, fev. 2023. Disponível em: <<https://doi.org/10.1016/j.agwat.2022.108056>>.

KAPLAN, G.; FINE, L.; LUKYANOV, V.; MANIVASAGAM, V. S.; TANNY, J.; ROZENSTEIN, O. Normalizing the Local Incidence Angle in Sentinel-1 Imagery to Improve Leaf Area Index, Vegetation Height, and Crop Coefficient Estimations. **Land**, v. 10, n. 7, p. 680, 28 jun. 2021. Disponível em: <<https://www.mdpi.com/2073-445X/10/7/680>>.

KE, Y.; IM, J.; PARK, S.; GONG, H. Spatiotemporal downscaling approaches for monitoring 8-day 30 m actual evapotranspiration. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 126, p. 79–93, abr. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.isprsjprs.2017.02.006>>.

KHABBAZAN, S.; VERMUNT, P.; STEELE-DUNNE, S.; RATERING ARNTZ, L.; MARINETTI, C.; VAN DER VALK, D.; IANNINI, L.; MOLIJN, R.; WESTERDIJK, K.; VAN DER SANDE, C. Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. **Remote Sensing**, v. 11, n. 16, p. 1887, 13 ago. 2019. Disponível em: <<https://www.mdpi.com/2072-4292/11/16/1887>>.

KIMBALL, B. A.; THORP, K. R.; BOOTE, K. J.; STOCKLE, C.; SUYKER, A. E.; EVETT, S. R.; BRAUER, D. K.; COYLE, G. G.; COPELAND, K. S.; MAREK, G. W.; COLAIZZI, P. D.; ACUTIS, M.; ALIMAGHAM, S.; ARCHONTOULIS, S.; BABACAR, F.; BARCZA, Z.; BASSO, B.; BERTUZZI, P.; CONSTANTIN, J.; DE ANTONI MIGLIORATI, M.; DUMONT, B.; DURAND, J.; FODOR, N.; GAISER, T.; GAROFALO, P.; GAYLER, S.; GIGLIO, L.; GRANT, R.; GUAN, K.; HOOGENBOOM, G.; JIANG, Q.; KIM, S.; KISEKKA, I.; LIZASO, J.; MASIA, S.; MENG, H.; MEREU, V.; MUKHTAR, A.; PEREGO, A.; PENG, B.; PRIESACK, E.; QI, Z.; SHELIA, V.; SNYDER, R.; SOLTANI, A.; SPANO, D.; SRIVASTAVA, A.; THOMSON, A.; TIMLIN, D.; TRABUCCO, A.; WEBBER, H.; WEBER, T.; WILLAUME, M.; WILLIAMS, K.; VAN DER LAAN, M.; VENTRELLA, D.; VISWANATHAN, M.; XU, X.; ZHOU, W. Simulation of evapotranspiration and yield of maize: An Inter-comparison among 41 maize models. **Agricultural and Forest Meteorology**, v. 333, n. March, p. 109396, abr. 2023. Disponível em: <<https://doi.org/10.1016/j.agrformet.2019.02.037>>.

KOCIAN, A.; CARMASSI, G.; CELA, F.; CHESSA, S.; MILAZZO, P.; INCROCCI,

L. IoT based dynamic Bayesian prediction of crop evapotranspiration in soilless cultivations. **Computers and Electronics in Agriculture**, v. 205, n. January, p. 107608, fev. 2023. Disponível em: <<https://doi.org/10.1016/j.compag.2022.107608>>.

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; TEAM, R. C. **Classification and Regression Training**. 2020. Disponível em: <<https://cran.r-project.org/package=caret>>.

LOOS, C.; GAYLER, S.; PRIESACK, E. Assessment of water balance simulations for large-scale weighing lysimeters. **Journal of Hydrology**, v. 335, n. 3–4, p. 259–270, mar. 2007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022169406006251>>.

LUNDBERG, S. M.; LEE, S. I. A unified approach to interpreting model predictions. **Advances in Neural Information Processing Systems**, v. 2017-Decem, n. Section 2, p. 4766–4775, 2017.

MAKWANA, J. J.; TIWARI, M. K.; DEORA, B. S. Development and comparison of artificial intelligence models for estimating daily reference evapotranspiration from limited input variables. **Smart Agricultural Technology**, v. 3, n. September 2022, p. 100115, fev. 2023. Disponível em: <<https://doi.org/10.1016/j.atech.2022.100115>>.

MAYER, M. **shapviz: SHAP Visualizations** 2023. Disponível em: <<https://github.com/mayer79/shapviz>>.

MECIKALSKI, J. R.; DIAK, G. R.; ANDERSON, M. C.; NORMAN, J. M. Estimating Fluxes on Continental Scales Using Remotely Sensed Data in an Atmospheric–Land Exchange Model. **Journal of Applied Meteorology**, v. 38, n. 9, p. 1352–1369, set. 1999. Disponível em: <[http://journals.ametsoc.org/doi/10.1175/1520-0450\(1999\)038%3C1352:EFOCSU%3E2.0.CO;2](http://journals.ametsoc.org/doi/10.1175/1520-0450(1999)038%3C1352:EFOCSU%3E2.0.CO;2)>.

MUSYIMI, P. K.; SAHBENI, G.; TIMÁR, G.; WEIDINGER, T.; SZÉKELY, B. Actual Evapotranspiration Estimation Using Sentinel-1 SAR and Sentinel-3 SLSTR Data Combined with a Gradient Boosting Machine Model in Busia County, Western Kenya. **Atmosphere**, v. 13, n. 11, p. 1927, 18 nov. 2022. Disponível em: <<https://www.mdpi.com/2073-4433/13/11/1927>>.

NAGLER, P.; GLENN, E.; NGUYEN, U.; SCOTT, R.; DOODY, T. Estimating Riparian and Agricultural Actual Evapotranspiration by Reference Evapotranspiration and MODIS Enhanced Vegetation Index. **Remote Sensing**, v. 5, n. 8, p. 3849–3871, 5 ago. 2013. Disponível em: <<http://www.mdpi.com/2072-4292/5/8/3849>>.

NASIRZADEHDIZAJI, R.; BALIK SANLI, F.; ABDIKAN, S.; CAKIR, Z.; SEKERTEKIN, A.; USTUNER, M. Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. **Applied Sciences**, v. 9, n. 4, p. 655, 15 fev. 2019. Disponível em: <<https://www.mdpi.com/2076-3417/9/4/655>>.

NGUYEN, L. H.; ROBINSON, S.; GALPERN, P. Medium-resolution multispectral satellite imagery in precision agriculture: mapping precision canola (*Brassica napus* L.) yield using Sentinel-2 time series. **Precision Agriculture**, v. 23, n. 3, p. 1051–1071, 29 jun. 2022. Disponível em: <<https://doi.org/10.1007/s11119-022-09874-7>>.

NORMA, J. M.; KUSTAS, W. P.; HUMES, K. S. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. **Agricultural and Forest Meteorology**, v. 77, n. 3–4, p. 263–293, dez. 1995. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/016819239502265Y>>.

NORMAN, J. M.; KUSTAS, W. P.; PRUEGER, J. H.; DIAK, G. R. Surface flux

estimation using radiometric temperature: A dual-temperature-difference method to minimize measurement errors. **Water Resources Research**, v. 36, n. 8, p. 2263–2274, ago. 2000. Disponível em: <<http://doi.wiley.com/10.1029/2000WR900033>>.

OCHEGE, F. U.; LUO, G.; OBETA, M. C.; OWUSU, G.; DUULATOV, E.; CAO, L.; NSENGIYUMVA, J. B. Mapping evapotranspiration variability over a complex oasis-desert ecosystem based on automated calibration of Landsat 7 ETM+ data in SEBAL. **GIScience & Remote Sensing**, v. 56, n. 8, p. 1305–1332, 17 nov. 2019. Disponível em: <<https://doi.org/10.1080/15481603.2019.1643531>>.

OLMEDO, G. F.; ORTEGA-FARIAS, S.; FONSECA-LUENGO, D.; LA FUENTE-SÁIZ, D. de; FUENTES-PENAILILLO, F.; MUNAFÓ, M. V. **water: Actual Evapotranspiration with Energy Balance Models**. Disponível em: <<https://cran.r-project.org/web/packages/water/water.pdf>>. Acesso em: 3 jan. 2021.

OUAADI, N.; JARLAN, L.; KHABBA, S.; LE PAGE, M.; CHAKIR, A.; ER-RAKI, S.; FRISON, P.-L. Are the C-band backscattering coefficient and interferometric coherence suitable substitutes of NDVI for the monitoring of the FAO-56 crop coefficient? **Agricultural Water Management**, v. 282, n. October 2022, p. 108276, maio 2023. Disponível em: <<https://doi.org/10.1016/j.agwat.2023.108276>>.

PELTA, R.; BEERI, O.; TARSHISH, R.; SHILO, T. Forecasting seasonal plot-specific crop coefficient (Kc) protocol for processing tomato using remote sensing, meteorology, and artificial intelligence. **Precision Agriculture**, v. 23, n. 6, p. 1983–2000, 23 dez. 2022. Disponível em: <<https://doi.org/10.1007/s11119-022-09910-6>>.

PENMAN, H. L. Natural evaporation from open water, bare soil and grass. **Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences**, v. 193, n. 1032, p. 120–145, 22 abr. 1948. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rspa.1948.0037>>.

PERIASAMY, S. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. **Remote Sensing of Environment**, v. 217, n. April, p. 537–549, nov. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0034425718304140>>.

QIN, L.; YAN, C.; YU, L.; CHAI, M.; WANG, B.; HAYAT, M.; SHI, Z.; GAO, H.; JIANG, X.; XIONG, B.; MAO, P.; QIU, G. Y. High-resolution spatio-temporal characteristics of urban evapotranspiration measured by unmanned aerial vehicle and infrared remote sensing. **Building and Environment**, v. 222, n. April, p. 109389, ago. 2022. Disponível em: <<https://doi.org/10.1016/j.buildenv.2022.109389>>.

R CORE TEAM. **R: A language and environment for statistical computing** Vienna, Austria. R Foundation for Statistical Computing, 2020.

RAZA, A.; SHOAI, M.; FAIZ, M. A.; BAIG, F.; KHAN, M. M.; ULLAH, M. K.; ZUBAIR, M. Comparative Assessment of Reference Evapotranspiration Estimation Using Conventional Method and Machine Learning Algorithms in Four Climatic Regions. **Pure and Applied Geophysics**, v. 177, n. 9, p. 4479–4508, 2020.

REICHSTEIN, M.; CAMPS-VALLS, G.; STEVENS, B.; JUNG, M.; DENZLER, J.; CARVALHAIS, N.; PRABHAT. Deep learning and process understanding for data-driven Earth system science. **Nature**, v. 566, n. 7743, p. 195–204, 13 fev. 2019. Disponível em: <<http://dx.doi.org/10.1038/s41586-019-0912-1>>.

REN, S.; GUO, B.; WANG, Z.; WANG, J.; FANG, Q.; WANG, J. Optimized spectral

index models for accurately retrieving soil moisture (SM) of winter wheat under water stress. **Agricultural Water Management**, v. 261, n. October 2021, p. 107333, mar. 2022. Disponível em: <<https://doi.org/10.1016/j.agwat.2021.107333>>.

ROERINK, G. ; SU, Z.; MENENTI, M. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. **Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere**, v. 25, n. 2, p. 147–157, jan. 2000. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1464190999001288>>.

SAGGI, M. K.; JAIN, S. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. **Computers and Electronics in Agriculture**, v. 156, n. November 2018, p. 387–398, jan. 2019. Disponível em: <<https://doi.org/10.1016/j.compag.2018.11.031>>.

SALEEM, A.; AWANGE, J. L. Coastline shift analysis in data deficient regions: Exploiting the high spatio-temporal resolution Sentinel-2 products. **CATENA**, v. 179, n. September 2018, p. 6–19, ago. 2019. Disponível em: <<https://doi.org/10.1016/j.catena.2019.03.023>>.

SANTOS, E. P. dos; BHOGAPURAPU, N. R. eupassarinho/sentinel-1-SAR-vegetation-indices: v0.1. 20 nov. 2022. Disponível em: <<https://doi.org/10.5281/zenodo.7339421#ZFVgu21Qwp4.mendeley>>. Acesso em: 5 maio. 2023.

SANTOS, N. V. dos; DEMATTÊ, J. A. M.; SILVERO, N. E. Q. Improving the monitoring of sugarcane residues in a tropical environment based on laboratory and Sentinel-2 data. **International Journal of Remote Sensing**, v. 42, n. 5, p. 1768–1784, 4 mar. 2020. Disponível em: <<https://doi.org/10.1080/01431161.2020.1842542>>.

SANTOS, R. A. dos; CHARTUNI MANTOVANI, E.; FILGUEIRAS, R.; INÁCIO FERNANDES-FILHO, E.; CRISTIELLE BARBOSA DA SILVA, A.; PERONI VENANCIO, L. Actual Evapotranspiration and Biomass of Maize from a Red–Green–Near-Infrared (RGNIR) Sensor on Board an Unmanned Aerial Vehicle (UAV). **Water**, v. 12, n. 9, p. 2359, 22 ago. 2020. Disponível em: <<https://www.mdpi.com/805378>>. Acesso em: 22 ago. 2020.

SATTARI, M. T.; APAYDIN, H.; BAND, S. S.; MOSAVI, A.; PRASAD, R. Comparative analysis of kernel-based versus ANN and deep learning methods in monthly reference evapotranspiration estimation. **Hydrology and Earth System Sciences**, v. 25, n. 2, p. 603–618, 10 fev. 2021. Disponível em: <<https://hess.copernicus.org/articles/25/603/2021/>>.

SENAY, G. B.; BUDDE, M. E.; VERDIN, J. P. Enhancing the Simplified Surface Energy Balance (SSEB) approach for estimating landscape ET: Validation with the METRIC model. **Agricultural Water Management**, v. 98, n. 4, p. 606–618, fev. 2011. Disponível em: <<http://dx.doi.org/10.1016/j.agwat.2010.10.014>>.

SHANG, K.; YAO, Y.; DI, Z.; JIA, K.; ZHANG, X.; FISHER, J. B.; CHEN, J.; GUO, X.; YANG, J.; YU, R.; XIE, Z.; LIU, L.; NING, J.; ZHANG, L. Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau. **Remote Sensing of Environment**, v. 289, n. December 2022, p. 113519, maio 2023. Disponível em: <<https://doi.org/10.1016/j.rse.2023.113519>>.

SHORACHI, M.; KUMAR, V.; STEELE-DUNNE, S. C. Sentinel-1 SAR Backscatter Response to Agricultural Drought in The Netherlands. **Remote Sensing**, v. 14, n. 10, p. 2435, 19 maio 2022. Disponível em: <<https://www.mdpi.com/2072-4292/14/10/2435>>.

SILVA, V. de P. R. da; GARCÊZ, S. L. A.; SILVA, B. B. da; ALBUQUERQUE, M. F.

De; ALMEIDA, R. S. R. Métodos de estimativa da evapotranspiração da cultura da cana-de-açúcar em condições de sequeiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 5, p. 411–417, maio 2015. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-43662015000500411&lng=pt&tlng=pt>.

SPARKS, A. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. **Journal of Open Source Software**, v. 3, n. 30, p. 1035, 19 out. 2018. Disponível em: <<http://joss.theoj.org/papers/10.21105/joss.01035>>.

SRIVASTAVA, A.; SAHOO, B.; RAGHUWANSHI, N. S.; SINGH, R. Evaluation of Variable-Infiltration Capacity Model and MODIS-Terra Satellite-Derived Grid-Scale Evapotranspiration Estimates in a River Basin with Tropical Monsoon-Type Climatology. **Journal of Irrigation and Drainage Engineering**, v. 143, n. 8, 2017.

STAPLETON, A.; EICHELMANN, E.; ROANTREE, M. A framework for constructing machine learning models with feature set optimisation for evapotranspiration partitioning. **Applied Computing and Geosciences**, v. 16, n. September, 2022.

STUMPF, A.; MICHÉA, D.; MALET, J.-P. Improved Co-Registration of Sentinel-2 and Landsat-8 Imagery for Earth Surface Motion Measurements. **Remote Sensing**, v. 10, n. 2, p. 160, 23 jan. 2018. Disponível em: <<http://www.mdpi.com/2072-4292/10/2/160>>.

SUN, Q.; GU, X.; SUN, L.; YANG, G.; ZHOU, L.; GUO, W. Dynamic change in rice leaf area index and spectral response under flooding stress. **Paddy and Water Environment**, v. 18, n. 1, p. 223–233, 10 jan. 2020. Disponível em: <<https://doi.org/10.1007/s10333-019-00776-5>>.

TAIZ, L.; ZEIGER, E.; MOLLER, I. max; MURPHY, A. **Plant Physiology & Development**. Sunderland, MA, USA: Sinauer Associates Incorporated, 2015. 858 p.

TANG, D.; FENG, Y.; GONG, D.; HAO, W.; CUI, N. Evaluation of artificial intelligence models for actual crop evapotranspiration modeling in mulched and non-mulched maize croplands. **Computers and Electronics in Agriculture**, v. 152, n. July, p. 375–384, 2018. Disponível em: <<https://doi.org/10.1016/j.compag.2018.07.029>>.

TASUMI, M. **Progress in Operational Estimation of Regional Evapotranspiration using Satellite Imagery**. 2003. University of Idaho, 2003.

TEIXEIRA, A. H. de C. Determining Regional Actual Evapotranspiration of Irrigated Crops and Natural Vegetation in the São Francisco River Basin (Brazil) Using Remote Sensing and Penman-Monteith Equation. **Remote Sensing**, v. 2, n. 5, p. 1287–1319, 6 maio 2010. Disponível em: <<http://www.mdpi.com/2072-4292/2/5/1287>>.

TEIXEIRA; BASTIAANSEN, W. G. M.; AHMAD, M. D.; BOS, M. G. Reviewing SEBAL input parameters for assessing evapotranspiration and water productivity for the Low-Middle São Francisco River basin, Brazil. Part B: Application to the regional scale. **Agricultural and Forest Meteorology**, v. 149, n. 3–4, p. 477–490, 2009.

THORNTHWAITE, C. W. An Approach toward a Rational Classification of Climate. **Geographical Review**, v. 38, n. 1, p. 55, jan. 1948. Disponível em: <<https://www.jstor.org/stable/210739?origin=crossref>>.

TIKHAMARINE, Y.; MALIK, A.; KUMAR, A.; SOUAG-GAMANE, D.; KISI, O. Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches. **Hydrological Sciences Journal**, v. 64, n. 15, p. 1824–1842, 2019. Disponível em: <<https://doi.org/10.1080/02626667.2019.1678750>>.

VIRNODKAR, S. S.; PACHGHARE, V. K.; PATIL, V. C.; JHA, S. K. Remote sensing and machine learning for crop water stress determination in various crops: a critical review. **Precision Agriculture**, v. 21, n. 5, p. 1121–1155, 17 out. 2020. Disponível em: <<https://doi.org/10.1007/s11119-020-09711-9>>.

VREUGDENHIL, M.; NAVACCHI, C.; BAUER-MARSCHALLINGER, B.; HAHN, S.; STEELE-DUNNE, S.; PFEIL, I.; DORIGO, W.; WAGNER, W. Sentinel-1 Cross Ratio and Vegetation Optical Depth: A Comparison over Europe. **Remote Sensing**, v. 12, n. 20, p. 3404, 16 out. 2020. Disponível em: <<https://www.mdpi.com/2072-4292/12/20/3404>>.

VREUGDENHIL, M.; WAGNER, W.; BAUER-MARSCHALLINGER, B.; PFEIL, I.; TEUBNER, I.; RÜDIGER, C.; STRAUSS, P. Sensitivity of Sentinel-1 backscatter to vegetation dynamics: An Austrian case study. **Remote Sensing**, v. 10, n. 9, p. 1–19, 2018.

XIANG, K.; LI, Y.; HORTON, R.; FENG, H. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. **Agricultural Water Management**, v. 232, n. August 2019, p. 106043, abr. 2020. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0378377419315616>>.

XU, T.; GUO, Z.; XIA, Y.; FERREIRA, V. G.; LIU, S.; WANG, K.; YAO, Y.; ZHANG, X.; ZHAO, C. Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. **Journal of Hydrology**, v. 578, n. July, p. 124105, nov. 2019. Disponível em: <<https://doi.org/10.1016/j.jhydrol.2019.124105>>.

YANG, M.; HUANG, C.; KANG, X.; QIN, S.; MA, L.; WANG, J.; ZHOU, X.; LV, X.; ZHANG, Z. Early Monitoring of Cotton Verticillium Wilt by Leaf Multiple “Symptom” Characteristics. **Remote Sensing**, v. 14, n. 20, p. 5241, 20 out. 2022. Disponível em: <<https://www.mdpi.com/2072-4292/14/20/5241>>.

YANG, Y.; SU, H.; ZHANG, R.; TIAN, J.; LI, L. An enhanced two-source evapotranspiration model for land (E-TEML): Algorithm and evaluation. **Remote Sensing of Environment**, v. 168, p. 54–65, out. 2015. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S003442571530050X>>.

ZHANG, F.; ZHOU, G. Estimation of vegetation water content using hyperspectral vegetation indices: A comparison of crop water indicators in response to water stress treatments for summer maize. **BMC Ecology**, v. 19, n. 1, p. 1–12, 2019. Disponível em: <<https://doi.org/10.1186/s12898-019-0233-0>>.

ZHOU, Z.; MAJEED, Y.; DIVERRES NARANJO, G.; GAMBACORTA, E. M. T. Assessment for crop water stress with infrared thermal imagery in precision agriculture: A review and future prospects for deep learning applications. **Computers and Electronics in Agriculture**, v. 182, n. November 2020, p. 106019, mar. 2021. Disponível em: <<https://doi.org/10.1016/j.compag.2021.106019>>.