

UNIVERSIDADE FEDERAL DE VIÇOSA

**Monitoramento da qualidade da água utilizando sensoriamento remoto e
aprendizado de máquina**

Rafael Luis Silva Dias
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2025**

RAFAEL LUIS SILVA DIAS

**Monitoramento da qualidade da água utilizando sensoriamento remoto e
aprendizado de máquina**

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: Ricardo S. Silva Amorim

Coorientadores: Demetrius D. da Silva
Elpidio I. F. Filho

**VIÇOSA - MINAS GERAIS
2025**

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

T

D541m
2025

Dias, Rafael Luís Silva, 1993-
Monitoramento da qualidade da água utilizando
sensoriamento remoto e aprendizado de máquina / Rafael Luís
Silva Dias. – Viçosa, MG, 2025.
1 tese eletrônica (103 f.): il. (algumas color.).

Inclui apêndice.

Orientador: Ricardo Santos Silva Amorim.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Engenharia Agrícola, 2025.

Referências bibliográficas: f.62-70.

DOI: <https://doi.org/10.47328/ufvbbt.2025.672>

Modo de acesso: World Wide Web.

1. Desenvolvimento de recursos hídricos. 2. Água -
Qualidade - Paraopeba, Rio (MG). 3. Desastres ambientais.
4. Sensoriamento remoto. 5. Aprendizado do computador.
I. Amorim, Ricardo Santos Silva, 1971-. II. Universidade
Federal de Viçosa. Departamento de Engenharia Agrícola.
Programa de Pós-Graduação em Engenharia Agrícola. III. Título.

CDD 22. ed. 333.910098151

RAFAEL LUIS SILVA DIAS

**Monitoramento da qualidade da água utilizando sensoriamento remoto e
aprendizado de máquina**

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: 21 de julho de 2025.

Assentimento:

Rafael Luis Silva Dias
Autor

Ricardo Santos Silva Amorim
Orientador

Essa tese foi assinada digitalmente pelo autor em 20/10/2025 às 14:44:39 e pelo orientador em 20/10/2025 às 23:24:09. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **X2KU.URIM.V7PC** e clique no botão 'Validar documento'.

A Deus e à minha família, **dedico**.

AGRADECIMENTOS

Primeiramente agradeço a Deus por sempre estar ao meu lado durante a minha caminhada, protegendo e dando alento nos momentos mais críticos. E pela intercessão de Nossa Senhora em minha vida.

Aos meus pais Walmir e Maria, minha fonte de amor, carinho e que nunca pouparam esforços e sempre abriram mão de muitos dos seus prazeres para garantir, a mim e ao meu irmão, uma boa formação. A meu irmão Bruno, por ser meu parceiro número um, sempre me apoiando incondicionalmente.

À minha noiva Fernanda, pelo amor, companheirismo e amizade, que desde o início não poupou esforços para me incentivar a encarar este desafio e, durante todo esse tempo.

Ao meu orientador Ricardo, pelos conhecimentos e exemplo de profissionalismo a mim repassados, pela confiança e apoio durante a orientação. Aos meus coorientadores Demetrius e Elpídio por me auxiliarem nos momentos de dificuldade, sem vocês não teria conseguido.

Ao meu estagiário Ronam que foi peça chave para a execução deste trabalho e a toda equipe multidisciplinar que auxiliou em cada etapa, em especial ao Gustavo (Ganso) do Departamento de Solos e a equipe do IGAM.

Aos meus amigos e colegas do Centro Referência em Recursos Hídricos (CRRH) pela amizade, companheirismo e por compartilharem bons momentos e as trocas de conhecimentos em nosso dia a dia.

Enfim, o meu muito obrigado a todos que me ajudaram e apoiaram de alguma forma durante o doutorado.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Seja quem você for,
seja qual for a posição social que você tenha na vida,
a mais alta ou a mais baixa,
tenha sempre como meta muita força,
muita determinação e sempre faça tudo
com muito amor e com muita fé em Deus,
que um dia você chega lá.
De alguma forma você chega lá.”
(Ayrton Senna)

“Por mais que as coisas avancem,
o básico, sempre será o fundamental. ”

RESUMO

DIAS, Rafael Luis Silva, D.Sc., Universidade Federal de Viçosa, julho de 2025. **Monitoramento da qualidade da água utilizando sensoriamento remoto e aprendizado de máquina.** Orientador: Ricardo Santos Silva Amorim. Coorientadores: Demetrius David da Silva e Elpidio Inacio Fernandes Filho.

O método convencional de monitoramento da qualidade das águas (QA) continentais é realizado por meio de amostragens pontuais. Embora confiável, esse método apresenta limitações expressivas na representação da variabilidade espacial e temporal dos parâmetros de QA, além de ser oneroso e demandar grande esforço operacional, principalmente num país de dimensões continentais como o Brasil. Nesse contexto, torna-se essencial a busca de alternativas complementares que ampliem a eficiência e a abrangência do monitoramento qualitativo das águas superficiais. A utilização de imagens orbitais surge, portanto, como uma alternativa promissora aprimorar a avaliação contínua da QA. No entanto, o uso de imagens orbitais ainda enfrenta limitações relacionadas às resoluções espectral, espacial, temporal e radiométrica, o que pode comprometer a precisão das avaliações. Nesse cenário, a constelação de nanossatélites Planet surge como uma alternativa viável, oferecendo alta frequência de revisita e elevada resolução espacial. Diante desse contexto, o objetivo deste trabalho foi desenvolver uma metodologia que permita o monitoramento, com elevada frequência, dos parâmetros de qualidade da água (T, SST, Chla, P, N, DQO, OD e Fe) em ambientes lênticos e lóticos, baseada em produtos de sensoriamento remoto e técnicas de aprendizado de máquina. Contudo, as imagens captadas pelos sensores PlanetScope (PS) frequentemente apresentam inconsistências radiométricas devido à variabilidade entre os diferentes satélites da constelação, comprometendo a homogeneidade da série temporal. Desta forma, para alcançar o referido objetivo, a pesquisa foi estruturada em dois capítulos: (i) Normalização radiométrica relativa para a constelação de nanossatélites PS com base em imagens Sentinel-2 (S2); e (ii) Proposição de modelos de aprendizado de máquina aplicados à predição de parâmetros da qualidade da água na bacia do rio Paraopeba. No primeiro capítulo foi desenvolvido um modelo de normalização radiométrica das imagens PS com base em imagens S2, aplicando cinco algoritmos de aprendizado de máquina: Multiple Linear Regression (MLR), Model Averaged Neural Network (avNNet), Random Forest (RF), k-Nearest Neighbors (KNN) e Support Vector Machines com kernel de função de base radial (SVM-RBF), em um conjunto de dados que abrange o período de 2017 a 2022. No segundo capítulo foi

proposta uma metodologia para predição da QA em ambientes lênticos e lóticos, utilizando dados de sensoriamento remoto e técnicas de aprendizado de máquina. Para isso, foram utilizados os dados (2016 a 2023) de 24 estações de monitoramento da QA localizadas ao longo do rio Paraopeba e no reservatório de Três Marias. Os dados compilados foram organizados em três subconjuntos de dados: (i) dados derivados de imagens S2; (ii) dados provenientes de imagens brutas PS; e (iii) dados obtidos a partir das imagens PS normalizadas com a metodologia do capítulo 1. Quatro algoritmos foram testados: avNNet, RF, KNN e SVM-RBF. A avaliação do desempenho dos modelos foi realizada utilizando-se as seguintes métricas estatísticas: raiz do erro quadrático médio – RMSE, erro médio absoluto – MAE, coeficiente de correlação de concordância de Lin – CCC e o coeficiente de determinação - R^2 . A normalização radiométrica desenvolvida para o sensor PS, com base nos dados S2, mostrou-se necessária e adequada para atenuar as inconsistências da constelação de nanossatélites da Planet. O modelo SVM-RBF apresentou o melhor desempenho para a normalização radiométrica de todas as bandas do sensor PS, especialmente para as de maior comprimento de onda (bandas 3 e 4), alcançando baixos valores de RMSE (0,032 a 0,034) e elevados valores de R^2 e CCC, os quais variaram de 0,77 a 0,90 e 0,87 a 0,95, respectivamente. Os dados provenientes dos sensores PS normalizados apresentaram melhor desempenho na predição dos parâmetros de QA em comparação aos dados PS brutos e do S2. O modelo desenvolvido com base no algoritmo RF demonstrou maior robustez em comparação aos modelos baseados nos algoritmos Cubist, KNN e SVM-RBF. Dessa forma, a metodologia proposta se mostrou robusta e capaz de complementar o monitoramento tradicional da qualidade da água, especialmente para os parâmetros T, SST, Fe, P e OD, oferecendo alta resolução temporal e espacial. Essa abordagem possibilita um avanço no monitoramento da qualidade das águas superficiais, bem como na avaliação dos impactos de eventos atípicos (extremos) sobre a qualidade dos recursos hídricos.

Palavras-chave: águas continentais; desastres ambientais; eventos extremos; monitoramento ambiental; gestão dos recursos hídricos

ABSTRACT

DIAS, Rafael Luis Silva, D.Sc., Universidade Federal de Viçosa, July, 2025. **Water quality monitoring using remote sensing and machine learning**. Adviser: Ricardo Santos Silva Amorim. Co-advisers: Demetrius David da Silva and Elpidio Inacio Fernandes Filho.

The conventional approach for monitoring continental water quality (WQ) relies on point sampling. While reliable, this method presents significant limitations in capturing the spatial and temporal variability of WQ parameters and is both costly and operationally demanding, particularly in a country of continental dimensions such as Brazil. Consequently, the development of complementary approaches that enhance the efficiency and coverage of WQ monitoring is essential. In this context, the use of orbital imagery emerges as a promising alternative to address these gaps and improve continuous WQ assessment. However, orbital imagery still faces constraints related to spectral, spatial, temporal, and radiometric resolutions, which may compromise analytical accuracy. The Planet nanosatellite constellation offers a viable solution, providing high revisit frequency and fine spatial resolution. Nonetheless, images acquired by PlanetScope (PS) sensors frequently exhibit radiometric inconsistencies due to variability among the constellation's satellites, which affects the temporal homogeneity of the data series. In this framework, the present study aimed to develop a methodology for high-frequency monitoring of water quality parameters (T, SST, Chla, P, N, COD, DO, and Fe) in lentic and lotic environments, integrating remote sensing products with machine learning techniques. The research was structured in two chapters: (i) Relative radiometric normalization of the PS nanosatellite constellation based on Sentinel-2 (S2) imagery; and (ii) Machine learning models for predicting water quality parameters in the Paraopeba River basin. In the first chapter, a radiometric normalization model for PS imagery was developed using S2 data, applying five machine learning algorithms: Multiple Linear Regression (MLR), Model Averaged Neural Network (avNNet), Random Forest (RF), k-Nearest Neighbors (kNN), and Support Vector Machines with a radial basis function kernel (SVM-RBF). The analysis used a dataset spanning from 2017 to 2022. The second chapter proposed a methodology for predicting WQ in lentic and lotic environments based on remote sensing and machine learning. Data from 24 WQ monitoring stations along the Paraopeba River and the Três Marias reservoir were used, covering the period from 2016 to 2023. Three distinct datasets were organized: (i) data derived from S2 imagery; (ii) data from raw PS images; and (iii) data from PS

images normalized using the methodology developed in Chapter 1. Four algorithms were evaluated: avNNet, RF, kNN, and SVM-RBF. Model performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Lin's Concordance Correlation Coefficient (CCC), and the coefficient of determination (R^2). The radiometric normalization developed for the PS sensor based on S2 imagery proved necessary and effective in mitigating inconsistencies within the Planet nanosatellite constellation. The SVM-RBF model achieved the best performance for radiometric normalization across all PS bands, particularly for the longer-wavelength bands (bands 3 and 4), with RMSE values of approximately 0.032 and 0.034, and R^2 and CCC ranging from 0.77–0.90 and 0.87–0.95, respectively. Normalized PS data outperformed both raw PS and S2 data in predicting WQ parameters. Among the machine learning models, RF demonstrated the greatest robustness compared to Cubist, kNN, and SVM-RBF. The proposed methodology, therefore, proved robust and capable of complementing traditional water quality monitoring, particularly for T, SST, Fe, P, and DO parameters, offering high temporal and spatial resolution. This approach represents a significant advancement in monitoring anthropogenic and natural impacts on aquatic ecosystems.

Keywords: inland waters; environmental disasters; extreme events; environmental monitoring ; water resources management

SUMÁRIO

1	INTRODUÇÃO GERAL	11
	REFERÊNCIAS.....	15
2	ARTIGOS CIENTÍFICOS	19
2.1	ARTIGO 1 - Normalização Radiométrica Relativa para a constelação de Nanossatélites PlanetScope com base em imagens Sentinel-2	19
2.1.1	INTRODUÇÃO	20
2.1.2	MATERIAL E MÉTODOS.....	23
2.1.3	RESULTADOS E DISCUSSÃO	35
2.1.4	CONCLUSÕES	45
	REFERÊNCIAS.....	46
2.2	ARTIGO 2 -Predição de parâmetros da qualidade da água na bacia do rio Paraopeba com base em produtos de sensoriamento remoto e aprendizado de máquina.....	53
2.2.1	INTRODUÇÃO	54
2.2.2	MATERIAL E MÉTODOS.....	57
2.2.3	RESULTADOS E DISCUSSÃO	71
2.2.4	CONCLUSÕES	86
	REFERÊNCIAS.....	87
3	CONCLUSÕES GERAIS	101
	APÊNDICE A – COVARIÁVEIS UTILIZADAS PARA IMPLEMENTAÇÃO DOS MODELOS	102

1 INTRODUÇÃO GERAL

As águas superficiais são fundamentais para a vida terrestre. Desta forma, o acesso à água potável, segura e limpa é essencial para a saúde humana, o desenvolvimento econômico e social e, além disso, é reconhecido como o requisito mais básico para sustentar a subsistência, sem distinção de nacionalidade, religião, cor ou riqueza de qualquer população (BARBOSA; NOVO; MARTINS, 2019; EVERARD, 2019; PAPA et al., 2023). No ano de 2015, dada a sua importância, a Organização das Nações Unidas (ONU) listou a garantia de acesso a água potável e ao saneamento para todos como um dos Objetivos de Desenvolvimento Sustentável (ODS - 6), dentre os 17 a serem alcançados até 2030 (ONU, 2015). Observando todas as variáveis que envolvem esse objetivo, será um desafio atingir essa meta, tendo em vista principalmente a escassez de água em diversas partes do mundo, a distribuição desigual dos recursos hídricos e a poluição das águas (BO et al., 2021).

Apesar do Brasil possuir aproximadamente 12% da disponibilidade de água doce do planeta, a água não é distribuída de forma igualitária pelo território nacional, tanto espacialmente como temporalmente. A região hidrográfica Amazônica detém 83% da disponibilidade hídrica de superfície do país, enquanto a região hidrográfica do Paraná, localizada principalmente na região Centro-Sul, é a segunda maior e concentra apenas cerca de 6% da disponibilidade hídrica. Outro ponto importante de se avaliar está relacionado com a densidade populacional de cada região. A região hidrográfica do Paraná abastece cerca de 32% da população brasileira, e a Amazônica, apenas cerca de 8% da população (ANA, 2017; BARBOSA; NOVO; MARTINS, 2019). Assim, mesmo com abundância hídrica em termos absolutos, a distribuição desigual intensifica os conflitos entre os múltiplos usuários da água em diversas regiões do país.

Buscando alternativas para a gestão dos recursos hídricos superficiais e visando mitigar os desafios relacionados à distribuição desigual da água, a partir do século XX, o monitoramento da qualidade das águas passou a ser mais sistemático e rigoroso, com o objetivo de assegurar a disponibilidade de água em quantidade e qualidade adequadas para os diversos usos (ENGMAN; GURNEY, 1991; SLOGGETT et al., 1995). O Monitoramento da qualidade das águas superficiais é de extrema importância, pois além do entendimento da situação real da água é possível avaliar as mudanças e tendências espaciais ocorridas ao longo da série histórica, identificar

lançamentos clandestinos, dar subsídios para a fiscalização e auxiliar o estabelecimento de políticas públicas para melhorar a gestão dos recursos hídricos (BILGIN; KONANÇ, 2016; CALAZANS et al., 2018; DONG et al., 2015).

Historicamente, a metodologia para a determinação dos parâmetros de qualidade da água é realizada por meio de coletas pontuais de amostras de água em campo e posterior análises em laboratórios (RAMARAJ; SIVAKUMAR, 2023; YANG et al., 2022). Ainda que o procedimento tradicional demonstre alta acurácia e seja reconhecido em todo o mundo, a implantação de uma rede de monitoramento da qualidade da água convencional requer alto investimento por parte do órgão gestor de recursos hídricos e, uma vez implementada, raramente são realizados estudos de reavaliação da eficácia do monitoramento (CALAZANS et al., 2018; CHAWLA; KARTHIKEYAN; MISHRA, 2020; TAVAKOL et al., 2017). Além disso, a metodologia convencional de amostragens pontuais não é capaz de identificar as mudanças efêmeras e espaço-temporais que ocorrem na qualidade da água, o que é vital para definir estratégias de gestão frente a eventos atípicos (DUAN et al., 2013a, 2013b; GHOLIZADEH; MELESSE; REDDI, 2016).

Tendo em vista todas as limitações da metodologia convencional, a utilização de técnicas de sensoriamento remoto vem sendo amplamente difundida nos últimos anos para fins de monitoramento contínuo de parâmetros de qualidade da água. As imagens orbitais oferecem uma cobertura sinóptica espacial e temporal do corpo d'água que é inatingível por métodos convencionais (ADJOVU et al., 2023; NOVO, 2010; TAVAKOL et al., 2017; TIAN et al., 2023; YIGIT AVDAN et al., 2019). Diversos pesquisadores veem dedicando seus esforços para investigar a utilização de imagens proveniente de sensores ópticos orbitais na predição de parâmetros de qualidade da água, tais como clorofila-a (Chla), turbidez (T), pH, matéria orgânica dissolvida colorida (CDOM), sólidos suspensos totais (SST) e profundidade do disco de Secchi (SSD) (ARANGO; NAIRN, 2020; DIAS et al., 2021; HAJIGHOLIZADEH et al., 2021; RAMARAJ; SIVAKUMAR, 2023; YANG et al., 2022; YIGIT AVDAN et al., 2019).

No entanto, o que se observa é que pesquisas que visam estimar parâmetros não opticamente ativos, como, por exemplo, fósforo (P), nitrogênio (N), demanda química de oxigênio (DQO), oxigênio dissolvido (OD) e ferro (Fe) ainda são escassas (SAGAN et al., 2020; TIAN et al., 2023). A principal explicação para essa escassez está na dificuldade de identificar características espectrais independentes nos dados dos sensores que possam ser correlacionadas diretamente com parâmetros não

opticamente ativos (FERDOUS; RAHMAN, 2020; SWAIN; SAHOO, 2017; XIONG et al., 2020). Assim, torna-se necessário explorar a alta correlação entre parâmetros não opticamente ativos e parâmetros opticamente ativos, investigar as relações com variáveis ambientais e empregar algoritmos de aprendizado de máquina para a estimativa dos parâmetros não ópticos (GAO et al., 2015; LI et al., 2017b; MATHEW; SRINIVASA RAO; MANDLA, 2017).

Outro aspecto que merece atenção é que, apesar dos avanços e dos benefícios oferecidos pelos sensores ópticos orbitais, ainda persistem limitações importantes na determinação de parâmetros de qualidade da água. Entre esses desafios, destacam-se a frequência de revisita dos satélites, as resoluções espacial e espectral das imagens, além da qualidade radiométrica dos dados gerados (ASNER, 2010; DEKKER, 2001; SADEH et al., 2021a). Um exemplo disso são as imagens Landsat, que possuem resolução de 30 m, porém, são limitadas a um período de revisita de 16 dias (KEITH et al., 2018; NASA, 2023). As imagens MODIS (*Moderate Resolution Imaging Spectroradiometer*) apresentam revisitas diárias a um mesmo ponto, mas possuem resolução espacial de 250 m a 1 km (LI et al., 2017a). Já as imagens Sentinel-2 (S2), que fazem parte do programa Copernicus, possuem resolução espacial de 10 a 60 m, revisitas a cada 5 dias a um mesmo ponto e excelente calibração radiométrica (CLAVERIE et al., 2018; ESA, 2023; JIANG et al., 2022).

Assim, uma alternativa para o monitoramento espaço-temporal mais preciso seria a utilização da constelação de nanossatélites PlanetScope (PS), a qual possui grande número de nanossatélites CubeSat 3U, fornecendo imagens diárias em quatro bandas espectrais distintas (vermelho - R, verde - G, azul - B e infravermelho próximo - NIR) com resolução espacial de 3,7 m desde o início de 2016 (PLANET TEAM, 2021). Porém, vários estudos tem indicado que as imagens PS apresentam inconsistências na calibração entre seus satélites, especialmente quando se pretende trabalhar com grandes series temporais (JIANG et al., 2022; LEACH; COOPS; OBRKNEZEV, 2019; SADEH et al., 2021b; WALDNER et al., 2019).

Diante do exposto, o uso de técnicas de sensoriamento remoto com dados de alta resolução temporal, espacial e radiométrica são características necessárias para estudar fenômenos e/ou processos efêmeros a partir do monitoramento dos parâmetros de qualidade da água. Um monitoramento contínuo e detalhado dos parâmetros de qualidade da água é uma ferramenta poderosa, especialmente quando

se deseja avaliar o impacto de eventos extremos, como o rompimento da barragem de rejeitos da mina que ocorreu no córrego do Feijão no município de Brumadinho/MG.

Sendo assim, o presente trabalho, cujo objetivo geral foi propor uma metodologia que permita o monitoramento com elevada frequência da qualidade da água (T, SST, Chla, P, N, DQO, OD e Fe) em ambientes lênticos e lóticos, baseada em técnicas de sensoriamento remoto e de aprendizado de máquina, foi subdividida em dois capítulos.

No primeiro capítulo, intitulado “Normalização Radiométrica Relativa para a constelação de Nanossatélites PlanetScope com base em imagens Sentinel-2”, foi proposta uma metodologia que permite corrigir as inconsistências das imagens PS com base na imagem de referência S2.

Enquanto no segundo capítulo, intitulado “Predição de parâmetros da qualidade da água na bacia do rio Paraopeba com base em aprendizado de máquina”, foi proposta uma metodologia alternativa para o monitoramento da qualidade das águas continentais por meio de imagens orbitais.

REFERÊNCIAS

- ADJOVU, G. E. et al. Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters. **Remote Sensing**, v. 15, n. 7, 2023.
- ANA. **Agência Nacional de Águas. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas**. Brasília - DF, 2017.
- ARANGO, J. G.; NAIRN, R. W. Prediction of Optical and Non-Optical Water Quality Parameters in Oligotrophic and Eutrophic Aquatic Systems Using a Small Unmanned Aerial System. **Drone**, 2020.
- ASNER, G. P. Cloud cover in Landsat observations of the Brazilian Amazon. **International Journal of Remote Sensing**, v. 22, p. 3855–3862, 2010.
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos**. 1º ed. São José dos Campos: INPE, 2019.
- BILGIN, A.; KONANÇ, M. U. Evaluation of surface water quality and heavy metal pollution of Coruh River Basin (Turkey) by multivariate statistical methods. **Environmental Earth Sciences**, v. 75, p. 1–18, 2016.
- BO, Y. et al. Additional surface-water deficit to meet global universal water accessibility by 2030. **Journal of Cleaner Production**, v. 320, n. August, p. 128829, 2021.
- CALAZANS, G. M. et al. The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. **Environmental Monitoring and Assessment**, v. 190, n. 8, 2018.
- CHAWLA, I.; KARTHIKEYAN, L.; MISHRA, A. K. A review of remote sensing applications for water security: Quantity, quality, and extremes. **Journal of Hydrology**, v. 585, n. March, p. 124826, 2020.
- CLAVERIE, M. et al. The Harmonized Landsat and Sentinel-2 surface reflectance data set. **Remote Sensing of Environment**, v. 219, n. August 2017, p. 145–161, 2018.
- DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing**. (P. V. U. Amsterdam, Ed.) **Proefschrift Vrije Universiteit Amsterdam**, 2001.
- DIAS, R. L. S. et al. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatics**, n. August, p. 101414, 2021.
- DONG, J. et al. A survey of smart water quality monitoring system. **Environmental Science and Pollution Research**, v. 22, p. 4893–4906, 2015.
- DUAN, W. et al. Spatiotemporal evaluation of water quality incidents in Japan between 1996 and 2007. **Chemosphere**, v. 93, n. 6, p. 946–953, 2013a.

DUAN, W. et al. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River , Japan , 1985 to 2010. **Science of the Total Environment**, The, v. 461–462, p. 499–508, 1 set. 2013b.

ENGMAN, E. T.; GURNEY, R. J. **Remote sensing in hydrology**. Londres: Chapman and Hall Ltd, 1991.

ESA. **Guia de Missão Sentinel 2.** , 2023. Disponível em: <<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>

EVERARD, M. Meeting global drinking water needs. **Nature Sustainability**, v. 2, n. 5, p. 360–361, 2019.

FERDOUS, J.; RAHMAN, M. T. U. Developing an empirical model from Landsat data series for monitoring water salinity in coastal Bangladesh. **Journal of Environmental Management**, v. 255, n. October 2019, p. 109861, 2020.

GAO, Y. et al. Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques. **Journal of Environmental Management**, v. 151, p. 33–43, 2015.

GHOLIZADEH, M. H.; MELESSE, A. M.; REDDI, L. **A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors (Switzerland)**MDPI AG, , 16 ago. 2016.

HAJIGHOLIZADEH, M. et al. Land–lake linkage and remote sensing application in water quality monitoring in lake okeechobee, florida, usa. **Land**, v. 10, n. 2, p. 147, 2021.

JIANG, J. et al. Multi-sensor and multi-platform consistency and interoperability between UAV, Planet CubeSat, Sentinel-2, and Landsat reflectance data. **GIScience and Remote Sensing**, v. 59, n. 1, p. 936–958, 2022.

KEITH, D. et al. Monitoring algal blooms in drinking water reservoirs using the landsat-8 operational land imager. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2818–2846, 2018.

LEACH, N.; COOPS, N. C.; OBRKNEZEV, N. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. **Computers and Electronics in Agriculture**, v. 164, n. January, p. 104893, 2019.

LI, J. et al. Recovering low quality MODIS-Terra data over highly turbid waters through noise reduction and regional vicarious calibration adjustment: A case study in Taihu Lake. **Remote Sensing of Environment**, v. 197, p. 72–84, 2017a.

LI, Y. et al. Monitoring spatiotemporal variations in nutrients in a large drinking water reservoir and their relationships with hydrological and meteorological conditions based on Landsat 8 imagery. **Science of the Total Environment**, v. 599–600, p. 1705–1717, 1 dez. 2017b.

MATHEW, M. M.; SRINIVASA RAO, N.; MANDLA, V. R. Development of regression equation to study the Total Nitrogen, Total Phosphorus and Suspended Sediment using remote sensing data in Gujarat and Maharashtra coast of India. **Journal of Coastal Conservation**, v. 21, n. 6, p. 917–927, 1 dez. 2017.

NASA. **Detalhes da missão do Landsat 8**. , 2023. Disponível em: <<https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-mission-details>>

NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 4. ed. São Paulo: Blucher, 2010.

ONU. **Objetivos de Desenvolvimento Sustentável**. , 2015.

PAPA, F. et al. **Water Resources in Africa under Global Change: Monitoring Surface Waters from Space**. [s.l: s.n.]. v. 44

PLANET TEAM. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA, USA. **Planet Labs Inc**, p. 1–96, 2021.

RAMARAJ, M.; SIVAKUMAR, R. Remote Sensing and Nonlinear Auto-regressive Neural Network (NARNET) Based Surface Water Chemical Quality Study: A Spatio-Temporal Hybrid Novel Technique (STHNT). **Bulletin of Environmental Contamination and Toxicology**, v. 110, n. 1, p. 1–10, 2023.

SADEH, Y. et al. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3 m surface reflectance and wheat LAI monitoring. **International Journal of Applied Earth Observation and Geoinformation**, v. 96, p. 102260, 2021a.

SADEH, Y. et al. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3 m surface reflectance and wheat LAI monitoring. **International Journal of Applied Earth Observation and Geoinformation**, v. 96, p. 102260, 2021b.

SAGAN, V. et al. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. **Earth-Science Reviews**, v. 205, p. 103187, 2020.

SLOGGETT, D. et al. Operational uses of ocean colour data-Perspectives for the OCTOPUS Programme. **Sensors and Environmental Applications of Remote Sensing**, 1995.

SWAIN, R.; SAHOO, B. Mapping of heavy metal pollution in river water at daily time-scale using spatio-temporal fusion of MODIS-aqua and Landsat satellite imageries. **Journal of Environmental Management**, v. 192, p. 1–14, 2017.

TAVAKOL, M. et al. Application of multivariate statistical methods to optimize water quality monitoring network with emphasis on the pollution caused by fish farms. **Iranian Journal of Public Health**, v. 46, n. 1, p. 83, 2017.

TIAN, S. et al. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. **Environmental Science and**

Pollution Research, v. 30, n. 7, p. 18617–18630, 2023.

WALDNER, F. et al. High temporal resolution of leaf area data improves empirical estimation of grain yield. **Scientific Reports**, v. 9, n. 1, p. 1–14, 2019.

XIONG, Y. et al. Remotely assessing and monitoring coastal and inland water quality in China: Progress, challenges and outlook. **Critical Reviews in Environmental Science and Technology**, v. 50, n. 12, p. 1266–1302, 2020.

YANG, H. et al. A review of remote sensing for water quality retrieval: progress and challenges. **Remote Sensing**, v. 14, n. 8, p. 1770, 2022.

YIGIT AVDAN, Z. et al. Monitoring the water quality of small water bodies using high-resolution remote sensing data. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 553, 2019.

2 ARTIGOS CIENTÍFICOS

2.1 ARTIGO 1 - Normalização Radiométrica Relativa para a constelação de Nanossatélites PlanetScope com base em imagens Sentinel-2

Resumo: A detecção e caracterização de mudanças contínuas na superfície terrestre se tornou uma ferramenta essencial para o planejamento e desenvolvimento da humanidade. Atenta a essas necessidades, a Planet vem lançando centenas de nanossatélites conhecidos como “Dove”. Apesar das vantagens proporcionadas pela resolução espacial e temporal dos nanossatélites, é notável que suas imagens ainda apresentam inconsistências na resolução radiométrica, limitando a sua utilização. Neste sentido, objetiva-se com o presente trabalho propor um modelo para a normalização radiométrica de imagens do sensor PlanetScope (PS), utilizando como referência imagens do sensor *Multispectral Instrument/Sentinel 2* (MSI/S2). Para isso, foi utilizado um extenso banco de dados, composto por imagens do sensor MSI/S2 (referência) e imagens adquiridas nas mesmas datas e em horários aproximados pelo sensor PS, abrangendo todas as gerações disponíveis (PS2, PSB.SD e PS2.SD), no período de 2017 a 2022. O processo de amostragem foi realizado banda a banda, sendo utilizado dois métodos: o *Conditioned Latin Hypercube Sampling* (cLHS) e visualização estatística. Em seguida, foram aplicados cinco algoritmos de aprendizado de máquina, abrangendo algoritmos lineares e não lineares, baseados em regras e em árvores de decisão, que são: *Multiple Linear Regression* (MLR), *Model Averaged Neural Network* (avNNet), *Random Forest* (RF), *k-Nearest Neighbors* (KNN) e *Support Vector Machines* com *Kernel* de Função de Base Radial (SVM-RBF). Para cada um desses modelos, realizou-se a seleção das covariáveis. Para avaliar o desempenho dos modelos, foram utilizados os seguintes índices estatísticos: raiz do erro quadrático médio – RMSE, erro médio absoluto – MAE, coeficiente de correlação de concordância de Lin – CCC e o coeficiente de determinação - R^2 . Além disso, foram aplicados os testes de Kruskal-Wallis e Dunn para definição do modelo com o melhor desempenho. Os resultados evidenciaram que o modelo SVM-RBF obteve o melhor ajuste aos dados para todas as bandas do sensor PS, produzindo resultados mais acurados para as bandas de maior comprimento de onda (banda 3 e 4). Nessas bandas, o RMSE variou na ordem de 0,032 e 0,034 e R^2 e CCC variando de 0,77 – 0,90 e 0,87 – 0,95, respectivamente. Com base nos resultados obtidos verificou-se que este estudo contribui expressivamente para otimizar a utilização das imagens do

sensor PS em diversas aplicações, fornecendo uma abordagem detalhada e robusta para a normalização radiométrica. Esses resultados têm implicações importantes para o monitoramento eficiente das mudanças na superfície terrestre e podem influenciar positivamente o uso desses dados em aplicações práticas e científicas.

Palavras-chave: CubeSat; Série temporal; Sensoriamento remoto; Aprendizado de máquina.

2.1.1 INTRODUÇÃO

A crescente preocupação com a segurança hídrica e ambiental despertou uma atenção especial para o monitoramento e a caracterização da superfície terrestre nos domínios do espaço e do tempo (BEVEN; CLOKE, 2012; YOUNG et al., 2021). Com isso, os sistemas de sensores embarcados em satélites tiveram um desenvolvimento exponencial nos últimos anos (HOUBORG; MCCABE, 2016; ROY et al., 2014). Contudo, a maioria dos sensores possuem limitações nas resoluções espectrais, espaciais e/ou temporais (LATTE; LEJEUNE, 2020). Devido a requisitos construtivos, os satélites projetados para capturar imagens de alta resolução espacial frequentemente apresentam baixa resolução temporal, uma vez que estas exigem um tempo de revisita mais longo (WEGMUELLER; LEACH; TOWNSEND, 2021).

Nesse contexto, a Planet desenvolveu uma extensa constelação de nanossatélites, chamados "*Dove*". As imagens fornecidas pelos nanossatélites proporcionam imagens diárias da superfície terrestre com resolução espacial de 3,7 metros. Esses satélites são compactos e de baixo peso (4 kg) e, desta forma, são de rápida construção e não requerem um veículo específico para o seu lançamento, uma vez que podem ser transportados como carga útil secundária em outras missões (PLANET TEAM, 2020). Devido a essas vantagens, o sensor PlanetScope (PS) tem sido amplamente utilizado em diversas aplicações ecológicas e ambientais, especialmente em casos em que é necessária a detecção de alterações momentâneas ou em áreas de pequena extensão (CSILLIK et al., 2019; FRANCINI et al., 2020; HELMAN et al., 2018; SHIN et al., 2021; WANG et al., 2021; WU et al., 2021).

Embora o sensor PS apresente grandes vantagens em termos de resolução espacial (3,7 m) e frequência de imagens (1 dia), ainda apresentam inconsistências

radiométricas, como constatado em alguns estudos, por exemplo, Houborg e McCabe, (2018a) e por Leach; Coops; Obrknezev (2019). Essas inconsistências dificultam sua utilização, pois um mesmo alvo invariante pode apresentar respostas radiométricas diferentes sob as mesmas condições ao longo da série histórica. Tais inconsistências podem ser atribuídas a dois fatores principais: a calibração dos sensores nas diferentes versões da plataforma Planet (PS2, PS2.SD e PSB.SD) e os efeitos atmosféricos e ambientais (FRAZIER; HEMINGWAY, 2021). Assim, uma alternativa para superar essas limitações é a aplicação de técnicas de correção radiométrica (XU et al., 2021).

Segundo Zhu (2017), a correção radiométrica se divide em dois grupos principais: correção radiométrica absoluta (CRA); e normalização radiométrica relativa (NRR). A correção radiométrica absoluta emprega parâmetros relacionados à correção atmosférica e dados síncronos de observação de satélite para realizar uma calibração precisa do sensor (YANG; LO, 2000). A aplicação da CRA ainda enfrenta obstáculos consideráveis devido à dificuldade na obtenção de parâmetros de correção atmosférica para imagens históricas, em função do baixo tempo de vida dos satélites (XU et al., 2021). Por outro lado, a NRR utiliza informações radiométricas derivadas das próprias imagens, realizando o ajuste dos valores dos pixels da imagem alvo para coincidir com os da imagem de referência (ZHANG; WU; DU, 2014).

O método NRR baseia-se na identificação por pares de alvos pseudo-invariantes entre as imagens, a partir dos quais é construído um modelo de normalização radiométrica (SYARIZ et al., 2019). Os objetos pseudo-invariantes, por definição, são caracterizados como alvos que possuem refletância relativamente constante durante as datas de aquisição de várias imagens, sendo de grande importância para estabelecer um modelo de normalização radiométrica (SCHOTT; SALVAGGIO; VOLCHOK, 1988). Houborg e McCabe (2018b) desenvolveram uma metodologia de normalização radiométrica para imagens PS, MODIS e Landsat, baseada em regras para derivar a refletância espectral, sendo utilizada com sucesso para monitorar mudanças em índices espectrais da agricultura na Arábia Saudita. Leach, Coops e Obrknezev (2019) tiveram êxito ao aplicar uma abordagem de normalização temporal de séries de imagens PS, com base em imagens Landsat, para áreas impactadas por incêndios florestais.

Embora diversas metodologias tenham sido propostas para a normalização de imagens PS, abordagens mais avançadas que envolvam a avaliação de algoritmos de

aprendizado de máquina, séries temporais PS completas e o sensor *Multispectral Instrument* S2 (MSI) como base para corrigir inconsistências radiométricas entre as aquisições do PS ainda são escassas na literatura. Para atender aos requisitos da NRR, propõe-se neste trabalho a utilização das imagens do sensor MSI/S2 como referência, fornecidas pela Agência Espacial Europeia (ESA). As imagens S2 destacam-se como uma escolha promissora para servir como sensor base na normalização, devido à sua excelente calibração radiométrica entre sensores, evidenciada em vários estudos (GASCON et al., 2017; LAMQUIN et al., 2019; LI et al., 2019; LOUIS et al., 2016). Além disso, o sensor MSI/S2 possui 13 bandas espectrais, frequência de imagens de 5 dias e resolução espacial variada, abrangendo intervalos de 10 a 60 metros (AGUIRRE-GUTIÉRREZ et al., 2021; DU et al., 2016; LATTE; LEJEUNE, 2020). Sendo assim, este estudo foi realizado com o objetivo de propor um método eficaz de normalização radiométrica para a constelação de nanossatélites PS, utilizando imagens do sensor MSI/S2 como referência.

2.1.2 MATERIAL E MÉTODOS

2.1.2.1 Área de estudo

Para a aplicação e a avaliação da metodologia proposta no presente trabalho foi utilizada uma área de aproximadamente 2.000 km², abrangendo parte do reservatório de Três Marias e de sua área de drenagem, situado na porção alta da bacia hidrográfica do São Francisco (Figura 1), na região central do estado de Minas Gerais - Brasil (ANA, 2017). O reservatório de Três Marias possui um espelho d'água com área de 1.054,6 km², abrangendo nove municípios do Estado e desempenhando um papel fundamental na economia da região (CEMIG, 2021).

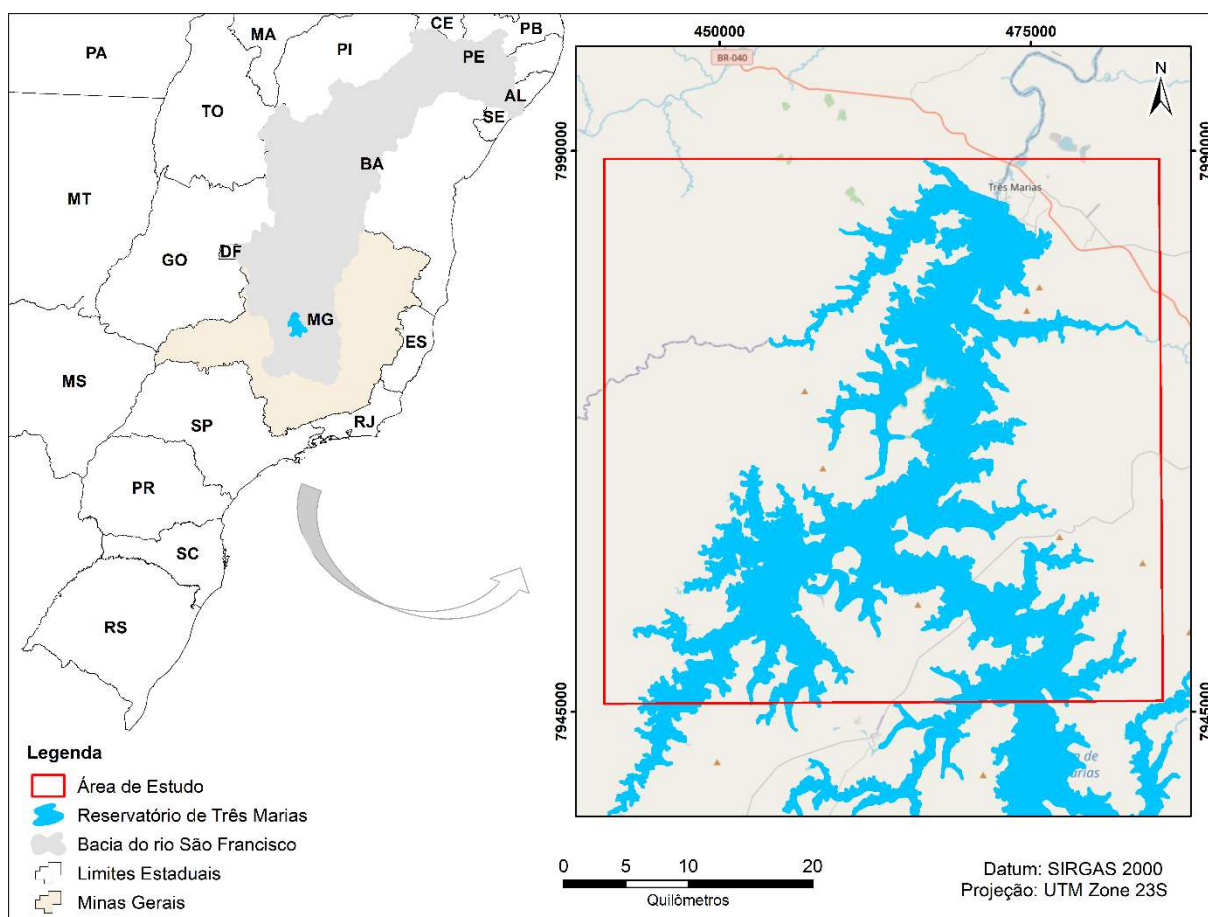


Figura 1 - Localização da área utilizada no presente estudo.

Essa região apresenta uma altitude média de 720 metros e relevo predominantemente forte ondulado a montanhoso. Em relação às classes de solo, observa-se o predomínio de Latossolos Vermelhos, Latossolos Vermelho-Amarelos, Cambissolos Háplicos e Cambissolos Húmicos (EMBRAPA, 2006). O clima é

classificado como Aw tropical, caracterizado por duas estações bem definidas, um verão chuvoso de outubro a março e um inverno seco de abril a setembro (ALVARES et al., 2014).

Esta área foi escolhida por apresentar grande variedade de tipos de solo, condições topográficas, cobertura vegetal e presença de corpos d'água continentais, fatores primordiais para a elaboração e construção de um modelo e para normalização de imagens de satélite (DE CARVALHO JÚNIOR et al., 2013; SCHOTT; SALVAGGIO; VOLCHOK, 1988).

2.1.2.2 Dados espectrais

Para o presente estudo foram utilizadas imagens do sensor *Multispectral Instrument* (MSI), embarcado nos satélites Sentinel 2A (S2A) e Sentinel 2B (S2B), como referência para a normalização radiométrica. Foram utilizadas as imagens do sensor PlanetScope, presente nas três diferentes gerações do CubeSat Planet (PS2, PS2.SD e PSB.SD), como imagem alvo. Na sequência são apresentados os detalhes dos produtos utilizados, bem como dos processamentos aplicados no presente estudo.

2.1.2.3 Imagens PlanetScope

A Planet é uma empresa comercial que opera uma constelação de pequenos nanossatélites com fator de forma CubeSat 3U ($10 \times 10 \times 30$ cm) de aproximadamente 4 kg de massa. Atualmente, a Planet opera mais de 180 nanossatélites '*Doves*', que fornecem imagens diárias da superfície terrestre com resolução espacial de 3,7m. As cenas são capturadas em bandas de comprimento de onda do vermelho (610 - 700 nm), verde (500 - 590 nm), azul (420 - 530 nm), infravermelho próximo (780 - 860 nm) (NIR) e resolução espacial de aproximadamente 3,7 m. Segundo a Planet Team (2021), as imagens PS são fornecidas ortorretificadas na projeção *Universal Transverse Mercator* (UTM) e corrigidas geometricamente, com precisão posicional de 10 m.

2.1.2.4 Imagens Sentinel-2

Os satélites S2A e S2B são conhecidos como satélites gêmeos que estão na mesma órbita e separados 180° um do outro. Foram lançados em 23 de junho de 2015 e 7 de março de 2017, respectivamente, pela Agência Espacial Europeia (ESA), e

estão em operação até hoje (MÜLLER-WILM, 2016). Com o trabalho combinado dos satélites (S2A ou S2B), as imagens S2 apresentam frequência de imagens de 5 dias (AGUIRRE-GUTIÉRREZ et al., 2021; DU et al., 2016). Para esse estudo foram utilizadas 10 bandas espectrais do *Multispectral Instrument* S2 (MSI), sendo três bandas no comprimentos de onda do visível (centrados em 490 nm, 560 nm e 665 nm) com resolução espacial de 10 m; uma banda de infravermelho próximo (NIR) (centrada em 842nm) com resolução espacial de 10 m; quatro bandas no comprimentos de onda *red edge* (centradas em 705 nm, 740 nm, 783 nm, 865 nm) com resolução espacial de 20 m; e duas bandas no comprimento de onda do infravermelho de onda curta (SWIR) (centradas em 1610 nm e 2190 nm) com resolução espacial de 20 m.

2.1.2.5 Pré-processamento das imagens

As imagens do Sentinel-2 foram adquiridas diretamente no site da ESA (<https://scihub.copernicus.eu>). Para esse trabalho foram selecionadas 15 cenas com baixa cobertura de nuvem, do nível 1C (refletância do Topo da atmosfera), em diferentes estações do ano no período de 2017 a 2022.

As imagens adquiridas da ESA foram submetidas ao processo de correção atmosférica Acolite (VANHELLEMONT; RUDDICK, 2016), para remover os efeitos de atenuação causados por constituintes atmosféricos ativos, como dispersão molecular e de aerossol e absorção por vapor de água, ozônio, oxigênio e dióxido de carbono (GAO et al., 2009). Ressalta-se que esse processo de correção atmosférica é amplamente indicado para aplicação em águas continentais (LAI et al., 2022; SCHROEDER et al., 2006; SONG et al., 2020).

Com as imagens corrigidas, foi possível converter as imagens do produto nível 1C em nível 2A (refletância de superfície), reduzindo a influência das condições atmosféricas presentes no momento de aquisição das cenas (MARTINS et al., 2017). Em seguida, as bandas do sensor *MSI* foram empilhadas, sendo descartadas as bandas 1 (Aerossol), banda 9 (Vapor de água) e banda 10 (*Cirrus*). Por fim, realizou-se o recorte para o quadrante de interesse utilizando o pacote “*terra*” em linguagem *R* (HIJMANS et al., 2022; R CORE TEAM, 2020).

Para aquisição das imagens PS foi utilizada a Interface de Programação de Aplicativos (API) disponível em <https://www.planet.com/explorer>. Foram selecionadas cenas das três gerações de sensores do PS disponíveis (PS2, PS2.SD e PSB.SD).

Todas as cenas escolhidas apresentam baixa nebulosidade e foram capturadas na mesma data e em horários semelhantes de imageamento do Sentinel-2 (S2). Destaca-se que, devido à captura em faixas das imagens PS, foi necessário realizar mosaico das diversas faixas para compor uma única cena. É importante destacar também que as imagens do PlanetScope não são gratuitas, no entanto, podem ser adquiridas sem custo com afiliação universitária e inscrição no programa *Planet Education and Research* em go.planet.com/research.

Posteriormente, realizou-se o corregristo das imagens PS. O corregristo de imagens entre sensores é um requisito essencial ao lidar com dados de sensoriamento remoto. O registro incorreto, seja inter ou intrassensorial, afetará qualquer análise subsequente da imagem, especialmente no processamento de dados multissensoriais ou multitemporais (SCHEFFLER et al., 2017).

Com o objetivo de minimizar erros de corregristo entre sensores diferentes, cenas PS consecutivas foram alinhadas por meio da otimização de sua autocorrelação espacial, conforme descrito por Houborg e McCabe (2018b). Essa técnica automatizada realiza a correção dos deslocamentos de pixel nas direções x e y em relação a uma imagem de referência, maximizando a qualidade do ajuste.

Na Tabela 1 pode-se observar que a diferença de horário entre o imageamento dos satélites PS e S2 foi baixa minimizando, portanto, a variação da condição atmosférica e as disparidades nas condições do ambiente aquático, contribuindo, dessa maneira, para minimizar as incertezas na comparação entre os diferentes sensores (PAHLEVAN et al., 2019).

Tabela 1: Informações dos dados adquiridos nas plataformas PlanetScope e Sentinel-2 (continua)

Data	PlanetScope			Sentinel-2	
	Instrumento	Satélite	Hora	Satélite	Hora
30/08/2017	PS2	0e1f	12:36:33	S2A	13:12:42
		100b	12:31:01		
		103d	12:29:14		
		1050	14:26:19		
09/10/2017	PS2	0f52	12:31:55	S2A	13:14:11
		0f34	12:32:49		
		0f17	12:37:36		
28/12/2017	PS2	103b	12:36:13	S2A	13:12:36
		0f43	12:35:38		
		1006	12:42:16		
01/07/2018	PS2	0f44	13:46:46	S2B	13:12:41
		0f3c	13:44:32		
		1013	12:43:21		

Tabela 1: Informações dos dados adquiridos nas plataformas PlanetScope e Sentinel-2
(conclusão)

29/09/2018	PS2	104b	13:33:25	S2A	13:12:39
		0f31	12:45:17		
		0e2f	12:43:05		
		0e26	12:42:00		
22/01/2019	PS2	1039	12:49:27	S2B	13:12:39
		1050	13:15:15		
27/04/2019	PS2	105F	12:59:14	S2A	12:49:02
		1105	12:52:22		
		1005	12:51:38		
27/05/2019	PS2	104e	12:56:55	S2A	13:12:49
		0f25	12:52:14		
		1004	12:50:25		
19/09/2019	PS2	101b	12:50:54	S2B	13:12:42
		0f17	12:56:21		
		103c	12:52:40		
17/03/2020	PS2	1040	12:57:04	S2A	13:12:41
	PSB.SD	2275	12:26:06		
28/09/2020	PS2.SD	2304	12:32:47	S2B	13:12:45
		106d	13:59:23		
		0f2a	11:39:15		
31/01/2021	PS2	1052	11:18:03	S2A	13:12:43
		241c	13:15:52		
		2264	12:32:51		
22/03/2021	PSB.SD	2416	13:17:35	S2A	13:12:41
	PS2	0f44	11:09:09		
	PS2.SD	106c	13:36:00		
18/09/2021	PSB.SD	2402	13:15:32	S2B	13:14:17
		2405	13:14:38		
		2262	12:35:07		
		2432	12:28:13		
27/03/2022	PSB.SD	2416	13:08:18	S2A	13:12:51
	PSB.SD	2442	12:17:24		

2.1.2.6 Amostragem dos dados

Os procedimentos para identificação dos alvos pseudo-invariantes foram realizados em duas etapas para coleta de amostras: a primeira consistiu na identificação dos alvos pseudo-invariantes através do método de visualização estatística, banda a banda, de forma individualizada (DU; TEILLET; CIHLAR, 2002). A segunda consistiu na aplicação do método automático *Conditioned Latin Hypercube Sampling* (cLHS) (MCKAY; BECKMAN; CONOVER, 1979; MINASNY; MCBRATNEY, 2006). O cLHS foi escolhido devido à sua capacidade de preenchimento de espaço, proporcionando resultados de alto desempenho com um conjunto menor de dados. Isso reduz tanto o custo de armazenamento quanto o tempo de treinamento, motivo pelo qual o cLHS é considerado um dos mecanismos mais avançados em termos de amostragem (IORDANIS; KOUKOUVINOS; SILOU, 2024).

Para o primeiro método foram utilizadas as 15 imagens S2 com as bandas empilhadas. Inicialmente, converteu-se os valores dos pixels em pontos de forma individualizada para cada banda. Posteriormente, calculou-se o desvio padrão entre o mesmo pixel nos diferentes anos. Então, foi realizado o somatório dos desvios para todas as bandas de um mesmo bloco e calculado o desvio padrão. Os pixels que apresentaram menores valores de desvio padrão foram analisados individualmente, obedecendo aos critérios de pixels pseudo-invariantes. Nesses critérios de seleção dos pixels pseudo-invariantes são analisadas as pequenas mudanças na elevação, para as quais a espessura da atmosfera sobre cada alvo é aproximadamente a mesma; quantidades mínimas de vegetação, devido à alta suscetibilidade de mudanças ao longo do tempo; posicionamento em áreas planas, para minimizar a influência das mudanças no ângulo do sol entre as imagens; e uma ampla variedade de valores de brilho (DE CARVALHO JÚNIOR et al., 2013; SCHOTT; SALVAGGIO; VOLCHOK, 1988).

Para a definição das amostras utilizando o método cLHS foram utilizadas as 15 cenas S2 abrangendo diferentes períodos ao longo dos anos de 2017 e 2022. O processo considerou 1.000.000 iterações, sendo monitorado o ponto de convergência da função objetiva do método Metropolis Hastings, garantindo estabilidade e identificação de pontos representativos de um hipercubo latino. O conjunto de dados ideal, selecionado por meio da aplicação do cLHS, foi obtido quando se alcançou o estado estacionário da função objetivo, que ocorreu com 500.000 iterações (MCKAY; BECKMAN; CONOVER, 1979; MINASNY; MCBRATNEY, 2006; SENA et al., 2021).

2.1.2.7 Modelagem para normalização radiométrica

Visando obter estimativas robustas nos modelos de normalização radiométrica, a base de dados foi subdividida em 4 conjuntos: B1 – Azul, B2 – Verde, B3 – Vermelho e B4 – NIR. Dessa forma, construiu-se um modelo preditivo individual para cada banda. Na Figura 2 está apresentado o fluxograma das etapas realizadas no processo de normalização radiométrica da constelação PS, utilizando como referência o sensor MSI/S2.

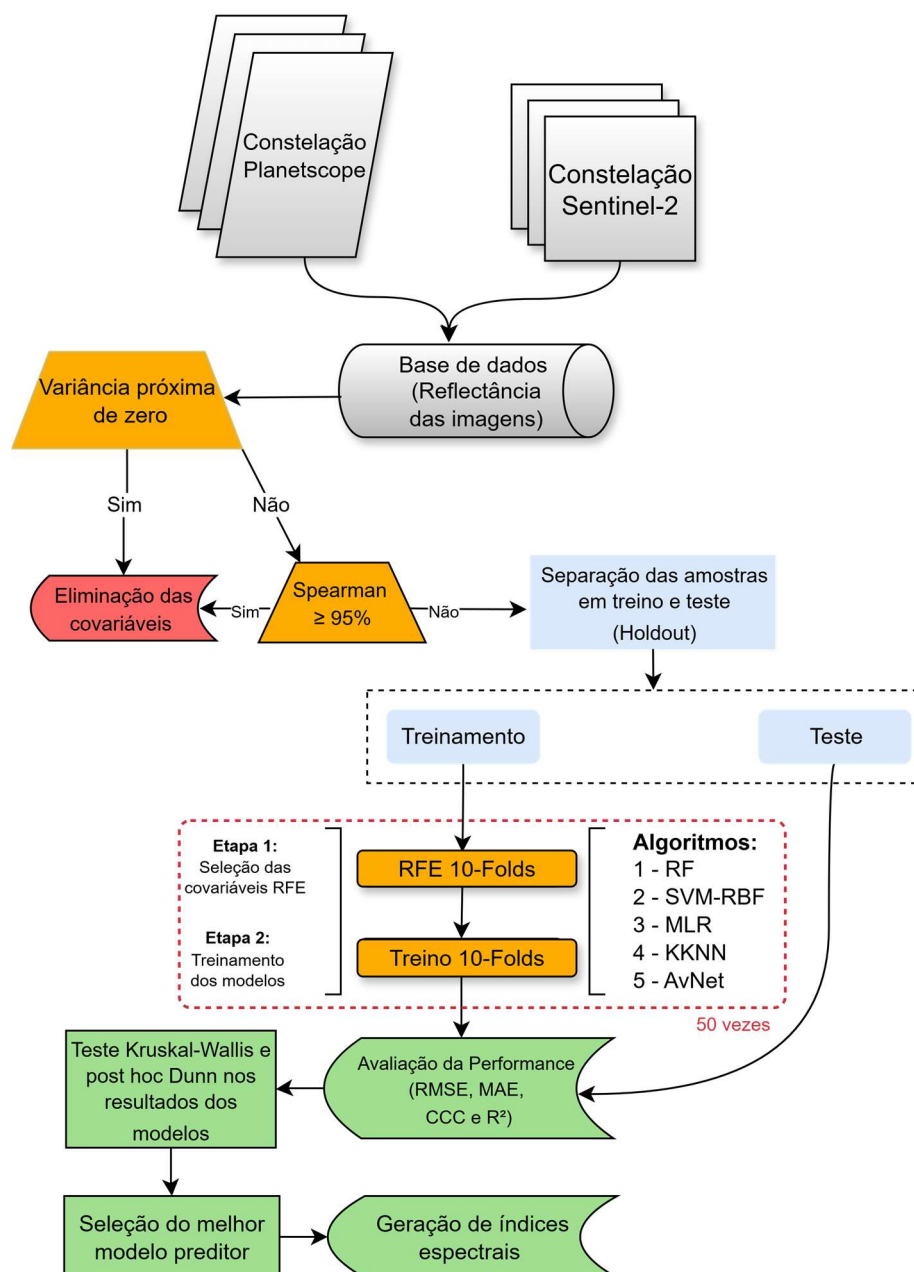


Figura 2: Fluxograma esquemático das etapas realizadas no processo de a normalização radiométrica da Constelação PlanetScope (PS).

Devido à grande quantidade de dados coletados, foi necessário aplicar o processo de subamostragem. A subamostragem é um método que permite aproximar a distribuição amostral dos dados originais, reduzindo a quantidade total de dados disponíveis para treinamento e, conseqüentemente, diminuindo o tempo de processamento, sem comprometer o desempenho do modelo final (POLITIS et al., 1999). As amostras selecionadas foram subdivididas aleatoriamente, sendo 5% destinadas ao treinamento e 95% ao teste, e esse procedimento foi repetido 50 vezes.

É importante salientar que, com a utilização do cLHS, os 5% selecionados para o treinamento conseguem representar a variabilidade do conjunto amostral de forma igual. Além disso, a utilização de um conjunto de treinamento maior elevaria os custos computacionais exigidos para processar o conjuntos de dados (WANG et al., 2020).

Para modelar a normalização dos dados PS, foram selecionados cinco algoritmos diferentes de aprendizado de máquina: *Multiple Linear Regression* – MLR (BEYCA et al., 2019), *Model Averaged Neural Network* – avNNet (RIPLEY, 1996), *Random Forest* – RF (BREIMAN, 2001), *k-Nearest Neighbors* – KKNn (ZHANG; ZHOU, 2007) e *Support Vector Machines* com *Kernel* de Função de Base Radial – SVM-RBF (CORTES; VAPNIK, 1995). Esses algoritmos têm como principal diferença a forma como gerenciam os dados em seu conjunto de treinamento (MURPHY, 2013). Além disso, apresentam uma variedade de regras que os classificam como algoritmos lineares, não lineares, baseados em regras, em árvore de decisão, kernel e redes neurais, alguns dos quais são capazes de estimar valores fora do intervalo dos dados de treinamento (DIAS et al., 2021). O teste de algoritmos com diferentes regras favorece a identificação daquele que melhor se ajusta ao conjunto de dados disponível (KUHN; JOHNSON, 2013).

Durante o treinamento, os hiperparâmetros internos de cada modelo foram ajustados utilizando validação cruzada repetida (*repeated cross-validation*), com 10 dobras (*folds*) e 10 repetições, aplicadas a cada hiperparâmetro de cada algoritmo (*tuneLength*). Os valores iniciais e as faixas de busca de cada parâmetro interno foram definidos de acordo com as recomendações dos desenvolvedores do pacote caret, conforme descrito no manual em seu capítulo 6, “Modelos descritos”, disponível em <https://topepo.github.io/caret/available-models.html>. Os valores dos hiperparâmetros finais, obtidos após a otimização, estão detalhados na Tabela 2.

Tabela 2: Hiperparâmetros otimizados para cada modelo desenvolvido.

Bandas	KKNn (kmax/distance/kernel)	RF (mtry)	SVM- RBF(sigma/C)	avNNet (size/decay)
B2	13 / 2 / optimal	2	1,6528 / 4	9 / 0.001
B3	13 / 2 / optimal	3	2,0429 / 4	7 / 0.001
B4	13 / 2 / optimal	2	1,9495 / 4	9 / 0.001
B8	13 / 2 / optimal	2	0,9831 / 4	9 / 0.001

2.1.2.8 Seleção das covariáveis

A seleção de covariáveis é frequentemente usada para reduzir a dimensionalidade dos dados e remover covariáveis ruidosas, redundantes e com baixa variância; aumentando, assim, a parcimônia dos modelos (DIAS et al., 2021; MUÑOZ-ROMERO et al., 2020; REUNANEN, 2003). Desta forma, esse procedimento foi realizado visando selecionar um subconjunto das covariáveis originais que minimizasse a redundância e maximizasse a relevância. Isso resulta na redução do número de covariáveis dos modelos, utilizando apenas aquelas mais importantes para prever a variável dependente em análise.

Inicialmente, foi aplicado o processo de eliminação das covariáveis com variância nula, uma vez que não contribuíam expressivamente para o desempenho do modelo (DA SILVEIRA et al., 2022). Neste processo utilizou-se a função “*nearZeroVar*” do pacote *Caret* (KUHN, 2020).

Posteriormente, aplicou-se o coeficiente de correlação de *Spearman* (LEE RODGERS; NICEWANDER, 1988) e as covariáveis com correlação igual ou superior a 95% foram avaliadas em pares, sendo removida aquela que apresentam o maior valor absoluto de correlação com todas as outras covariáveis. Esses procedimentos foram aplicados para as amostras coletadas das 4 bandas do sensor PS (B1 – Azul, B2 – Verde, B3 – Vermelho e B4 – NIR).

2.1.2.9 Avaliação dos modelos

Com o objetivo de avaliar o desempenho dos modelos ajustados para cada banda, os valores normalizados nas imagens do sensor, obtidos via modelagem, foram comparados aos valores observados nas imagens de referência S2, por meio das seguintes métricas estatísticas: raiz do erro quadrático médio – RMSE (Equação 1), erro médio absoluto – MAE (Equação 2), coeficiente de correlação de concordância de Lin – CCC (Equação 3) e o coeficiente de determinação - R^2 (Equação 4), (ARLOT; CELISSE, 2010; KENNEDY; NEVILLE, 1986; LIN, 1989).

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad (2)$$

$$CCC = \frac{2rV_rV_R}{(\bar{P} - \bar{O})^2 + V_r + V_R} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

Onde: P_i representa os valores preditos pelo modelo; O_i representa os valores observados; $\bar{O}$ representa a média dos valores observados; $\bar{P}$ representa a média dos valores preditos; V_r e V_R são as variações dos valores preditos e estimados, respectivamente e n representa o número de pares de observações.

O conjunto de métricas supracitado foi adotado para envolver métricas com características diferentes na avaliação dos modelos. O RMSE eleva ao quadrado a diferença entre os valores preditos e observados, tornando-o um bom indicador de valores discrepantes em dados preditos (CHAI; DRAXLER, 2014). Por outro lado, o MAE não apresenta grande sensibilidade a outliers (WILLMOTT; MATSUURA, 2005). O CCC e R^2 foram utilizados para avaliar o comportamento da correlação entre os valores normalizados, obtidos via modelagem, com os valores observados nas imagens de referência S2. A principal diferença entre o CCC e o R^2 está nos critérios locais e de escala, visto que o CCC mede a proximidade da linha de melhor ajuste do modelo a uma reta de 45 graus (1:1) partindo da origem (LIN, 1989; MORSE-MCNABB; HASAN; KARUNARATNE, 2023). Segundo Altman (1990), o CCC e R^2 podem ser interpretados de forma semelhante, sendo considerados moderados para valores acima de 0,5, fortes para valores acima de 0,7 e muito fortes para valores acima de 0,9.

Para fins de comparação, calculou-se, ainda, os valores dos índices estatísticos RMSE e MAE (Equação 5 e 6) para o modelo nulo. O modelo nulo (NULL) considera o uso do valor médio das amostras coletadas. Ele representa a abordagem mais simples, retornando uma única média para os resultados numéricos com base no conjunto de treinamento. Modelos com desempenho semelhante ou pior em relação ao modelo nulo são descartados. Os melhores modelos devem mostrar RMSE e MAE inferiores aos valores de NULL MAE e NULL RMSE, indicando que o modelo é

superior à simples utilização dos valores médios, evidenciando ganho na utilização do modelo de aprendizado de máquina (MELLO et al., 2022).

$$NULL_RMSE = \left[\frac{1}{n} \sum_{i=1}^n (\bar{O}_{mt} - O_{iT})^2 \right]^{\frac{1}{2}} \quad (5)$$

$$NULL_MAE = \frac{\sum_{i=1}^n |\bar{O}_{mt} - O_{iT}|}{n} \quad (6)$$

Onde: $\bar{O}_{mt}$ representa a média das amostras de treinamento; O_{iT} representa os valores das amostras teste e n representa o número de amostras (loop).

Por fim, os modelos foram comparados entre si por meio de testes estatísticos não paramétricos. Como cada modelagem gerou 50 valores para cada métrica estatística (RMSE, MAE, CCC e R^2), o objetivo desta etapa foi avaliar se os modelos apresentaram diferenças estatisticamente significativas e identificar quais modelos apresentam diferenças estatísticas significativas entre si. Para isso, foram utilizados o teste de Kruskal-Wallis (1952) (teste não paramétrico para três ou mais grupos de variáveis contínuas) e o teste de Dunn (1964) (teste *post hoc pairwise* do teste de Kruskal-Wallis), adotando-se um intervalo de confiança de 95% (MCKIGHT; NAJAB, 2010).

2.1.2.10 Índices Espectrais

Com base nos modelos gerados para cada banda e com o objetivo de realizar uma avaliação abrangente e aplicada, foram escolhidos cinco índices espectrais: *Normalized Difference Vegetation Index* – NDVI (Equação 7), *Soil-adjusted Vegetation Index* – SAVI (Equação 8), *Normalized Difference Water Index* – NDWI (Equação 9), *Normalized Difference Pond Index* – NDTI (Equação 10) e *Iron* (Equação 11). Para a geração desses índices utilizou-se o modelo que apresentou o melhor desempenho na comparação entre os valores preditos e observados. Este modelo foi submetido a 50 iterações, gerando assim 50 valores para cada pixel. O valor final atribuído a cada pixel foi obtido através do cálculo da média dos valores resultantes (HEWSON;

CUDAHY; HUNTINGTON, 2001; HUETE, 1988; LACAUX et al., 2007; MCFEETERS, 1996a; ROUSE et al., 1974).

$$NDVI = \frac{(B_{NIR} - B_{vermelho})}{(B_{NIR} + B_{vermelho})} \quad (7)$$

$$SAVI = \frac{(B_{NIR} - B_{vermelho})}{(B_{NIR} + B_{vermelho} + 0,5)} \times (1,5) \quad (8)$$

$$NDWI = \frac{(B_{verde} - B_{NIR})}{(B_{verde} + B_{NIR})} \quad (9)$$

$$NDTI = \frac{(B_{vermelho} - B_{verde})}{(B_{vermelho} + B_{verde})} \quad (10)$$

$$Iron = \frac{B_{vermelho}}{B_{azul}} \quad (11)$$

Onde: B_{Azul} representa a banda azul; B_{Verde} representa as bandas do verde; $B_{Vermelho}$ representa a banda vermelho; B_{NIR} representa as bandas do infravermelho

2.1.3 RESULTADOS E DISCUSSÃO

Na figura 3 são apresentados os pontos amostrais obtidos após a aplicação dos métodos de amostragem para uma cena. No total foram coletados 750.050 pontos amostrais nas 15 cenas selecionadas, sendo 750.000 pontos (representados em laranja) obtidos utilizando-se o método cLHS e 50 pontos (em verde) obtidos utilizando-se a avaliação estatística de pixels pseudo-invariantes banda a banda.

Embora o número de amostras obtidas por meio da identificação de alvos pseudo-invariantes seja pequeno em comparação com o total de amostras, tal abordagem é essencial no processo analítico. Utilizando este método, foi possível coletar amostras com valores próximos às mínimas e máximas refletâncias (0 e 1), que são de grande importância para o treinamento de modelos. Essas amostras proporcionam uma maior amplitude de dados, o que é fundamental para algoritmos baseados em árvores de decisão, os quais têm dificuldade em extrapolar o intervalo para o qual foram treinados (DIAS et al., 2021; ZHANG; LI; DENG, 2020).

Além disso, destaca-se que a quantidade de amostras coletadas representa 3,75% do total de pixels na área de estudo (20 milhões), reforçando uma das características principais do método cLHS, que é sua capacidade de otimizar a distribuição das amostras no espaço de dados, resultando em desempenho eficaz com um conjunto relativamente pequeno de dados em comparação com outros métodos de amostragem. Essa eficiência é crucial em aplicações práticas que envolvem grandes conjuntos de dados (MINASNY; MCBRATNEY, 2006).

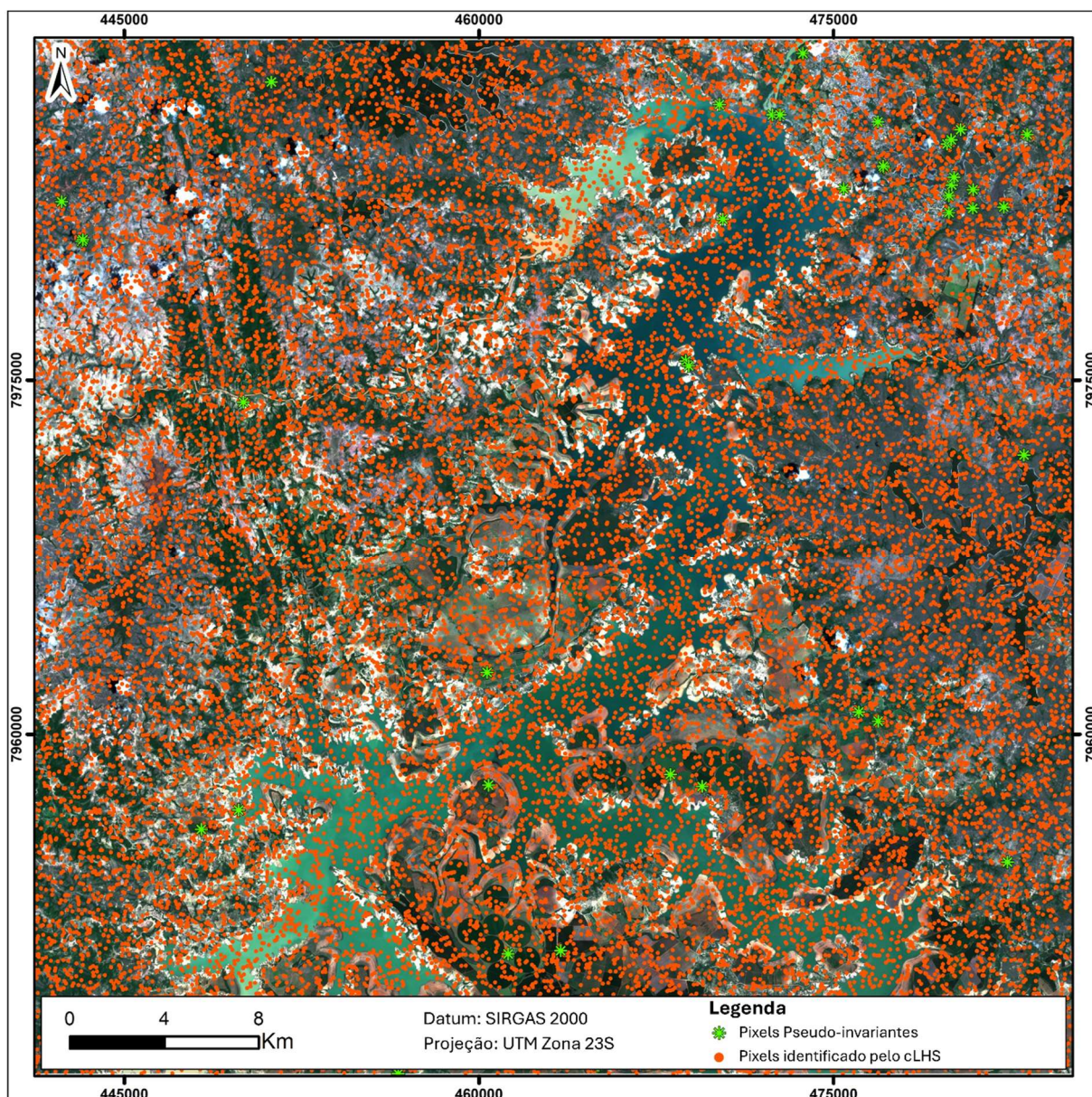


Figura 3: Espacialização das amostras no quadrante em estudo, onde os pontos em laranja representam os locais identificados no método cLHS e em verdes aqueles obtidos por meio da avaliação estatística de pixels pseudo-invariantes banda a banda.

Fonte: Imagem Sentinel-2

Na Figura 4, são apresentadas as covariáveis mais importantes para cada modelo nas quatro bandas analisadas do sensor *MSI/S2* (B1 – Azul, B2 – Verde, B3 – Vermelho e B4 – NIR). De forma geral, o processo de seleção busca identificar o menor número possível de covariáveis relevantes para os modelos. Neste caso, observou-se que, ao trabalhar com uma quantidade reduzida de covariáveis de entrada, não houve eliminação de nenhum dado. Ao analisar individualmente as bandas para a normalização do sensor PS, verifica-se que a para as 20 situações avaliadas, a banda 3 foi a variável selecionada de maior importância em 13 situações,

seguida pela banda 4 que foi a variável de maior importância em sete situações (Figura 4).

Avaliando a importância das covariáveis para os algoritmos de predição, observa-se que a covariável B3 foi a mais importante em três das quatro bandas para os algoritmos SVM-RBF, KKNN, avNNet e MLR. Enquanto para o algoritmo RF, a covariável mais importante foi a B4 em três das quatro bandas. Esse fato é atribuído à maneira como os modelos realizam o processo de generalização e separação dos conjuntos (DIAS et al., 2021). Esses resultados evidenciam-se a importância de investigar o comportamento das variáveis mais relevantes, fornecendo uma compreensão mais aprofundada do processo de normalização radiométrica. Essa análise oferece *insights* valiosos para potenciais otimizações e aprimoramentos em futuras aplicações desses modelos.

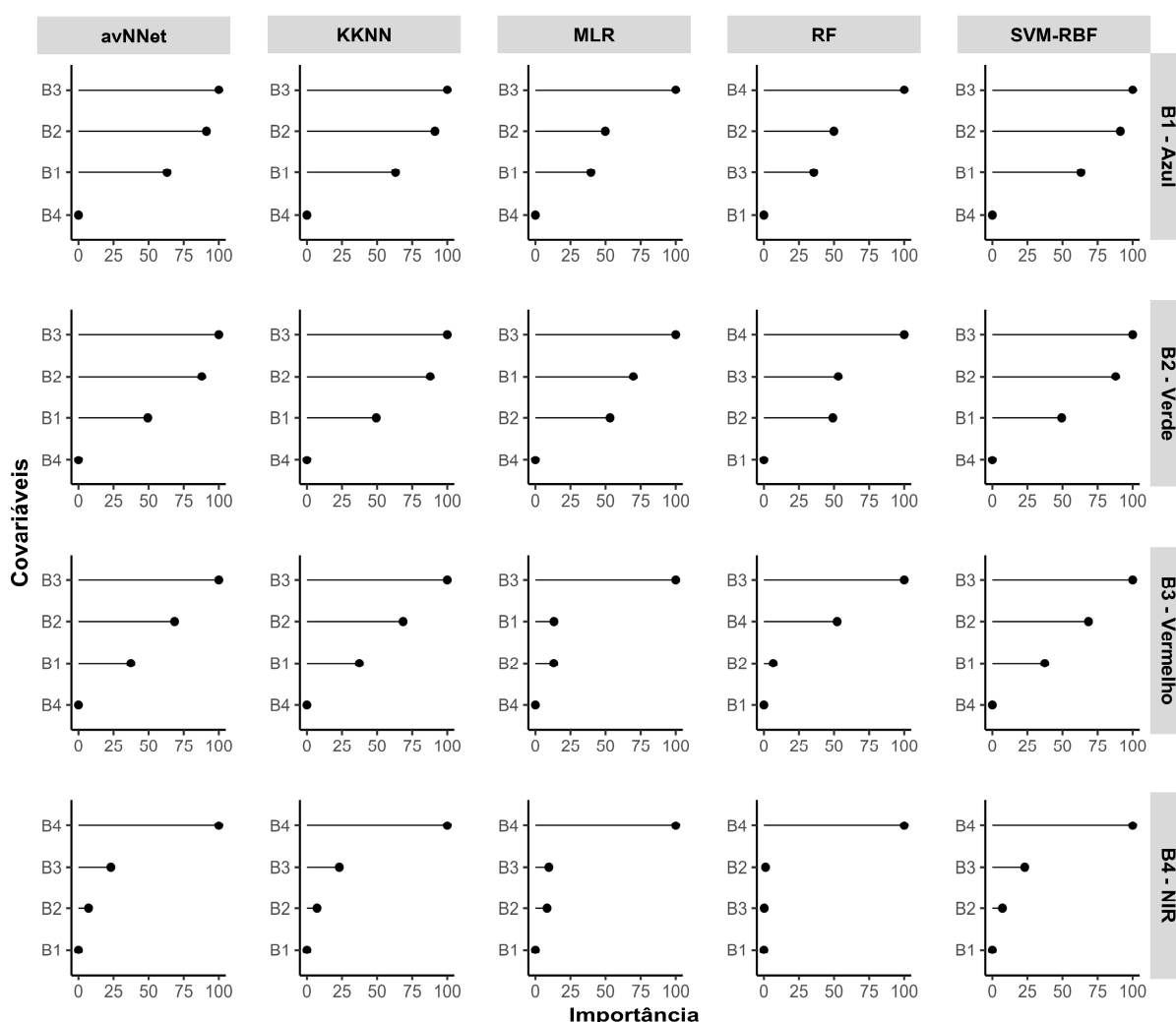


Figura 4: Importância das covariáveis para a normalização do sensor PlanetScope com base no sensor MSI, para os algoritmos RF, avNNet, KKNN, MLR e SVM-RBF aplicados nas bandas B1, B2, B3 e B4.

Na Figura 5 é apresentada a dispersão dos valores de refletância observados e preditos com a normalização radiométrica, em uma relação linear 1:1. Observa-se que as bandas com maior comprimento de onda exibem uma maior concordância com a relação linear 1:1 em todos os modelos ajustados, destacando assim a eficácia dos algoritmos implementados. É relevante destacar que o intervalo de valores para cada banda foi respeitado, indicando consistência da normalização radiométrica ao preservar as características específicas de cada faixa espectral. É possível observar também que os maiores erros absolutos de predição ocorrem com predominância nos pixels com valores mais elevados, o que pode ser atribuído à escassez de amostras representativas nessas regiões devido ao padrão dos dados trabalhados.

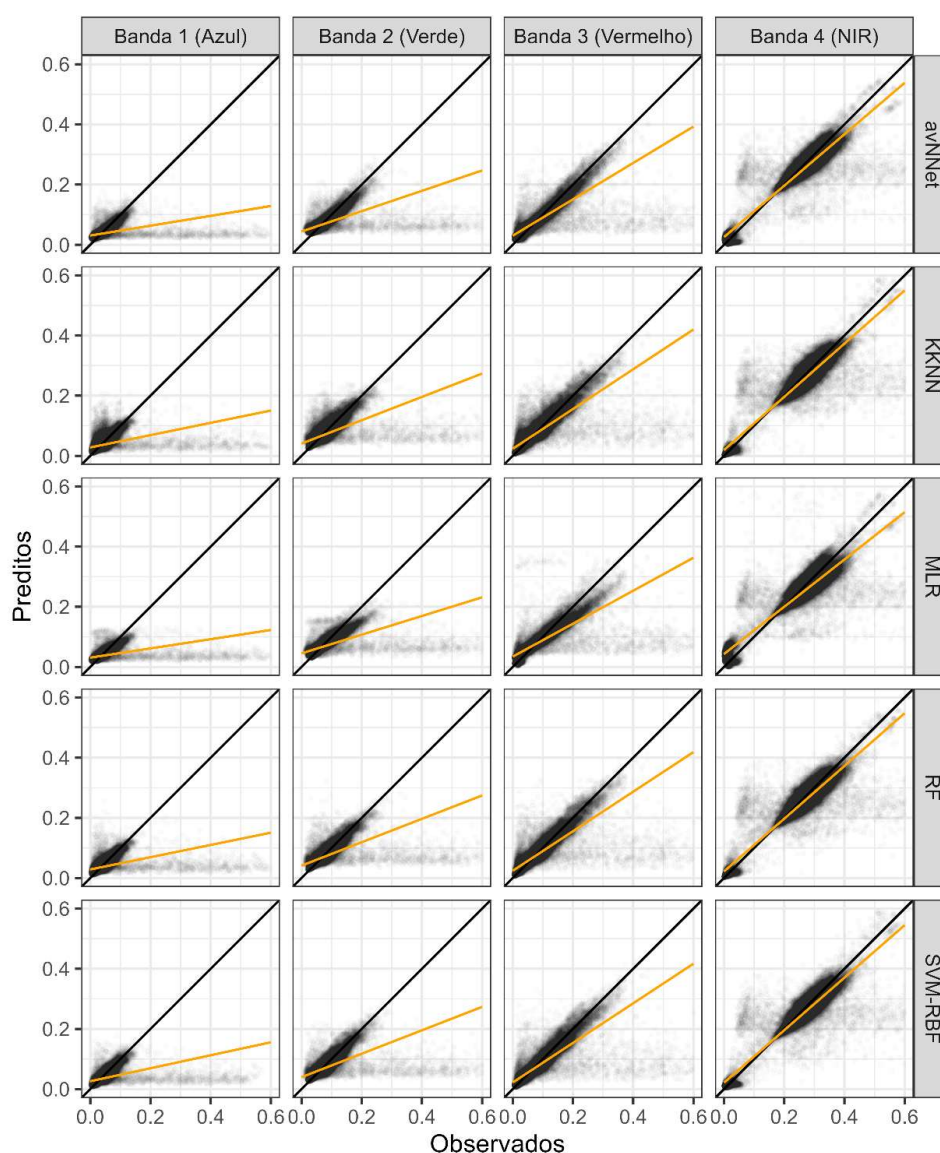


Figura 5: Comparação entre os valores de refletância preditos e observados para as bandas 1,2,3 e 4 utilizando os modelos de aprendizado de máquina RF, avNNet, KNN, MLR e SVM-RBF. Em preto a linha 1:1 e em dourado a linha de tendência.

Na Tabela 3 estão apresentados os resultados das métricas utilizadas para avaliação do desempenho dos modelos de aprendizado de máquina empregados na normalização radiométrica das imagens PS. Observa-se que as métricas de ajuste no treinamento dos modelos RF, avNNet, KNN, MLR e SVM-RBF obtidos com a validação cruzada repetida (*repeatedcv*) foram notavelmente semelhantes, indicando a robustez da modelagem, uma vez que a validação cruzada foi empregada para ajuste dos hiperparâmetro (ARLOT; CELISSE, 2010; DOS SANTOS et al., 2022). Além disso, todos os modelos apresentaram valores de MAE e RMSE inferiores ao modelo nulo (*Null Model*), ou seja, as estimativas dos modelos são mais indicadas do que adotar os valores médios; uma vez que o desempenho do *Null Model* é um limite mínimo de desempenho a ser alcançado pelos modelos (KUHN; JOHNSON, 2013; MELLO et al., 2022).

Os resultados obtidos mostram alta concordância entre os parâmetros de desempenho dos modelos nas etapas de treinamento e teste, diante da considerável variação nos conjuntos de entrada dos algoritmos, indicando uma redução expressiva do risco de *overfitting* quando da aplicação a um conjunto de dados desconhecidos. Essa conclusão é reforçada pela estratégia de teste *holdout*, na qual as estimativas de RMSE e MAE entre os conjuntos de treinamento e teste foram semelhantes. Esses resultados evidenciam que a técnica empregada revelou-se apropriada para o estudo proposto, proporcionando uma avaliação robusta e confiável do desempenho do modelo (ARLOT; CELISSE, 2010; DIAS et al., 2021; FILGUEIRAS et al., 2019).

É pertinente destacar que os coeficientes R^2 e CCC exibiram concordância (com valores de treinamento menores que os de teste) em ambos os conjuntos de dados para todas as bandas avaliadas. Esta observação ganha relevância ao considerar que o conjunto de dados foi dividido em 5% (37.537 amostras) para treinamento e 95% (713.213 amostras) para teste.

Conforme destacado por Wang et al. (2020), a utilização de subamostragem em casos de análise de grandes conjuntos de dados frequentemente resulta em desempenho preditivo superior quando comparada à abordagem de amostragem tradicional. Isso se deve à redução da redundância nos dados proporcionada pelo processo de subamostragem, simplificando a tarefa de modelagem e, conseqüentemente, melhorando a precisão das previsões. Além disso, destaca-se que, ao realizar o treinamento com uma subamostragem adequada, é possível reduzir

expressivamente os custos computacionais associados ao processamento de conjuntos de dados extensos.

Ao analisar os resultados dos índices estatísticos aplicados, destaca-se que os modelos que demonstraram melhor desempenho foram o RF e SVM-RBF para todas as bandas consideradas. Os modelos avNNet e KNN, embora não tenham superado RF e SVM-RBF, apresentaram boas métricas, indicando uma modelagem robusta ao conjunto de dados analisado. O modelo MLR, apesar de apresentar valores menos favoráveis, demonstrou potencial para ser utilizado, visto que sua aplicabilidade ocorre de forma direta através de uma equação linear, desde que os valores estejam dentro do intervalo desta pesquisa. Essa consideração destaca a importância de escolher o modelo mais apropriado, considerando não apenas o desempenho, mas também as características específicas do conjunto de dados e as demandas computacionais.

Avaliando os resultados de maneira individual para cada banda (tabela 3 e Figura 5), é evidente que as Banda 3 (Vermelho) e Banda 4 (NIR), com maiores comprimentos de onda, demonstram desempenho superior, refletido por baixos valores de RMSE e MAE e valores elevados de CCC e R^2 .

Os piores desempenho dos modelos para as Banda 1 (Azul) e Banda 2 (Verde) podem ser explicados pelo fenômeno de espalhamento *rayleigh* que atuam em comprimentos de onda mais curtos (BARBOSA; NOVO; MARTINS, 2019). Essa susceptibilidade ao espalhamento é atribuída a diversas razões intrínsecas ao espectro eletromagnético e à interação com a atmosfera. Além disso, devido à interação com as moléculas atmosféricas, essas bandas são mais suscetíveis à dispersão de luz pelas partículas do ar, característica fundamental do espalhamento *rayleigh* (LAI et al., 2022; SCHROEDER et al., 2006). Sendo assim, os modelos preditivos desenvolvidos para as Bandas 1 (Azul) e 2 (Verde) apresentaram valores inferiores em comparação com as bandas de maior comprimento de onda (Banda 3 e Banda 4).

Tabela 3: Métricas estatísticas dos modelos RF, avNNet, KKNn, MLR e SVM-RBF

Modelos	Treinamento				Teste					
	R ²	CCC	RMSE	MAE	R ²	CCC	RMSE	NULL RMSE	MAE	NULL MAE
Banda 1 (Azul) - Sensor PS										
SVM-RBF	0,325	0,488	0,0083	0,0268	0,311	0,475	0,0084		0,0272	
RF	0,316	0,496	0,0093	0,0269	0,306	0,488	0,0093		0,0273	
KKNn	0,291	0,486	0,0096	0,0275	0,280	0,476	0,0097	0,0327	0,0279	0,0175
avNNet	0,296	0,453	0,0106	0,0273	0,281	0,440	0,0106		0,0277	
MLR	0,276	0,430	0,0109	0,0276	0,262	0,416	0,011		0,0281	
Banda 2 (Verde) - Sensor PS										
SVM-RBF	0,5497	0,7081	0,0281	0,0107	0,5397	0,7021	0,0285		0,0107	
RF	0,5435	0,7079	0,0282	0,0115	0,5346	0,7030	0,0285		0,0115	
KKNn	0,5247	0,6991	0,0288	0,0119	0,5148	0,6933	0,0292	0,0418	0,0119	0,0265
avNNet	0,5013	0,6652	0,0294	0,0141	0,4897	0,6577	0,0299		0,0141	
MLR	0,4770	0,6441	0,0302	0,0147	0,4656	0,6361	0,0306		0,0147	
Banda 3 (Vermelho) - Sensor PS										
SVM-RBF	0,7712	0,8702	0,0317	0,0140	0,7674	0,8684	0,0320		0,0140	
RF	0,7680	0,8692	0,0318	0,0147	0,7650	0,8680	0,0321		0,0147	
KKNn	0,7589	0,8646	0,0324	0,0151	0,7548	0,8628	0,0328	0,0661	0,0151	0,0488
avNNet	0,7456	0,8530	0,0333	0,0178	0,7411	0,8507	0,0337		0,0178	
MLR	0,7189	0,8360	0,0350	0,0198	0,7140	0,8334	0,0354		0,0198	
Banda 4 (NIR) - Sensor PS										
SVM-RBF	0,9002	0,9471	0,0342	0,0183	0,8993	0,9467	0,0345		0,0184	
RF	0,8982	0,9464	0,0346	0,0188	0,8974	0,9462	0,0348		0,0188	
KKNn	0,8955	0,9452	0,0350	0,0190	0,8943	0,9447	0,0353	0,1086	0,0191	0,0832
avNNet	0,8915	0,9427	0,0357	0,0204	0,8907	0,9382	0,0367		0,0212	
MLR	0,8746	0,9330	0,0384	0,0242	0,8736	0,9326	0,0386		0,0242	

Onde: RMSE - raiz do erro quadrático médio, MAE - erro médio absoluto, CCC - coeficiente de correlação de concordância de Lin, R² - coeficiente de determinação e NULL RMSE – raiz do erro quadrático médio do modelo nulo e NULL MAE - erro médio absoluto do modelo nulo.

Nota: As linhas em negrito representam os modelos com melhor desempenho para cada parâmetro.

Os resultados do teste de Kruskal-Wallis aplicados às métricas de avaliação revelaram diferenças estatisticamente significativas entre os algoritmos testados para a normalização dos dados do sensor PS (Figura 6). Em seguida, foi realizado o teste de Dunn, que permitiu identificar que o modelo SVM-RBF se destacou, apresentando o melhor desempenho entre os algoritmos avaliados. Tal conclusão pode ser corroborada ao se analisar o conjunto completo de índices, visto que, nas bandas B1 (Azul) e B2 (Verde), o modelo gerado pelo RF obteve o melhor valor de CCC. No entanto, ao se considerar as demais métricas de desempenho (MAE, RMSE e R²), observa-se que o SVM-RBF apresenta o melhor desempenho. Esses resultados evidenciam que, apesar de um bom ajuste em algumas métricas específicas, o SVM-

RBF se destacou na capacidade preditiva geral, minimizando os erros e oferecendo melhor generalização dos dados. A análise abrangente dessas métricas fortalece a escolha do SVM-RBF como o modelo mais eficaz para o conjunto de dados em questão.

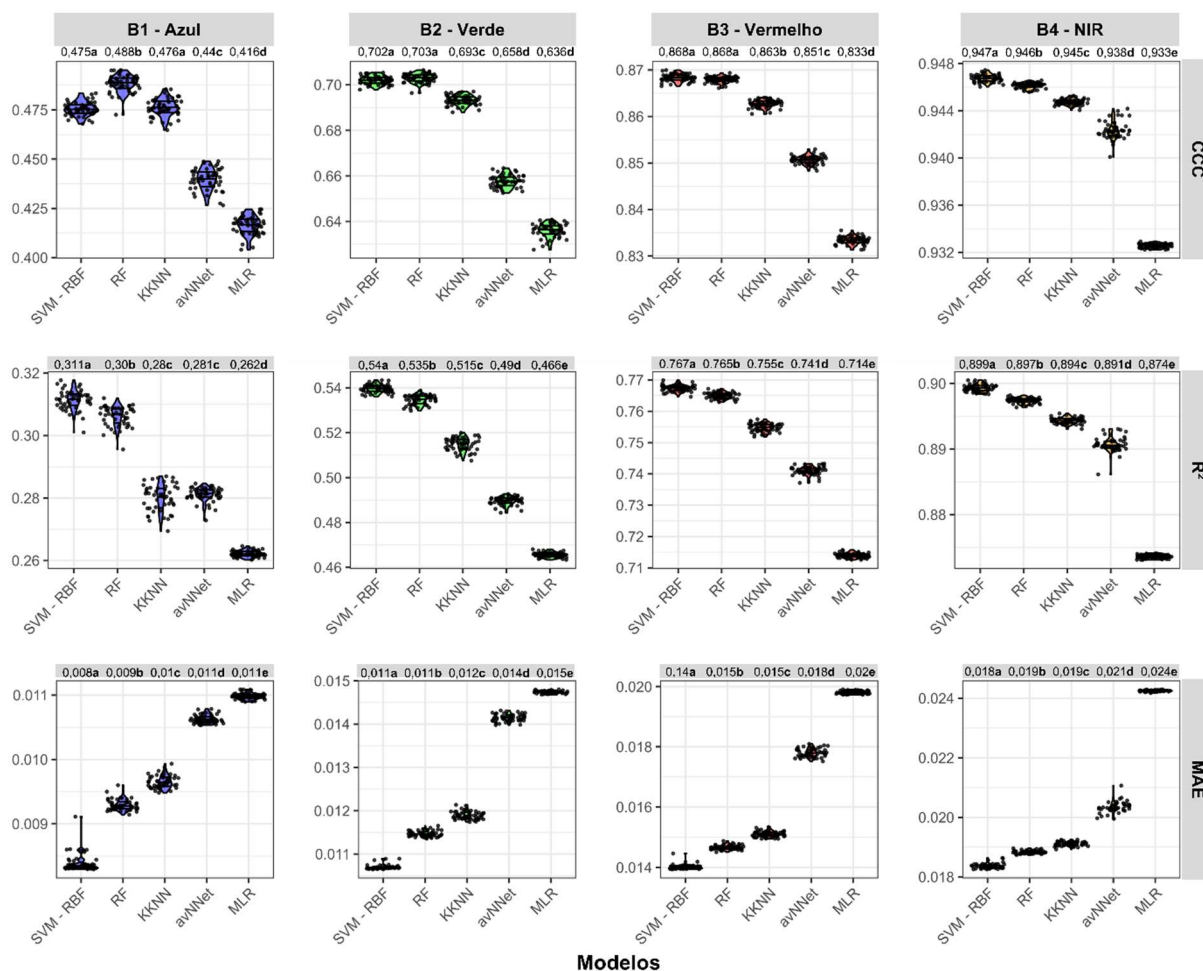


Figura 6: Violin plot demonstrando os resultados dos testes de Kruskal-Wallis e de Dunn aplicados aos índices estatísticos CCC e R^2 obtidos para os modelos de aprendizagem de máquina estudados.

Essa abordagem estatística robusta reforça a confiança nos resultados, indicando não apenas a existência de diferenças significativas entre os modelos, mas também destacando o SVM-RBF como a escolha mais eficiente para a normalização radiométrica dos dados provenientes do sensor PS. Esses resultados demonstram a robustez da metodologia adotada.

Na Figura 7, são apresentados os dados preditos e observados para os índices espectrais NDVI, NDWI, NDTI, SAVI e Iron, aplicados aos dados amostrados em regiões aquáticas e de solo/vegetação. Através dos gráficos e dos índices estatísticos

utilizados, foi possível estabelecer uma comparação entre os índices espectrais das bandas normalizadas e os valores observados dos índices gerados pelas bandas de referência S2.

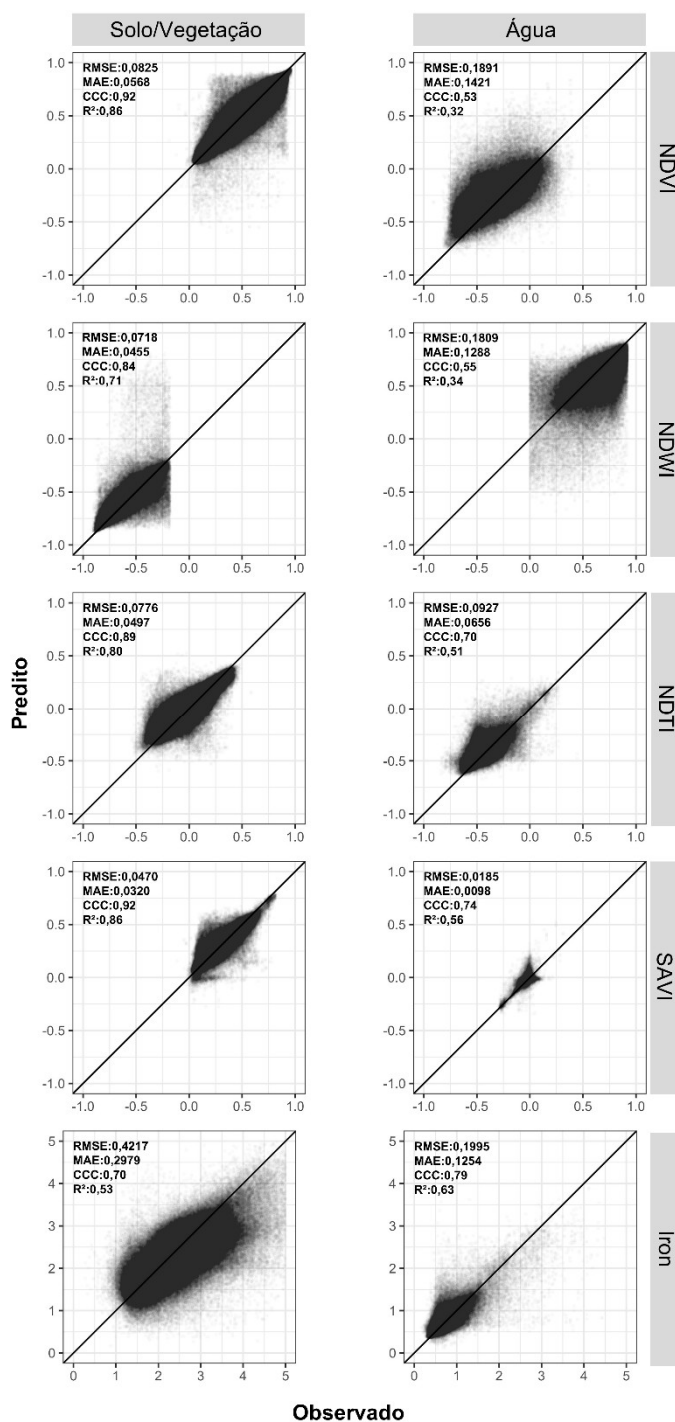


Figura 7: Dispersão dos dados preditos e observados e índices estatísticos utilizados para a avaliação das bandas normalizadas do sensor PS para os índices espectrais.

Na Figura 7, observa-se que os índices espectrais calculados a partir das bandas normalizadas apresentam uma tendência de correspondência com os índices gerados pelas bandas de referência S2, especialmente nas áreas de solo e vegetação. Contudo, nas regiões aquáticas, os índices espectrais demonstraram limitações significativas, atribuídas à elevada variabilidade inerente ao meio aquático. Essa variabilidade refletiu-se em valores de R^2 variando entre 0,32 e 0,63 e no CCC entre 0,53 e 0,79.

Nas áreas de solo e vegetação, a modelagem dos índices espectrais, utilizando bandas normalizadas, demonstrou boa correspondência com os dados observados. Os índices espectrais NDWI, NDVI, NDTI e SAVI apresentaram os melhores desempenhos, caracterizados por baixos valores de RMSE e MAE, além de R^2 variando de 0,71 a 0,86 e CCC entre 0,86 e 0,92. Em contraste, o índice Iron exibiu os piores resultados, com R^2 de 0,53 e CCC de 0,70, indicando limitações na sua capacidade preditiva.

Ao analisar os dados preditos e observados dos índices espectrais para as duas regiões, evidencia-se que, para uma avaliação abrangente do desempenho dos índices NDVI e NDWI, é essencial considerar a finalidade específica de sua aplicação. O NDVI, amplamente utilizado para monitorar a saúde da vegetação e estimar a biomassa em diversos ecossistemas (TUCKER, 1979), pode apresentar respostas distintas quando aplicado a ambientes aquáticos, dificultando sua interpretação e comparação nesses contextos (MATTHEWS, 2011). Por outro lado, o NDWI tem como principal objetivo a identificação e o monitoramento de corpos d'água em diferentes ambientes (GAO, 1996). No entanto, sua aplicação exclusiva a dados aquáticos, sem o contexto de dados terrestres, pode ser limitada, uma vez que a variabilidade espectral inerente ao meio aquático pode reduzir a precisão dos resultados (FISHER; FLOOD; DANAHER, 2016).

As diferentes abordagens empregadas evidenciam a versatilidade dos modelos desenvolvidos, com aplicações potenciais em estudos agrícolas, geológicos, geoquímicos, análise de parâmetros ambientais e monitoramento da qualidade da água. Esses modelos mostram-se particularmente úteis em situações que demandam imagens com alta resolução temporal e espacial, capazes de capturar fenômenos e processos efêmeros, como mudanças sazonais na cobertura vegetal, variações na qualidade da água e eventos climáticos extremos.

2.1.4 CONCLUSÕES

O procedimento metodológico adotado nesta pesquisa possibilitou a avaliação dos modelos de aprendizado de máquina e a identificação das covariáveis mais relevantes para a normalização das imagens do sensor PS. Além disso, permitiu determinar a contribuição de cada variável na predição do algoritmo, demonstrando ser aplicável a outras regiões do globo, desde que sejam respeitados os conjuntos de covariáveis e o intervalo dos dados.

A metodologia desenvolvida para a normalização radiométrica de imagens do sensor PlanetScope (PS), utilizando como referência imagens do sensor *Multispectral Instrument/Sentinel 2* (MSI/S2), possibilitou prever os dados de refletância com boa acurácia e precisão, evidenciando seu potencial para aplicação em situações que demandam imagens de alta resolução temporal e espacial, especialmente para capturar fenômenos e/ou processos efêmeros.

O modelo de aprendizado de máquina SVM-RBF (*Support Vector Machines* com *Kernel* de Função de Base Radial) apresentou os resultados mais promissores na normalização das imagens, destacando sua robustez diante de um banco de dados complexo. Isso evidencia a potencialidade do modelo SVM-RBF para a normalização radiométrica do sensor PS com um elevado grau de confiabilidade.

Recomenda-se que, em futuras pesquisas, seja realizada a avaliação de outros sensores (Landsat, MODIS) no processo de normalização, o que proporcionará uma compreensão mais abrangente da aplicabilidade da metodologia, contribuindo, dessa forma, para o processo de validação em diferentes contextos e condições. Além disso, deve-se realizar comparações entre o método desenvolvido e outras metodologias propostas na literatura, visando avaliar seu desempenho relativo em diferentes cenários.

REFERÊNCIAS

- AGUIRRE-GUTIÉRREZ, J. et al. Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. **Remote Sensing of Environment**, v. 252, n. May 2020, p. 112122, 2021.
- ALTMAN, D. G. **Practical statistics for medical research**. [s.l.] Chapman and Hall/CRC, 1990.
- ALVARES, C. A. et al. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2014.
- ANA. **Agência Nacional de Águas. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas**. Brasília - DF, 2017.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v. 4, p. 40–79, 2010.
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos**. 1º ed. São José dos Campos: INPE, 2019.
- BEVEN, K. J.; CLOKE, H. L. Comment on “hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al. **Water Resources Research**, v. 48, n. 1, p. 1–10, 2012.
- BEYCA, O. F. et al. Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. **Energy Economics**, v. 80, p. 937–949, 2019.
- BREIMAN, L. Random forests. **Machine learning**, v. 45.1, p. 27, 2001.
- CEMIG. **CEMIG Website**. , 2021. Disponível em: <<https://www.cemig.com.br>>. Acesso em: 22 set. 2021
- CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. **Geoscientific model development**, v. 7, n. 3, p. 1247–1250, 2014.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, v. 20.3, p. 25, 1995.
- CSILLIK, O. et al. Monitoring tropical forest carbon stocks and emissions using Planet satellite data. **Scientific Reports**, v. 9, n. 1, p. 1–12, 2019.
- DA SILVEIRA, V. A. et al. Modeling and mapping of Inselberg habitats for environmental conservation in the Atlantic Forest and Caatinga domains, Brazil. **Environmental Advances**, v. 8, n. October 2021, 2022.
- DE CARVALHO JÚNIOR, O. A. et al. Radiometric normalization of temporal images combining automatic detection of pseudo-invariant features from the distance and similarity spectral measures, density scatterplot analysis, and robust regression.

Remote Sensing, v. 5, n. 6, p. 2763–2794, 2013.

DIAS, R. L. S. et al. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatics**, n. August, p. 101414, 2021.

DOS SANTOS, E. P. et al. 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. February, 2022.

DU, Y. et al. Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. **Remote Sensing**, v. 8, n. 4, 2016.

DU, Y.; TEILLET, P. M.; CIHLAR, J. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. **Remote Sensing of Environment**, v. 82, n. 1, p. 123–134, 2002.

DUNN, O. J. Multiple comparisons using rank sums. **Technometrics**, v. 6, n. 3, p. 241–252, 1964.

EMBRAPA. **Sistema brasileiro de classificação de solos**. (EMBRAPA, Ed.)Rio de Janeiro, 2006.

FILGUEIRAS, R. et al. Crop NDVI monitoring based on sentinel 1. **Remote Sensing**, v. 11, n. 12, 2019.

FISHER, A.; FLOOD, N.; DANAHER, T. Comparing Landsat water index methods for automated water classification in eastern Australia. **Remote Sensing of Environment**, v. 175, p. 167–182, 2016.

FRANCINI, S. et al. Near-real time forest change detection using PlanetScope imagery. **European Journal of Remote Sensing**, v. 53, n. 1, p. 233–244, 2020.

FRAZIER, A. E.; HEMINGWAY, B. L. A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. **Remote Sensing 2021, Vol. 13, Page 3930**, v. 13, n. 19, p. 3930, 2021.

GAO, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote sensing of environment**, v. 58, n. 3, p. 257–266, 1996.

GAO, B. et al. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. **Remote sensing of environment**, v. 113, p. S17–S24, 2009.

GASCON, F. et al. Copernicus Sentinel-2A calibration and products validation status. **Remote Sensing**, v. 9, n. 6, 2017.

HELMAN, D. et al. Using time series of high-resolution planet satellite images to monitor grapevine stem water potential in commercial vineyards. **Remote Sensing**, v. 10, n. 10, p. 1–22, 2018.

HEWSON, R. D.; CUDAHY, T. J.; HUNTINGTON, J. F. **Geologic and alteration mapping at Mt Fitton, South Australia, using ASTER satellite-borne data**. IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217). **Anais...**2001.

HIJMANS, R. J. et al. Package 'terra'. **Maintainer: Vienna, Austria**, 2022.

HOUBORG, R.; MCCABE, M. F. High-Resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. **Remote Sensing**, v. 8, n. 9, 2016.

HOUBORG, R.; MCCABE, M. F. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. **Remote Sensing**, v. 10, n. 6, 2018a.

HOUBORG, R.; MCCABE, M. F. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. **Remote Sensing of Environment**, v. 209, n. February, p. 211–226, 2018b.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295–309, 1988.

IORDANIS, I.; KOUKOUVINOS, C.; SILOU, I. On the efficacy of conditioned and progressive Latin hypercube sampling in supervised machine learning. **Applied Numerical Mathematics**, n. December 2023, 2024.

KENNEDY, J. B.; NEVILLE, A. M. **Basic Statistical Methods for Engineers and Scientists**. Nova IorqueHarper and Row, , 1986.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, v. 47, n. 260, p. 583–621, 1952.

KUHN, M. **caret: Classification and Regression Training**. , 2020. Disponível em: <<https://cran.r-project.org/package=caret>>

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. [s.l.] Springer, 2013.

LACAUX, J. P. et al. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. **Remote Sensing of Environment**, v. 106, n. 1, p. 66–74, 2007.

LAI, W. et al. A portable algorithm to retrieve bottom depth of optically shallow waters from top-of-atmosphere measurements. **Journal of Remote Sensing**, 2022.

LAMQUIN, N. et al. An inter-comparison exercise of Sentinel-2 radiometric validations assessed by independent expert groups. **Remote Sensing of Environment**, v. 233, n. August, 2019.

LATTE, N.; LEJEUNE, P. PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. **Remote Sensing**, v. 12,

n. 15, 2020.

LEACH, N.; COOPS, N. C.; OBRKNEZEV, N. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. **Computers and Electronics in Agriculture**, v. 164, n. January, p. 104893, 2019.

LEE RODGERS, J.; NICEWANDER, W. A. Thirteen Ways to Look at the Correlation Coefficient. **The American Statistician**, v. 42, n. 1, p. 59–66, fev. 1988.

LI, W. et al. Generating red-edge images at 3M spatial resolution by fusing sentinel-2 and planet satellite products. **Remote Sensing**, v. 11, n. 12, p. 1–18, 2019.

LIN, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. **Biometrics**, v. 45, n. 1, p. 255, mar. 1989.

LOUIS, J. et al. Sentinel-2 SEN2COR: L2A processor for users. **European Space Agency**, v. SP-740, n. August, p. 9–13, 2016.

MARTINS, V. S. et al. Assessment of atmospheric correction methods for sentinel-2 MSI images applied to Amazon floodplain lakes. **Remote Sensing**, v. 9, n. 4, 2017.

MATTHEWS, M. W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. **International Journal of Remote Sensing**, v. 32, n. 21, p. 6855–6899, 2011.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. **Technometrics**, v. 42, n. 1, p. 55–61, 1979.

MCKIGHT, P. E.; NAJAB, J. Kruskal-wallis test. **The corsini encyclopedia of psychology**, p. 1, 2010.

MELLO, D. C. D. et al. A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. **Geoscientific Model Development**, v. 15, n. 3, p. 1219–1246, 2022.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, v. 32, n. 9, p. 1378–1388, 2006.

MORSE-MCNABB, E. M.; HASAN, M. F.; KARUNARATNE, S. A Multi-Variable Sentinel-2 Random Forest Machine Learning Model Approach to Predicting Perennial Ryegrass Biomass in Commercial Dairy Farms in Southeast Australia. **Remote Sensing**, v. 15, n. 11, 2023.

MÜLLER-WILM, U. Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual. **European Space Agency, (Special Publication) ESA SP**, v. 49, n. 2.2,

p. 1–51, 2016.

MUÑOZ-ROMERO, S. et al. Informative variable identifier: Expanding interpretability in feature selection. **Pattern Recognition**, v. 98, 2020.

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective. Chance Encounters: Probability in Education** London, England The MIT Press, , 2013.

PAHLEVAN, N. et al. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. **Remote Sensing of Environment**, v. 220, n. October 2018, p. 19–29, 2019.

PLANET TEAM. Planet Surface Reflectance Product v2. Planet Labs, Inc. Accessed 30.11.2020.

https://assets.planet.com/marketing/PDF/Planet_Surface_Reflectance_Technical_White_Paper.pdf. p. 1–18, 2020.

PLANET TEAM. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA, USA. **Planet Labs Inc**, p. 1–96, 2021.

POLITIS, D. N. et al. **Subsampling in the IID Case**. [s.l.] Springer, 1999.

R CORE TEAM. **R: A language and environment for statistical computing, version 3.3. 1. Vienna, Austria: R Foundation for Statistical Computing**. , 2020. Disponível em: <<https://www.r-project.org>>

REUNANEN, J. Overfitting in making comparisons between variable selection methods. **Journal of Machine Learning Research**, v. 3, p. 1371–1382, 2003.

RIPLEY, B. D. **Pattern recognition and neural networks**. [s.l.] Cambridge university press, 1996.

ROUSE, J. W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication**, v. 351, n. 1974, p. 309, 1974.

ROY, D. P. et al. Landsat-8: Science and product vision for terrestrial global change research. **Remote Sensing of Environment**, v. 145, p. 154–172, 2014.

SCHEFFLER, D. et al. AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data. **Remote Sensing**, v. 9, n. 7, 2017.

SCHOTT, J. R.; SALVAGGIO, C.; VOLCHOK, W. J. Radiometric Scene Normalization Using Pseudoinvariant Features. **Remote Sensing of Environment**, v. 16, p. 1–16, 1988.

SCHROEDER, T. A. et al. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. **Remote Sensing of Environment**, v. 103, n. 1, p. 16–26, 2006.

SENA, N. C. et al. Soil sampling strategy in areas of difficult access using the cLHS method. **Geoderma Regional**, v. 24, 2021.

SHIN, J. et al. High spatial-resolution red tide detection in the southern coast of Korea using U-Net from PlanetScope imagery. **Sensors**, v. 21, n. 13, 2021.

SONG, K. et al. Quantification of lake clarity in China using Landsat OLI imagery data. **Remote Sensing of Environment**, v. 243, p. 111800, 2020.

SYARIZ, M. A. et al. Spectral-consistent relative radiometric normalization for multitemporal Landsat 8 imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 147, n. November 2018, p. 56–64, 2019.

TUCKER, C. J. **Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. REMOTE SENSING OF ENVIRONMENT**, 1979.

VANHELLEMONT, Q.; RUDDICK, K. **Acolite for Sentinel-2: Aquatic applications of MSI imagery**. Proceedings of the 2016 ESA Living Planet Symposium, Prague, Czech Republic. **Anais...**2016.

WANG, H. et al. Technological Forecasting & Social Change Tension in big data using machine learning : Analysis and applications ☆. **Technological Forecasting & Social Change**, v. 158, n. May 2019, p. 120175, 2020.

WANG, J. et al. Automatic cloud and cloud shadow detection in tropical areas for PlanetScope satellite images. **Remote Sensing of Environment**, v. 264, n. May, p. 112604, 2021.

WEGMUELLER, S. A.; LEACH, N. R.; TOWNSEND, P. A. LOESS radiometric correction for contiguous scenes (LORACCS): Improving the consistency of radiometry in high-resolution satellite image mosaics. **International Journal of Applied Earth Observation and Geoinformation**, v. 97, p. 102290, 2021.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate research**, v. 30, n. 1, p. 79–82, 2005.

WU, S. et al. Monitoring tree-crown scale autumn leaf phenology in a temperate forest with an integration of PlanetScope and drone remote sensing observations. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 171, n. June 2020, p. 36–48, 2021.

XU, H. et al. A novel automatic method on pseudo-invariant features extraction for enhancing the relative radiometric normalization of high-resolution images. **International Journal of Remote Sensing**, v. 42, n. 16, p. 6155–6186, 2021.

YANG, X.; LO, C. P. Relative radiometric normalization performance for change detection from multi-date satellite images. **Photogrammetric Engineering and Remote Sensing**, v. 66, n. 8, p. 967–980, 2000.

YOUNG, S. L. et al. Perspective: The Importance of Water Security for Ensuring Food Security, Good Nutrition, and Well-being. **Advances in Nutrition**, v. 12, n. 4, p. 1058–1073, 2021.

ZHANG, L.; WU, C.; DU, B. Automatic Radiometric Normalization for Multitemporal Remote Sensing Imagery With Iterative Slow Feature Analysis. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 10, p. 6141–6155, 2014.

ZHANG, M. L.; ZHOU, Z. H. ML-KNN: A lazy learning approach to multi-label learning. **Pattern Recognition**, v. 40, n. 7, p. 2038–2048, 2007.

ZHANG, Y.; LI, A.; DENG, B. Data-driven predictive models for chemical durability of oxide glass under different chemical conditions. **npj Materials Degradation**, p. 1–11, 2020.

ZHU, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 370–384, 2017.

2.2 ARTIGO 2 -Predição de parâmetros da qualidade da água na bacia do rio Paraopeba com base em produtos de sensoriamento remoto e aprendizado de máquina

Resumo – O monitoramento da qualidade das águas superficiais (QA) é essencial para avaliar os recursos hídricos e identificar os seus padrões de qualidade. Os métodos tradicionais de monitoramento, baseados em estações convencionais de amostragens pontuais, embora confiáveis, são onerosos e apresentam limitações quanto à frequência e abrangência, dificultando a avaliação dos parâmetros de qualidade da água em escalas temporais e espaciais necessárias para identificar os efeitos de eventos extremos sobre a QA. Nesse contexto, o uso de imagens orbitais se apresenta como uma alternativa complementar viável para aprimorar a escala temporal e espacial do monitoramento da qualidade da água dos métodos tradicionais. No entanto, as limitações associadas às resoluções espectral, espacial, temporal e/ou radiométrica dessas imagens representam desafios expressivo para a precisão das predições. Assim, objetivou-se com este estudo propor uma metodologia para predição de parâmetros (ópticamente ativos e inativos) de qualidade da água em ambientes lóticos e lénticos utilizando dados de sensoriamento remoto e técnicas de aprendizado de máquina. Para o desenvolvimento do presente trabalho foram organizados e avaliados três bancos de dados de sensoriamento remoto: (i) dados obtidos a partir das imagens S2; (ii) dados obtidos a partir das imagens brutas PS; e (iii) dados obtidos a partir das imagens PS normalizadas utilizando a metodologia desenvolvida no capítulo 1. Os dados dos parâmetros de qualidade da água (QA) utilizados neste estudo foram obtidos de 24 estações de monitoramento localizadas ao longo da calha do rio Paraopeba e do reservatório de Três Marias, sendo utilizados a série de dados do período de 2016 a 2023. Para a predição dos parâmetros de QA foram utilizados quatro algoritmos de aprendizado de máquina: *Random Forest* (RF), *k-Nearest Neighbors* (KNN), *Support Vector Machines* com Kernel de Função de Base Radial (SVM-RBF) e *Cubist*. O desempenho dos modelos foi avaliado por meio dos índices estatísticos: raiz do erro quadrático médio – RMSE, erro médio absoluto – MAE, coeficiente de correlação de concordância de Lin – CCC e o coeficiente de determinação - R^2 . Os resultados evidenciaram que os modelos com dados normalizados do sensor PS apresentaram o melhor desempenho na estimativa dos

parâmetros de QA. Além disso, observou-se que os algoritmos baseados em árvore de decisão proporcionaram melhor generalização dos dados, superando os demais modelos utilizados. A metodologia proposta demonstrou-se adequada, validando não apenas a utilização dos dados do PS para esse tipo de análise, mas também fornecendo subsídios relevantes para sua aplicação em diferentes demandas ambientais.

Palavras-chave: Águas continentais. Eventos extremos. PlanetScope. Ambientes lênticos e lóticos.

2.2.1 INTRODUÇÃO

Os corpos d'água superficiais continentais, como são conhecidos os rios, lagos, reservatórios e riachos, são a principal fonte de água potável para o consumo humano, bem como para as necessidades recreativas, industriais, agricultura, produção de energia, transporte e pesca (OGASHAWARA; MISHRA; GITELSON, 2017; SABERIOON et al., 2020). Além disso, as águas superficiais continentais são indispensáveis para a manutenção do ciclo de vida no planeta, pois garantem a estabilidade da biodiversidade e do fluxo hidrológico (BARBOSA; NOVO; MARTINS, 2019). Diante disso, a avaliação e o monitoramento contínuos da quantidade e da qualidade das águas continentais é de fundamental importância para a conservação dos recursos hídricos, uma vez que permitem compreender os processos naturais e antrópicos que interferem nos sistemas aquáticos (ARANGO; NAIRN, 2020).

Convencionalmente, o monitoramento da qualidade das águas superficiais (QA) continentais é realizado por meio de amostragens pontuais em intervalos de tempo. Contudo, em casos de ocorrência de eventos extremos (atípicos) a rede de monitoramento básica pode não ser eficiente, pois a frequência de amostragem e a distribuição espacial das medições não permitem identificar a variabilidade espaço-temporal da QA com boa acurácia, muito menos as entradas das águas pluviais ao longo das margens dos lagos e reservatórios (GHOLIZADEH; ASSEFA; REDDI, 2016; WINSTON; DORSEY; HUNT, 2016). Portanto, o monitoramento especializado e contínuo mostra-se fundamental para complementar as abordagens convencionais, especialmente quando se deseja avaliar os impactos de eventos extremos, como rompimento de barragens, chuvas torrenciais, dentre outros.

Considerando a relação entre a refletância da água e sua qualidade, as técnicas de sensoriamento remoto e processamento de imagens têm se destacado como métodos alternativos eficazes para o monitoramento contínuo de parâmetros opticamente ativos de qualidade da água. Devido à sua cobertura espacial abrangente e temporalidade adequada, além do custo-benefício vantajoso, essas técnicas vêm sendo amplamente adotadas para a avaliação dos seguintes parâmetros de QA: clorofila-a (Chla), turbidez (T), matéria orgânica dissolvida colorida (CDOM), sólidos suspensos totais (SST) e profundidade do disco de Secchi (SDD) (ARANGO; NAIRN, 2020; BONANSEA et al., 2015; CUI et al., 2021; DIAS et al., 2021).

Por outro lado, ainda são escassos estudos que buscam estimar os parâmetros não opticamente ativos, como fósforo (P), nitrogênio (N), demanda química de oxigênio (DQO), oxigênio dissolvido (OD) e ferro (Fe) (SAGAN et al., 2020; TIAN et al., 2023). A principal dificuldade quando se trabalha com esses parâmetros não opticamente ativos é a sua baixa relação com as bandas espectrais, ou seja, são menos propensos a impactar as características dos espectros medidos pelos satélites (FERDOUS; RAHMAN, 2020; SWAIN; SAHOO, 2017; XIONG et al., 2020). Uma das alternativas para estimativa dos parâmetros de QA não opticamente ativos é explorar a sua alta correlação com parâmetros opticamente ativos. Além disso, deve-se investigar as relações com as variáveis ambientais combinadas com algoritmos de aprendizado de máquina para a estimativa indireta dos parâmetros não opticamente ativos (GAO et al., 2015; LI et al., 2017b; MATHEW; SRINIVASA RAO; MANDLA, 2017).

Essas limitações metodológicas somam-se às restrições inerentes às plataformas orbitais atuais como Landsat, MODIS e Sentinel-2, que apresentam resoluções espaciais de moderadas a baixas, variando de 10 a 1000 m. Associado a esse fator, as resoluções temporais dos satélites Landsat e Sentinel 2A e 2B são consideradas medianas, 10 e 16 dias, respectivamente (ESA, 2023; NASA, 2023). Assim, suas aplicações para detecção remota da QA em pequenos reservatórios, enseadas de grandes reservatórios e rios estreitos ou onde se necessita de imagens com maior frequência são limitadas (ANSPER; ALIKAS, 2018; NGUYEN; PHAM; DANG, 2019).

Considerando as limitações supracitadas, a Planet lançou uma grande quantidade de nanossatélites, conhecidos como “Dove”. Os Doves constituem uma constelação de nanossatélites com fator de forma CubeSat 3U (10 × 10 × 30 cm) e

capacidade de revisita diária a um mesmo alvo na superfície terrestre. Desde 2016 esses satélites fornecem imagens com resolução espacial de 3,7 m e espectral de 4 (bandas B1 - Azul [420 - 530 nm], B2 - Verde [500 - 590 nm], B3 - Vermelho [610 - 700 nm], e B4 – NIR [780 - 860 nm]) (PLANET TEAM, 2020). Além disso, tem um peso aproximado de apenas 4 kg e, por esse motivo, podem ser produzidos e lançados muito mais rápidos que os satélites tradicionais. Também não requerem um veículo de lançamento dedicado, podendo ser colocados em órbita como carga útil secundária (CRUSAN; GALICA, 2019).

Diante do exposto, objetivou-se com o presente trabalho propor uma metodologia para predição de parâmetros (opticamente ativos e inativos) de qualidade da água em ambientes lóticos e lênticos utilizando dados de sensoriamento remoto e técnicas de aprendizado de máquina.

2.2.2 MATERIAL E MÉTODOS

2.2.2.1 Área de estudo

Para aplicação e avaliação da metodologia proposta no presente trabalho foi utilizada uma área localizada na região central do estado de Minas Gerais (Figura 1), que contempla a bacia do rio Paraopeba e o reservatório de Três Marias. Segundo o IBGE (2021), o rio Paraopeba é um dos principais afluentes do rio São Francisco, com extensão de 510 km que abastece o reservatório de Três Marias após passar por 48 municípios mineiros.

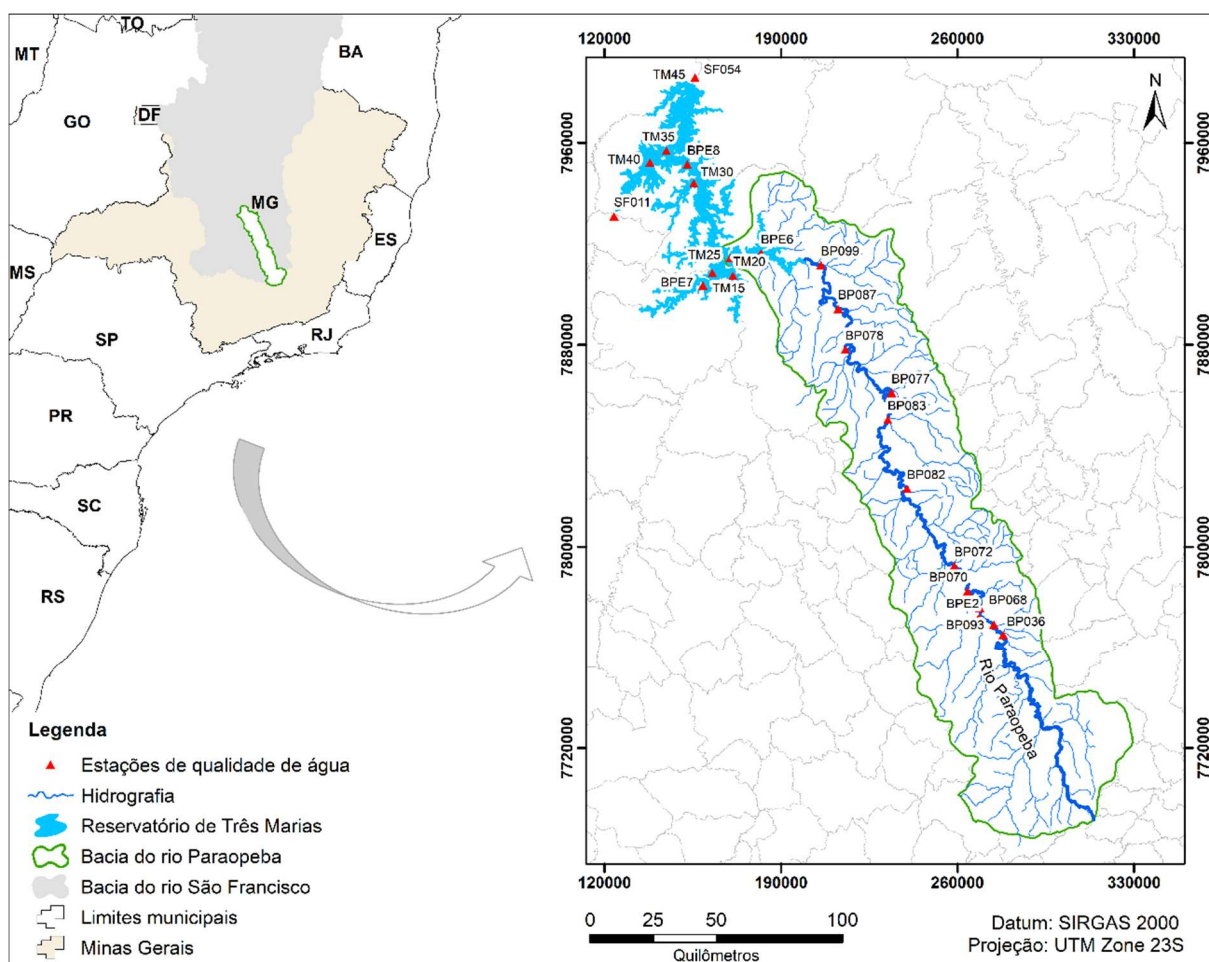


Figura 1: Localização da área de estudo.

A bacia do rio Paraopeba tem uma altitude média de 720 m, caracterizando-se por um relevo predominantemente forte ondulado e montanhoso (EMBRAPA, 1979). Na referida bacia as classes de solo predominantes são: Latossolos Vermelhos, Latossolos Vermelho-Amarelos, Cambissolos Háplicos e Cambissolos Húmicos. Segundo Alvares et al. (2013), o clima da região é classificado como Aw tropical, com

duas estações bem definidas: verão chuvoso de outubro a março; e inverno seco de abril a setembro.

A bacia do rio Paraopeba foi escolhida para o desenvolvimento deste trabalho por apresentar elevada diversidade pedológica, topográficas, cobertura vegetal e presença de corpos d'água continentais. Além disso, em 25 de janeiro de 2019 a bacia do rio Paraopeba sofreu com um dos maiores desastres socioambientais da história do Brasil que foi o rompimento da barragem de rejeitos (B-I) da mina do córrego do Feijão, localizada no município de Brumadinho. De acordo com o governo do estado de Minas Gerais (2024), esse rompimento ocasionou o despejo de aproximadamente 12 milhões de m³ de rejeitos, dos quais estima-se que cerca de 2 milhões de m³ permaneceram na área da antiga B-I; 7,8 milhões de m³ ficaram depositados na calha do ribeirão Ferro-Carvão até sua confluência com o rio Paraopeba e o restante (2,2 milhões de m³) atingiu a calha do rio Paraopeba.

2.2.2.2 Dados Hidrológicos

Para o desenvolvimento do trabalho foram utilizadas as séries de dados (2016 a 2023) dos parâmetros de QA de 24 estações de monitoramento distribuídas ao longo da calha do rio Paraopeba e no reservatório de Três Marias, incluindo a estação SF054, localizada a jusante da barragem (Figura 1 e Tabela 1). Os referidos dados contemplaram dos seguintes parâmetros de QA: turbidez, sólidos suspensos totais, clorofila-a, fósforo, nitrogênio, demanda química de oxigênio, oxigênio dissolvido e ferro.

Os dados das estações selecionadas foram obtidos do repositório institucional do Instituto Mineiro de Gestão das Águas (IGAM) (<http://repositorioigam.meioambiente.mg.gov.br>), órgão governamental responsável pelo monitoramento dos recursos hídricos no Estado de Minas Gerais. No monitoramento da rede básica, as campanhas de amostragem foram realizadas trimestralmente até dezembro de 2018. No entanto, após o rompimento da barragem em janeiro de 2019, as amostragens das estações localizadas na calha do rio Paraopeba passaram a ser mensais. Além disso, foram incorporados ao banco de dados os registros das estações de monitoramento situadas no reservatório de Três Marias, sob a responsabilidade da Companhia Energética de Minas Gerais (CEMIG).

Tabela 1: Descrição das estações de monitoramento das águas utilizadas no trabalho.

Estação	Município	Operadora	Longitude*	Latitude*
BP036	Brumadinho	Igam	591481,65	7766154,53
BP068	Mário Campos	Igam	582550,80	7777238,56
BP070	Betim	Igam	577842,67	7783718,41
BP072	Betim	Igam	571826,23	7794515,43
BP077	Papagaio	Igam	549063,79	7862981,60
BP078	Curvelo	Igam	530977,92	7880408,76
BP082	Esmeraldas	Igam	554515,30	7824940,83
BP083	Papagaio	Igam	549278,61	7858233,86
BP087	Curvelo	Igam	528445,63	7896560,69
BP093	Brumadinho	Igam	587971,75	7770652,15
BP099	Felixlândia	Igam	521656,61	7914060,68
BPE2	Brumadinho	Igam	582099,27	7773420,49
BPE6	Felixlândia	Igam	498228,23	7918958,61
BPE7	Abaeté	Igam	475069,83	7906716,13
BPE8	Três Marias	Igam	469656,80	7954599,84
SF011	Biquinhas	Igam	446188,74	7945567,26
SF054	Três Marias	Igam	473291,93	7988911,71
TM15	Abaeté	Cemig	478888,56	7911539,85
TM20	Pompéu	Cemig	487036,46	7910337,95
TM25	Pompéu	Cemig	486013,70	7917367,25
TM30	Morada Nova de Minas	Cemig	472184,16	7947127,11
TM35	Morada Nova de Minas	Cemig	461488,37	7960155,51
TM40	Morada Nova de Minas	Cemig	454887,44	7955627,32
TM45	Três Marias	Cemig	473237,48	7988849,00

*Datum Sirgas 2000 UTM 23S.

Fonte: (IGAM, 2021).

2.2.2.3 Aquisição e Processamento de Dados de Sensoriamento Remoto

Para a predição dos parâmetros de QA foram selecionadas imagens obtidas pelo sensor *Multispectral Instrument* (MSI), a bordo dos satélites Sentinel-2A (S2A) e Sentinel-2B (S2B), bem como pelo sensor PlanetScope, presente nas três diferentes gerações dos nanossatélites "Dove" da Planet (PS2, PS2.SD e PSB.SD). Adicionalmente, as imagens do sensor PS foram normalizadas, conforme metodologia proposta no capítulo 1. Na sequência serão apresentadas as especificações dos produtos, bem como a descrição dos processamentos aplicados aos mesmos.

a) **Sensor *Multispectral Instrument*/Sentinel-2**

No presente estudo foram utilizadas 10 bandas do sensor MSI/Sentinel-2, que estão acoplados aos satélites Sentinel-2A (S2A) e Sentinel-2B (S2B). Segundo Müller-Wilm (2016), as bandas multiespectrais estão distribuídas em diferentes regiões do espectro eletromagnético, sendo três bandas no visível (B2 - Azul [490 nm], B3 - Verde

[560 nm] e B4 - Vermelho [665 nm]); uma banda no infravermelho próximo (B8 - NIR [842 nm]); quatro bandas no *red edge* (B5 [705 nm], B6 [740 nm], B7 [783 nm] e B8a [865 nm]) e duas bandas no infravermelho de onda curta (B11 - SWIR [1610 nm] e B12 [2190 nm]). As bandas do visível e infravermelho próximo possuem resolução espacial de 10 m, enquanto as bandas do *red edge* e infravermelho de onda curta possuem resolução espacial de 20 m.

As cenas Sentinel-2 foram obtidas da Agência Espacial Europeia (ESA), disponíveis no site <https://scihub.copernicus.eu>. Foram selecionadas cenas do nível 1C sem a presença de nuvens (refletância do topo da atmosfera) e com tolerância de aquisição pelo sensor de dois dias em relação à coleta das amostras de água no período 2016 a 2023.

As etapas de pré-processamento utilizadas para as cenas do Sentinel-2 (S2A e S2B) foram as seguintes: **(i)** empilhamento das bandas por cenas e datas; **(ii)** mosaico das cenas de mesma data; **(iii)** correção atmosférica Acolite (VANHELLEMONT; RUDDICK, 2016) para remover os efeitos de atenuação causados por constituintes atmosféricos ativos, como dispersão molecular e de aerossol e absorção por vapor de água, ozônio, oxigênio e dióxido de carbono (GAO et al., 2009); **(iv)** cálculo dos índice espectrais (Tabela 2); **(v)** coleta e tabulação dos dados.

Tabela 2: Índices espectrais gerados através das imagens Sentinel 2

Índice	Equação	Autor
Aweish	$(B2 + (2.5 * B3)) - (1.5 * (B8 + B12)) - (0.25 * B12)$	(FEYISA et al., 2014)
Wri	$(B3+B4)/(B8+B12)$	(MUKHERJEE; SAMUEL, 2016)
SCI	$(B11- B8)/(B11+B8)$	(LI; LIU, 2008)
SAVI	$((B8- B4)/(B8+B4+0.5)) * (1.5)$	(HUETE, 1988)
FE_SI	$(B12/B11)$	(HENRICH et al., 2009)
FE_OX	$(B11/B8)$	(HENRICH et al., 2009)
FE2	$(B12/B8)+(B3/B4)$	(ROWAN; MARS, 2003)
FE3	$(B12/B8)+(B3/B4)$	(ROWAN; MARS, 2003)
NDTI	$(B11-B12)/(B11+B12)$	(DEVENTER et al., 1997)
NDMI	$(B8-B11)/(B8+B11)$	(ZHANG et al., 2016)
NDMI2	$(B4-B8A)/(B4+B8A)$	(WILSON; SADER, 2002)
Iron	$(B4/B2)$	(HEWSON; CUDAHY; HUNTINGTON, 2001)
Clay	$(B11/B12)$	(BOUSBIH et al., 2019)
Toming	$(B3/B4)$	(TOMING et al., 2016)
NDVI	$(B4-B8)/(B4+B8)$	(TUCKER, 1979)
NDWI	$(B3-B8)/(B3+B8)$	(MCFEETERS, 1996b)
IIA	$(4*B8)/(B3+(4*B8))$	(POLIDORIO; IMAI; TOMMASELLI, 2004)
IredEdge	$(B3+B4)/2$	(CLEVERS et al., 2002)
GLI	$(2*B3-B4-B2)/(2*B3+B4+B2)$	(COSTA; NUNES; AMPATZIDIS, 2020)
ARVI	$(B8-(2*B4-B2))/(B8+(2*B4)+B2)$	(KAUFMAN; TANRE, 1992)
NPQI	$(B2-B3)/(B2+B3)$	(BARNES et al., 1992)
GNDVI	$(B5-B2)/(B5+B2)$	(GITELSON; KAUFMAN; MERZLYAK, 1996)
NDVIVIS	$(B2-B3)/(B2+B3)$	(COSTA; NUNES; AMPATZIDIS, 2020)
AWEI	$4*(B3-B12)-(0.25*B8A+2.75*B11)$	(FEYISA et al., 2014)
ITBDN	$(B3-B2)/(B3+B2)$	(PIZANI; MAILLARD, 2023)

Onde: B corresponde à banda espectral utilizada da imagem Sentinel-2.

b) Sensor PlanetScope

As imagens PlanetScope (PS) são capturadas por pequenos nanossatélites construídos e operados pela empresa privada Planet. Os sensores PS são transportados por uma constelação de pequenos nanossatélites com fator de forma CubeSat 3U (10 × 10 × 30 cm). Atualmente, a Planet opera mais de 180 nanossatélites 'Doves', que fornecem imagens diárias da superfície terrestre com alta resolução espacial (3,7 m). As cenas são capturadas em bandas de comprimento de onda do B1 - Azul (420 - 530 nm), B2 - Verde (500 - 590 nm), B3 - Vermelho (610 - 700 nm), e B4 – NIR (780 - 860 nm). Segundo a Planet Team (2021), as imagens PS são

fornecidas ortorretificadas na projeção *Universal Transverse Mercator* (UTM) e corrigidas geometricamente, com precisão posicional de 10 m.

O download das imagens PS foi realizado através da Interface de Programação de Aplicativos (API) disponível em <https://www.planet.com/explorer>. Foram selecionadas cenas sem a presença de nuvens das três gerações de sensores disponíveis (PS2, PS2.SD e PSB.SD), que apresentavam data de aquisição coincidente com o dia da coleta das amostras de água no período de 2016 a 2023. Ressalta-se que as imagens do PS não são de livre acesso, no entanto, podem ser adquiridas sem custo mediante afiliação universitária e inscrição no programa *Planet Education and Research*, disponível em <https://go.planet.com/research>.

As etapas de pré-processamento das cenas PS foram as seguintes: (i) mosaico das faixas de aquisição; (ii) cálculo dos índices espectrais (Tabela 3); (iii) coleta e tabulação dos dados referentes aos pontos de monitoramento de qualidade da água.

Tabela 3: Índices espectrais gerados através das imagens PlanetScope

Índice	Equação	Autor
SAVI	$((B4-B3)/(B4+B3+0.5))^* (1.5)$	(HUETE, 1988)
NDTI_VIS	$(B3-B2)/(B3+B2)$	(LACAUX et al., 2007)
NDMI2	$(B3-B4)/(B3+B4)$	(ZHANG et al., 2016)
Iron	$(B3/B1)$	(WILSON; SADER, 2002)
Toming	$(B2/B3)$	(TOMING et al., 2016)
NDVI	$(B3-B4)/(B3+B4)$	(TUCKER, 1979)
NDWI	$(B2-B4)/(B2+B4)$	(MCFEETERS, 1996b)
HA	$(4*B4)/(B2+(4*B4))$	(POLIDORIO; IMAI; TOMMASELLI, 2004)
Irededge	$(B2+B3)/2$	(CLEVERS et al., 2002)
GLI	$(2*B2-B3-B1)/(2*B2+B3+B1)$	(COSTA; NUNES; AMPATZIDIS, 2020)
ARVI	$(B4-(2*B3-B1))/(B4+(2*B3)+B1)$	(KAUFMAN; TANRE, 1992)
NPQI	$(B1-B2)/(B1+B2)$	(BARNES et al., 1992)
GNDVI	$(B4-B1)/(B4+B1)$	(GITELSON; KAUFMAN; MERZLYAK, 1996)
NDVIVIS	$(B1-B2)/(B1+B2)$	(COSTA; NUNES; AMPATZIDIS, 2020)
ITBDN	$(B2-B1)/(B2+B1)$	(PIZANI; MAILLARD, 2023)

Onde: B corresponde à banda espectral utilizada da imagem PlanetScope.

c) Dados CHIRPS

Além dos dados das imagens PS e S2, foram também utilizados dados diários de precipitação estimados pelo projeto *Climate Hazards group Infrared Precipitation with Stations* (CHIRPS) (FUNK et al., 2015). O CHIRPS é um produto de reanálise que combina os dados de precipitação medidos em estações pluviométricas e dados de sensoriamento remoto orbital. O projeto CHIRPS disponibiliza dados de

precipitação em uma escala diária e global desde 1981, com uma resolução espacial de $0,05^\circ$ (≈ 5 km) (CHIRPS, 2024).

O processo de preparação dos dados de precipitação iniciou-se com a definição da área de contribuição a montante da estação selecionada. Em seguida, foi calculada a média dos valores dos pixels que se enquadravam nessa área. Posteriormente, realizou-se o download dos dados correspondentes ao dia da coleta das amostras de QA, além dos 14 dias anteriores, totalizando 15 dias de dados de precipitação.

O processo de preparação dos dados de precipitação iniciou-se com a definição da área de contribuição a montante da estação selecionada. Em seguida, foi calculada a média dos valores dos pixels que se enquadravam nessa área. Posteriormente, realizou-se o download dos dados correspondentes ao dia da coleta das amostras de QA, além dos 14 dias anteriores, totalizando 15 dias de dados de precipitação. Esses dados foram então acumulados em intervalos de 24, 48, 72 horas, até 360 horas. Por fim, as informações acumuladas foram adicionadas ao banco de dados como 15 variáveis independentes para a predição dos parâmetros de QA. Todas as etapas do pré-processamento das imagens e dados foram realizadas em linguagem de programação R (R CORE TEAM, 2020).

2.2.2.4 Aquisição dos valores de refletância

Após o pré-processamento das imagens, iniciou-se a etapa de refinamento e extração das amostras de refletâncias, pois a movimentação superficial da água pode gerar a reflexão da radiação solar direta (*sun glint*) e difusa (*sky glint*), o que interfere expressivamente na qualidade da resposta espectral das amostras (BARBOSA; NOVO; MARTINS, 2019; CARVALHO; CABÚS, 2020), que pode proporcionar superestimativa da refletância. Para minimizar a influência desta superestimativa na modelagem, adotou-se valores *NoData* para todos os pixels com refletância superior a 0,6 (DIAS et al., 2021).

Para a extração dos valores de refletância das imagens foram considerados dois cenários distintos: (i) No caso das imagens S2 foram extraídos os valores pontuais dos pixels e, para evitar interferências da margem do rio nos valores de refletância dos pixels com água, foram selecionadas apenas estações localizadas em seções de rio com uma largura mínima da calha de 30 metros (equivalente a três

pixels de S2); **(ii)** para as imagens PS foi utilizado uma janela 3x3, coletando a refletância média dessa área.

2.2.2.5 Modelagem dos parâmetros de QA por métodos de aprendizado de máquina

Com o objetivo de identificar o produto de sensoriamento remoto mais adequado para a predição dos parâmetros de QA, o processo de modelagem foi realizado considerando as seguintes bases de dados: **(i)** dados provenientes das imagens do Sentinel-2 (S2); **(ii)** dados das imagens do sensor PlanetScope (PS); e **(iii)** dados das imagens do sensor PS normalizadas, conforme metodologia proposta e descrita no Capítulo 1 desta tese.

Para a modelagem, a base de dados estruturada foi estratificada e, em seguida, subdividida aleatoriamente em dois conjuntos: 75% para treinamento e 25% para teste. A estratificação foi realizada para garantir a representatividade da amostra, seguindo dois critérios principais: **i)** proporção temporal: os conjuntos de treino e teste mantiveram a proporção de dados antes e depois do rompimento da barragem; **ii)** distribuição climática: os dados foram divididos proporcionalmente entre os períodos chuvoso e seco, respeitando a mesma proporção 75/25 dentro de cada período, com aleatorização aplicada apenas dentro de cada estrato.

Os resultados de desempenho dos modelos foram obtidos com base nos valores médios de 100 repetições realizadas nos conjuntos de treinamento e teste. Esse procedimento é eficaz para avaliar o desempenho do algoritmo, além de ajudar a identificar amostras problemáticas ou outliers nas coleções de dados (DIAS et al., 2021; MELLO et al., 2022; NEOGI; DAUWELS, 2022).

Na Figura 2 está apresentado o fluxograma das três etapas envolvidas na modelagem implementada: **(i)** identificação do conjunto ideal de covariáveis para cada algoritmo, removendo aquelas altamente correlacionadas ou com menor relevância para o treinamento; **(ii)** treinamento dos modelos utilizando o conjunto de variáveis selecionadas para cada algoritmo; e **(iii)** avaliação do desempenho dos modelos com um conjunto de dados distinto daquele utilizado no treinamento.

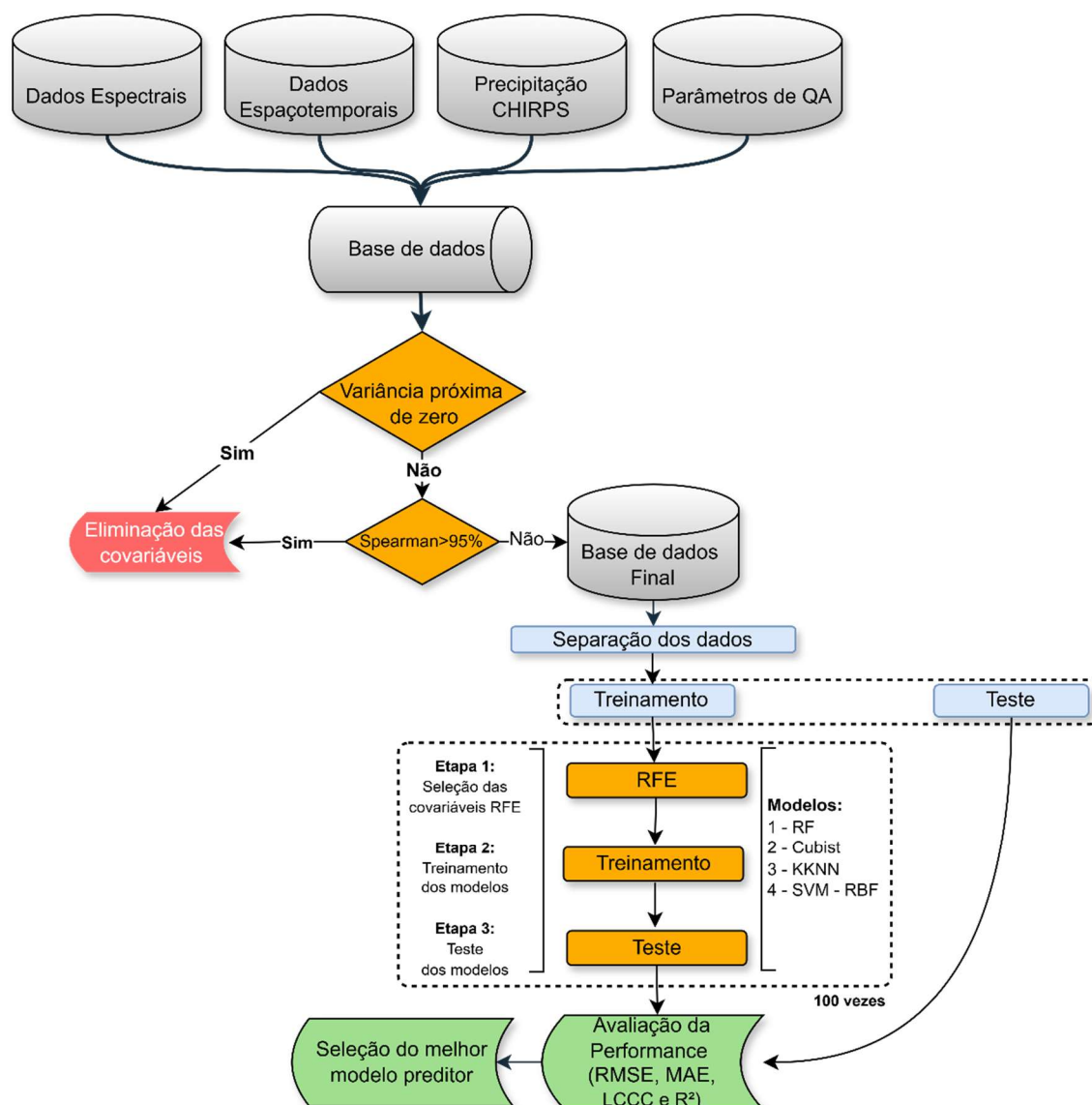


Figura 8: Fluxograma da metodologia desenvolvida para a predição dos parâmetros dos parâmetros de QA utilizando algoritmos de aprendizado de máquina e dados de sensoriamento remoto.

2.2.2.6 Seleção das covariáveis

A seleção de covariáveis é uma etapa do processo de modelagem que objetiva identificar o menor conjunto de covariáveis originais capaz de representar o fenômeno/processo modelado e minimizando a redundância, sendo usada para reduzir a dimensionalidade do espaço e remover covariáveis ruidosas, além de aumentar a parcimônia dos modelos (DIAS et al., 2021; MUÑOZ-ROMERO et al., 2020; REUNANEN, 2003).

Inicialmente, foi avaliada a variância das covariáveis, sendo eliminadas aquelas com valor nulo ou variabilidade muito baixa, com base nos critérios definidos na

metodologia proposta por (DA SILVEIRA et al., 2022). Essa avaliação foi realizada com a função “nearZeroVar”, do pacote *Caret* (KUHNS, 2020; MURPHY, 2013).

Em seguida, determinou-se o coeficiente de correlação de Spearman (LEE RODGERS; NICEWANDER, 1988). Para os pares de covariáveis com correlação maior ou igual a 95%, a variável com o maior valor absoluto de correlação com as demais variáveis foi removida.

Por fim, o algoritmo *Recursive Feature Elimination* (RFE) do pacote *Caret* foi implementado com o objetivo de eliminar as covariáveis que menos contribuíam para o modelo, com base em uma medida de importância para cada algoritmo (Kuhn e Johnson, 2013). Este algoritmo tem sido amplamente utilizado para o mapeamento de variáveis ambientais (DIAS et al., 2021; GHOSH; JOSHI, 2014; GOMES et al., 2019; MEYER et al., 2017, 2018; STEVENS et al., 2013).

Após a utilização do algoritmo RFE, obteve-se o conjunto de covariáveis ideal para cada algoritmo selecionado, as quais foram utilizadas nas etapas subsequentes da modelagem. Para o processamento da modelagem dos parâmetros de QA foram utilizados como preditores as seguintes covariáveis: as bandas individuais, razões de banda, índices espectrais (Tabela 2 e Tabela 3), dados de precipitação (acumulado de 2 a 15 dias anteriores a data da coleta) e dados sobre o período de aquisição da imagem (anterior ou posterior ao rompimento da barragem e período hidrológico). A lista das variáveis inseridas nos modelos está o Apêndice A.

2.2.2.6.1 Seleção dos modelos de aprendizado de máquina

Para a predição da concentração dos parâmetros de QA foram empregados os seguintes algoritmos: *Random Forest* – RF (BREIMAN, 2001), *Support Vector Machines* com *Kernel* de Função de Base Radial - SVM-RBF (CORTES; VAPNIK, 1995), *k-Nearest Neighbors* – KNN (ZHANG; ZHOU, 2007) e *Cubist* (ADAMS; STERLING, 1992). Estes algoritmos foram escolhidos por abranger os métodos de modelagem baseados em modelos linear, não linear, em regras, árvore de decisão e kernel (KUHNS; JOHNSON, 2013; MURPHY, 2013).

Os métodos RF e SVM-RBF já são amplamente empregados para a predição de parâmetros de QA (DU et al., 2018; SABERIOON et al., 2020; TIYASHA et al., 2021). O RF constrói uma coleção de N árvores de regressão que são aleatórias entre si, sendo considerada a predição média de todas as árvores para definir o valor final.

Por se tratar de um modelo de árvore, o RF é definido como um algoritmo não paramétrico (BREIMAN, 2001; FERREIRA et al., 2021).

Por outro lado, o SVM-RBF admite previsões com erro tolerável, controlado pelos vetores de suporte de modelagem e definido pelo hiperparâmetro C (cos) (CORTES; VAPNIK, 1995). O SVM-RBF, assim como o modelo RF, também é um modelo não paramétrico que se torna um método de regressão não linear ao utilizar uma função kernel não linear (MOUNTRAKIS; IM; OGOLE, 2011).

O algoritmo KKNn também é baseado na função Kernel, consistindo na identificação dos k pontos de treinamento mais próximos a uma nova amostra, usando uma métrica de distância, como a distância de Minkowski, que é uma forma geral das distâncias Euclidiana e de Manhattan (COVER; HART, 1967). Esse modelo não paramétrico atribui pesos aos vizinhos com base na distância, de modo que vizinhos mais próximos recebem maior peso. Isso evita ocorrência de suposições explícitas sobre as distribuições subjacentes dos dados (ABDULRAHMAN et al., 2020; DUDANI, 1976; SHEPARD, 1968).

O *Cubist* é um algoritmo de regressão baseado em árvores de decisão que utiliza uma abordagem híbrida, combinando a estrutura de uma árvore de decisão com modelos lineares ajustados em cada região da árvore (ADAMS; STERLING, 1992). O modelo constrói um conjunto de regras de decisão para dividir o espaço de atributos e, em seguida, aplica uma regressão linear local a cada região, permitindo uma modelagem mais flexível e interpretável das relações nos dados (HAFEEZ et al., 2019; HOUBORG; MCCABE, 2018c; NOI; DEGENER; KAPPAS, 2017).

A descrição mais detalhada dos algoritmos de aprendizado de máquina utilizado no presente estudo pode ser consultada nas publicações de Kuhn e Johnson (2013) e de Murphy (2013).

No treinamento, os hiperparâmetros internos de cada modelo foram ajustados utilizando validação cruzada repetida (*repeated cross-validation*), com 10 dobras (*folds*) e 10 repetições, aplicadas a cada hiperparâmetro de cada algoritmo (*tuneLength*). Os valores iniciais e as faixas de busca de cada parâmetro interno foram definidos de acordo com as recomendações dos desenvolvedores do pacote caret, conforme descrito no manual em seu capítulo 6, “Modelos descritos”, disponível em <https://topepo.github.io/caret/available-models.html>. Os valores dos hiperparâmetros finais, obtidos após a otimização de cada modelo, estão apresentados na Tabela 4.

Tabela 4: Hiperparâmetros otimizados para cada modelo desenvolvido.

Parâmetro	Cubist (committees/neighbors)	KKNN (kmax/distance/kernel)	RF (mtry)	SVM- RBF(sigma/C)
Planet				
chl _a	20 / 5	7 / 2 / optimal	10	0,02427 / 4
DQO	20 / 5	5 / 2 / optimal	6	18,5751 / 4
Fe	1 / 5	5 / 2 / optimal	2	0,50783 / 2
N	10 / 5	13 / 2 / optimal	2	0,04940 / 1
OD	1 / 9	5 / 2 / optimal	18	16,45009 / 4
PT	10 / 9	5 / 2 / optimal	60	0,00942 / 4
SST	10 / 9	7 / 2 / optimal	2	0,03363 / 2
T	10 / 5	13 / 2 / optimal	5	0,008999 / 2
Planet Norm				
chl _a	20 / 5	7 / 2 / optimal	10	0,02427 / 4
DQO	20 / 5	5 / 2 / optimal	6	18,5853 / 4
Fe	1 / 5	5 / 2 / optimal	2	0,50783 / 2
N	10 / 5	13 / 2 / optimal	2	0,04940 / 1
OD	1 / 9	5 / 2 / optimal	18	16,50009 / 4
PT	10 / 9	5 / 2 / optimal	60	0,00942 / 4
SST	10 / 9	7 / 2 / optimal	2	0,03363 / 2
T	10 / 5	13 / 2 / optimal	5	0,008999 / 2
Sentinel 2				
chl _a	1 / 5	5 / 2 / optimal	2	0,02736 / 4
DQO	1 / 9	7 / 2 / optimal	6	0,03702 / 4
Fe	20 / 5	5 / 2 / optimal	2	0,10574 / 4
N	20 / 9	7 / 2 / optimal	18	0,05874 / 4
OD	20 / 5	5 / 2 / optimal	5	0,05790 / 4
PT	20 / 5	11 / 2 / optimal	11	0,11027 / 4
SST	10 / 9	5 / 2 / optimal	2	0,01212 / 2
T	20 / 9	5 / 2 / optimal	15	0,00732 / 4

2.2.2.6.2 Avaliação dos modelos

Para avaliar o desempenho dos modelos ajustados, os valores estimados pelos modelos foram comparados com os valores observados nas estações de monitoramento de QA existentes na área de estudo (Tabela 1), por meio das seguintes métricas estatísticas: raiz do erro quadrático médio – RMSE (Equação 1), erro médio absoluto – MAE (Equação 2), coeficiente de correlação de concordância de Lin – CCC (Equação 3) e coeficiente de determinação - R^2 (Equação 4). Esse conjunto de métricas foi aplicado visando englobar diferentes características na avaliação de desempenho dos modelos ajustados (ARLOT; CELISSE, 2010; KENNEDY; NEVILLE, 1986; LIN, 1989).

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad (2)$$

$$CCC = \frac{2rV_rV_R}{(\bar{P} - \bar{O})^2 + V_r + V_R} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O})^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

Onde: P_i representa os valores preditos pelo modelo; O_i representa os valores observados; $\bar{O}$ representa a média dos valores observados; $\bar{P}$ representa a média dos valores preditos; V_r e V_R são as variações dos valores preditos e estimados, respectivamente e n representa o número de pares de observações.

O RMSE eleva ao quadrado a diferença entre os valores preditos e observados, penalizando erros extremos mais do que erros pequenos, tornando-o um bom indicador de valores discrepantes nos dados preditos (CHAI; DRAXLER, 2014). O MAE mede a média da diferença entre o valor observado com o predito. Como há valores positivos e negativos, é utilizado o módulo da diferença (WILLMOTT; MATSUURA, 2005). O CCC mede a proximidade da linha de melhor ajuste do modelo a uma reta de 45 graus (1:1) partindo da origem (LIN, 1989). Já o R^2 representa o percentual da variância dos dados que é explicado pelo modelo desenvolvido (MORSE-MCNABB; HASAN; KARUNARATNE, 2023).

As métricas RMSE e MAE fornecem resultados na mesma unidade da variável de interesse, o que facilita a interpretação dos erros. Para a análise dos resultados, modelos com menores valores de RMSE e MAE foram considerados mais precisos (BRUCE; BRUCE, 2019). Segundo Altman (1990), o CCC e o R^2 podem ser considerados moderados para valores entre 0,5 e 0,7, fortes para valores entre 0,7 e 0,9 e muito fortes para valores acima de 0,9.

Além de obter as métricas estatísticas de avaliação dos modelos obtidos pelos algoritmos de aprendizado de máquinas, calculou-se também os valores do RMSE (Equação 5) e do MAE (Equação 6) para o modelo “nulo”. O modelo nulo (NULL)

considera a predição dos parâmetros dado pelo valor médio das amostras coletadas. Ele representa a abordagem mais simples, retornando uma única média para os resultados numéricos com base no conjunto de treinamento.

Desta forma, modelos com desempenho semelhante ou pior em relação ao modelo nulo são mal avaliados. A escolha do modelo para a predição de um determinado parâmetro deve ocorrer quando os valores de RMSE e MAE forem inferiores aos observados para o modelo “nulo”, indicando que o modelo é superior à simples utilização dos valores médios, evidenciando ganho na utilização do modelo de aprendizado de máquina (MELLO et al., 2022).

$$NULL_RMSE = \left[\frac{1}{n} \sum_{i=1}^n (\bar{O}_{mt} - O_{iT})^2 \right]^{\frac{1}{2}} \quad (5)$$

$$NULL_MAE = \frac{\sum_{i=1}^n |\bar{O}_{mt} - O_{iT}|}{n} \quad (6)$$

Onde: $\bar{O}_{mt}$ representa a média das amostras de treinamento; O_{iT} representa os valores das amostras teste e n representa o número de amostras (loop).

2.2.3 RESULTADOS E DISCUSSÃO

Na Tabela 5 são apresentadas as estatísticas descritivas dos parâmetros de QA utilizados no presente estudo, evidenciando grande discrepância entre os valores mínimos e máximos, ou seja, uma elevada amplitude dos valores de QA observados. Além disso, observa-se que, em geral, as médias dos valores dos parâmetros são mais próximas dos valores mínimos, indicando uma certa assimetria na distribuição dos dados, com uma cauda longa em direção aos valores mais altos. Essa assimetria é frequentemente observada em dados ambientais, onde as medições tendem a se concentrar em valores baixos a moderados, mas eventos raros e extremos puxam a cauda da distribuição para cima (WILKS, 2011).

Tabela 5: Estatística descritiva dos parâmetros de QA para os dois bancos de dados utilizados na modelagem do estudo

Sensor	Estatísticas	Parâmetros							
		Chla	DQO	Fe	P	N	OD	SST	T
MSI/Sentinel-2	Máximo	8,460	57,400	5,241	0,40	1,240	14,800	528,00	809,00
	Mínimo	0,270	5,000	0,011	0,010	0,100	2,100	2,00	0,50
	Intervalo	8,190	52,4	5,230	0,390	1,140	12,700	526,00	808,50
	Média	1,932	17,661	0,317	0,072	0,406	7,481	31,33	34,08
	Mediana	1,340	17,000	0,174	0,050	0,380	7,700	12,00	11,20
	Desv. Padrão	3,718	9,090	0,544	0,080	0,238	1,445	58,20	77,18
PlanetScope	Máximo	8,69	59,000	21,846	0,480	1,310	14,800	716,00	982,00
	Mínimo	0,800	5,000	0,023	0,010	0,100	2,100	2,00	0,53
	Intervalo	7,89	54,000	21,823	0,470	1,210	12,700	714,00	981,47
	Média	2,202	20,777	0,677	0,098	0,431	7,481	78,83	87,93
	Mediana	1,600	20,000	0,253	0,060	0,380	7,500	29,00	30,45
	Desv. Padão	2,851	10,818	1,937	0,092	0,258	1,300	136,32	144,53

T – Turbidez, SST – Sólidos Suspensos Totais, Fe – Ferro, P – Fósforo, OD – Oxigênio Dissolvido, DQO – Demanda Bioquímica de Oxigênio, N – Nitrogênio e Chla – Clorofila-a.

Na Figura 3 são ranqueadas as covariáveis mais importantes para os modelos avaliados para a predição dos parâmetros de QA, considerando os três bancos de dados analisados (Sentinel-2, PlanetScope e PlanetScope Normalizada). A análise demonstrou que, por meio de processos de seleção de covariáveis, foi possível alcançar uma redução expressiva no número de covariáveis preditoras. Esse processo possibilitou reduzir de 151 covariáveis inicialmente selecionadas para dez variáveis para cada modelo, consideradas adequadas e ideais para a modelagem dos parâmetros de QA. Esses resultados corroboram com aqueles obtidos por Muñoz-Romero et al. (2020) e Stevens et al (2013), os quais demonstraram que a redução

da complexidade dos modelos, além de contribuir para a diminuição de custos computacionais, melhora a robustez e desempenho preditivo dos mesmos.

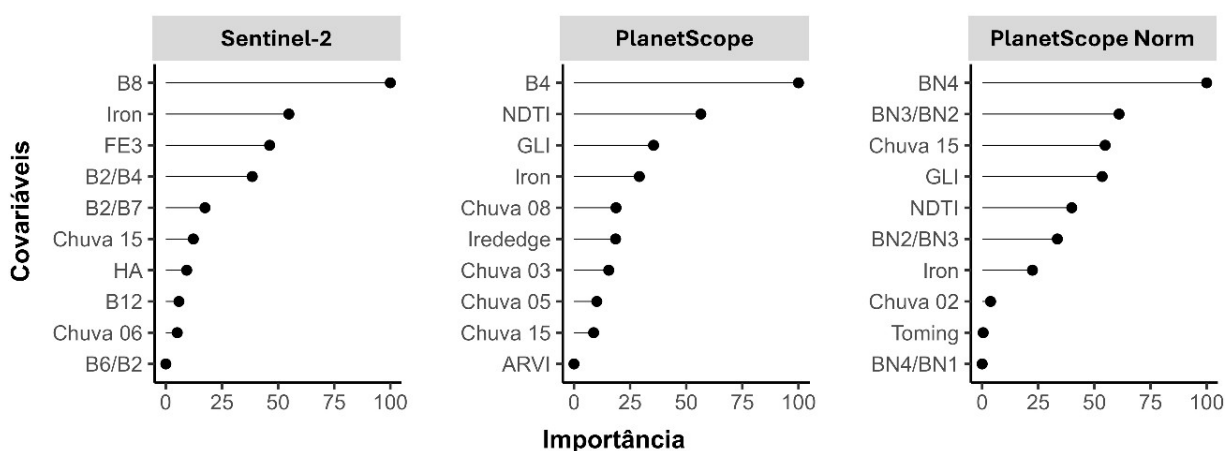


Figura 3: Ranking das covariáveis mais importantes para a modelagem dos parâmetros de QA nas três bases de dados analisadas.

Nota: As covariáveis listadas referem-se a bandas espectrais (ex: B4, BN4), razões entre bandas (ex: B2/B4, BN3/BN2), índices espectrais (ex: GLI, NDTI, ARVI), e variáveis ambientais como a precipitação acumulada nos dias anteriores à data da imagem (ex: Chuva 15 = chuva acumulada em 15 dias).

Observa-se na Figura 3 que, com exceção das bandas do comprimento de onda NIR (B8 para S2 e B4 para o PS), os índices e razões de bandas destacam-se como as covariáveis mais importantes para a predição dos parâmetros de QA em comparação com as bandas espectrais individuais. Esse resultado corrobora com aqueles obtidos por Sestini (1999) e Lillesand, Kiefer e Chipman (2015), que demonstraram que a combinação de bandas espectrais por meio de índices e razões entre bandas permitem discriminar diferenças sutis no comportamento espectral de diferentes alvos, enquanto as bandas espectrais individuais são capazes de identificar apenas variações mais evidentes e, desta forma, os índices e razões de bandas apresentam maior potencial para identificar características espectrais específicas de objetos ou fenômenos naturais.

A banda NIR e o índice Iron foram as covariáveis mais relevantes entre as demais, sendo selecionados com destaque na modelagem dos três bancos de dados avaliados. Além disso, os dados de precipitação estimados pelo produto CHIRPS foram selecionados em todas as modelagens avaliadas. Isso sugere que, com o aumento da precipitação e das vazões, ocorre um maior transporte de sedimento para o canal fluvial. Adicionalmente, os processos hidrodinâmicos promovem a ressuspensão dos materiais depositados no leito do canal, interferindo diretamente

nas características espectrais do corpo d'água, facilitando a predição dos parâmetros de interesse.

Na Tabela 6 são apresentados, para os dados do sensor MSI/S2, os resultados das métricas estatísticas utilizadas para avaliar o desempenho dos modelos de aprendizado de máquina aplicados à predição de parâmetros da QA na bacia do rio Paraopeba.

Tabela 6: Valores das métricas estatísticas utilizadas para avaliação de desempenho dos modelos *Cubist*, KNN, RF e SVM-RBF para a predição de parâmetros de qualidade da água (T – Turbidez, SST – Sólidos Suspensos Totais, Fe – Ferro, P – Fósforo, OD – Oxigênio Dissolvido, DQO – Demanda Bioquímica de Oxigênio, N – Nitrogênio e Chla – Clorofila-a) considerando os dados Sentinel-2.

Treinamento					Teste						
Modelos	RMSE	MAE	R ²	CCC	RMSE	MAE	R ²	CCC	NULL RMSE	NULL MAE	
T	Cubist	35,369	14,468	0,787	0,823	42,738	15,210	0,739	0,813	74,451	37,619
	KKNN	36,486	14,856	0,769	0,795	41,701	15,016	0,735	0,799		
	RF	33,953	14,322	0,785	0,824	39,176	14,625	0,750	0,824		
	SVM-RBF	39,347	16,527	0,746	0,761	48,495	17,575	0,656	0,709		
SST	Cubist	36,185	17,066	0,648	0,703	40,720	17,572	0,560	0,701	57,012	29,762
	KKNN	33,785	16,411	0,658	0,715	37,482	16,527	0,596	0,727		
	RF	34,250	16,800	0,643	0,709	37,878	16,881	0,585	0,719		
	SVM-RBF	34,652	16,899	0,640	0,674	40,932	17,780	0,533	0,630		
Fe	Cubist	0,422	0,208	0,369	0,490	0,474	0,213	0,278	0,449	0,519	0,269
	KKNN	0,438	0,215	0,338	0,466	0,482	0,219	0,242	0,418		
	RF	0,453	0,221	0,299	0,417	0,496	0,225	0,202	0,371		
	SVM-RBF	0,420	0,201	0,341	0,429	0,465	0,201	0,241	0,342		
P	Cubist	0,067	0,042	0,308	0,474	0,071	0,043	0,230	0,420	0,077	0,055
	KKNN	0,068	0,043	0,300	0,471	0,072	0,045	0,210	0,408		
	RF	0,065	0,042	0,325	0,478	0,069	0,043	0,249	0,427		
	SVM-RBF	0,067	0,040	0,299	0,420	0,070	0,040	0,228	0,371		
OD	Cubist	1,320	0,928	0,216	0,380	1,395	0,942	0,175	0,360	1,457	1,006
	KKNN	1,328	0,927	0,215	0,378	1,467	0,985	0,133	0,318		
	RF	1,218	0,836	0,288	0,419	1,325	0,864	0,203	0,361		
	SVM-RBF	1,267	0,857	0,232	0,377	1,382	0,899	0,152	0,316		
DQO	Cubist	8,438	6,634	0,192	0,351	8,641	6,746	0,139	0,313	8,978	7,471
	KKNN	8,869	7,067	0,146	0,295	9,124	7,200	0,093	0,252		
	RF	7,969	6,340	0,247	0,381	8,026	6,366	0,212	0,369		
	SVM-RBF	8,566	6,579	0,165	0,317	8,850	6,799	0,106	0,267		
N	Cubist	0,226	0,180	0,172	0,322	0,234	0,184	0,119	0,288	0,237	0,191
	KKNN	0,232	0,185	0,134	0,273	0,241	0,191	0,080	0,227		
	RF	0,213	0,171	0,220	0,352	0,221	0,176	0,159	0,313		
	SVM-RBF	0,231	0,183	0,137	0,280	0,241	0,190	0,082	0,234		
Chla	Cubist	3,406	1,520	0,106	0,156	4,231	1,548	0,047	0,069	3,031	1,364
	KKNN	2,989	1,390	0,149	0,183	3,524	1,397	0,075	0,113		
	RF	3,026	1,468	0,126	0,157	3,501	1,457	0,044	0,077		
	SVM-RBF	2,546	1,247	0,162	0,238	3,079	1,252	0,055	0,123		

Onde: RMSE - raiz do erro quadrático médio, MAE - erro médio absoluto, CCC - coeficiente de correlação de concordância de Lin, R² - coeficiente de determinação e NULL RMSE – raiz do erro quadrático médio do modelo nulo e NULL MAE - erro médio absoluto do modelo nulo.

Nota: As linhas em negrito representam os modelos com melhor desempenho para cada parâmetro.

A Tabela 6 evidencia a maior robustez dos modelos baseados em árvores, especificamente o RF e o *Cubist*, quando comparados aos modelos KNN e SVM-RBF. O RF foi o melhor modelo em cinco dos oito parâmetros avaliados, enquanto o *Cubist* se destacou como pelo menos o segundo modelo em seis dos oito parâmetros avaliados. Ambos se destacaram pelos menores valores de erro (RMSE e MAE) e pelos maiores valores de R^2 e CCC. Essa superioridade dos modelos baseados em árvores reflete a sua capacidade de capturar relações não lineares e interações complexas entre as variáveis (BREIMAN, 2001).

Ao analisar os parâmetros de qualidade de água de forma individual, observa-se que os conjuntos de dados que apresentaram melhor generalização dos algoritmos foram a T e o SST, com valores de CCC variando 0,82 e 0,72 e R^2 entre 0,75 e 0,59, respectivamente, e com baixos valores de RMSE e MAE. Em contrapartida, para os parâmetros de Fe, P, OD, DQO e N, os valores de CCC variaram de 0,44 a 0,31, enquanto os valores de R^2 oscilaram entre 0,27 e 0,15, indicando uma capacidade de predição bem limitada dos modelos. Quanto ao parâmetro Chla, os algoritmos enfrentaram grandes dificuldades em generalizar os dados, sendo que o algoritmo que apresentou o melhor desempenho, SVM-RBF (CCC = 0,12 e R^2 = 0,05), possui um RMSE pior do que aquele observado para o modelo “Nulo” na etapa de teste do modelo.

De forma geral, os modelos não apresentaram *overfitting*, o que pode ser comprovado pela proximidade e concordância entre os dados de treinamento e teste. Ao avaliar os valores de RMSE e MAE, observa-se que, com exceção dos modelos para o parâmetro clorofila-a (Chla), todos os outros demonstraram um ganho em comparação aos valores de NULL RMSE e NULL MAE. Os ganhos variaram de 47,38%, para o RMSE no parâmetro T, até 6,75% para o RMSE no parâmetro N. No que diz respeito ao MAE, os valores dos ganhos variaram de 61,12%, para o MAE de T, até 7,85% para o MAE de N. No caso do parâmetro Chla não foi observada vantagem na utilização do modelo desenvolvido em relação ao modelo nulo para RMSE, no entanto, para o MAE houve um ganho de 8,21% em comparação ao modelo nulo.

Na Tabela 7 são apresentados, para os dados do sensor PS, os resultados das métricas estatísticas utilizadas para a avaliação do desempenho dos modelos de

aprendizado de máquina aplicados à predição de parâmetros da QA na bacia do rio Paraopeba.

Tabela 7: Valores das métricas estatísticas utilizadas para a avaliação de desempenho dos modelos *Cubist*, KNN, RF e SVM-RBF para a predição de parâmetros de QA (T – Turbidez, SST – Sólidos Suspensos Totais, Fe – Ferro, P – Fósforo, OD – Oxigênio Dissolvido, DQO – Demanda Bioquímica de Oxigênio, N – Nitrogênio e Chla – Clorofila-a) considerando os dados do sensor PlanetScope.

	Modelos	Treinamento				Teste				NULL RMSE	NULL MAE
		RMSE	MAE	R ²	CCC	RMSE	MAE	R ²	CCC		
T	Cubist	65,317	33,266	0,825	0,878	66,664	32,608	0,796	0,878	140,817	96,021
	KKNN	67,917	34,881	0,803	0,861	76,514	36,639	0,726	0,835		
	RF	65,770	33,961	0,823	0,874	67,549	33,323	0,790	0,873		
	SVM-RBF	66,982	35,895	0,815	0,855	74,055	36,702	0,747	0,830		
SST	<i>Cubist</i>	76,373	38,527	0,734	0,804	88,274	40,492	0,648	0,768	136,236	86,864
	KKNN	74,968	38,483	0,724	0,798	87,313	41,017	0,627	0,758		
	RF	74,350	38,863	0,737	0,806	83,945	40,388	0,659	0,774		
	SVM-RBF	73,039	38,875	0,737	0,791	84,009	40,907	0,657	0,750		
Fe	Cubist	1,126	0,465	0,549	0,605	1,282	0,456	0,591	0,672	1,866	0,792
	KKNN	1,224	0,446	0,557	0,605	1,367	0,441	0,558	0,671		
	RF	1,408	0,615	0,369	0,421	1,642	0,629	0,315	0,408		
	SVM-RBF	1,172	0,507	0,549	0,573	1,449	0,534	0,505	0,541		
P	<i>Cubist</i>	0,077	0,053	0,342	0,513	0,078	0,053	0,303	0,505	0,091	0,071
	KKNN	0,078	0,053	0,324	0,498	0,079	0,053	0,283	0,487		
	RF	0,075	0,053	0,357	0,494	0,075	0,052	0,327	0,494		
	SVM-RBF	0,075	0,049	0,362	0,510	0,076	0,048	0,337	0,513		
OD	Cubist	1,130	0,788	0,252	0,367	1,139	0,781	0,294	0,476	1,304	0,852
	KKNN	1,159	0,798	0,257	0,385	1,198	0,807	0,270	0,488		
	RF	1,117	0,751	0,249	0,337	1,145	0,750	0,246	0,373		
	SVM-RBF	1,139	0,789	0,231	0,347	1,176	0,805	0,241	0,422		
DQO	<i>Cubist</i>	10,324	8,028	0,162	0,303	10,601	8,163	0,127	0,296	10,838	8,487
	KKNN	10,281	7,989	0,183	0,344	10,689	8,191	0,135	0,320		
	RF	9,378	7,403	0,263	0,400	9,681	7,564	0,222	0,389		
	SVM-RBF	9,904	7,699	0,209	0,360	10,397	8,013	0,146	0,322		
N	<i>Cubist</i>	0,226	0,174	0,265	0,417	0,236	0,179	0,202	0,386	0,257	0,202
	KKNN	0,236	0,181	0,223	0,391	0,248	0,189	0,157	0,350		
	RF	0,221	0,172	0,279	0,409	0,227	0,174	0,233	0,391		
	SVM-RBF	0,223	0,171	0,277	0,432	0,235	0,179	0,211	0,394		
Chla	<i>Cubist</i>	2,705	1,448	0,145	0,242	3,010	1,425	0,079	0,179	2,650	1,485
	KKNN	2,758	1,429	0,167	0,267	2,911	1,377	0,124	0,257		
	RF	2,392	1,360	0,159	0,270	2,711	1,361	0,071	0,183		
	SVM-RBF	2,235	1,192	0,217	0,337	2,577	1,212	0,117	0,235		

Onde: RMSE - raiz do erro quadrático médio, MAE - erro médio absoluto, CCC - coeficiente de correlação de concordância de Lin, R² - coeficiente de determinação e NULL RMSE – raiz do erro quadrático médio do modelo nulo e NULL MAE - erro médio absoluto do modelo nulo.

Nota: As linhas em negrito representam os modelos com melhor desempenho para cada parâmetro.

Observa-se na Tabela 7 que os modelos baseados em árvore apresentam destaque, com os algoritmos *Cubist* e RF dentre os dois melhores modelos em todos os oito parâmetros avaliados para os dados do sensor PS. Os parâmetros T, SST, Fe e P apresentaram valores de CCC variando entre 0,878 e 0,513, enquanto os valores de R^2 ficaram na faixa de 0,796 a 0,337, acompanhados de baixos índices de RMSE e MAE. Ademais, nota-se que os resultados de CCC e R^2 exibem uma forte concordância, evidenciando a robustez da metodologia de modelagem adotada.

Ao avaliar os valores de RMSE e MAE, observa-se pequenas variações entre os conjuntos de treinamento e teste, o que indica uma redução dos riscos de *overfitting*. Para o RMSE, conforme a ordem apresentada na Tabela 6, os modelos desenvolvidos para os parâmetros mostraram uma vantagem em relação ao RMSE do modelo nulo, variando de 52,65 a 11,82%. No caso do MAE, os modelos obtiveram desempenho superior, com uma vantagem variando de 66,04 a 13,82% em comparação com o modelo nulo. Conforme apresentado para os dados do sensor MSI/S2, apenas o modelo desenvolvido para o parâmetro Chla não demonstrou vantagem em relação ao modelo nulo, apresentando um RMSE 9,85% abaixo do valor médio e um MAE com leve vantagem de 7,28% em relação ao modelo NULL.

Na tabela 8 são apresentados, para os dados do sensor PS Normalizado, os resultados das métricas estatísticas utilizadas para a avaliação do desempenho dos modelos de aprendizado de máquina aplicados à predição de parâmetros da QA na bacia do rio Paraopeba. Observa-se que para todos os parâmetros avaliados os algoritmos RF e *Cubist* estão entre os dois melhores modelos, sendo que apenas nos parâmetros Fe, P e Chla esses algoritmos apresentaram desempenho inferior aos algoritmos SVM-RBF e KNN.

Ao analisar o desempenho dos modelos em relação aos parâmetros, verifica-se que os modelos para os parâmetros T, SST, Fe, P e OD apresentaram boa capacidade de generalização, com valores variando entre 0,918 e 0,553 para o índice estatístico CCC. Para o R^2 , os valores observados foram na faixa de 0,848 e 0,39. Ressalta-se que a boa concordância entre os dois índices constitui um fator importante para avaliar a robustez da metodologia aplicada.

Tabela 8: Valores das métricas estatísticas utilizadas para a avaliação de desempenho dos modelos *Cubist*, KNN, RF e SVM-RBF para a predição de parâmetros de QA (T – Turbidez, SST – Sólidos Suspensos Totais, Fe – Ferro, P – Fósforo, OD – Oxigênio Dissolvido, DQO – Demanda Bioquímica de Oxigênio, N – Nitrogênio e Chla – Clorofila-a) considerando os dados do sensor PlanetScope Normalizado.

	Modelos	Treinamento				Teste				NULL RMSE	NULL MAE
		RMSE	MAE	R ²	CCC	RMSE	MAE	R ²	CCC		
T	Cubist	55,939	29,051	0,852	0,927	56,389	29,982	0,848	0,918	140,817	96,021
	KNN	66,395	35,054	0,812	0,870	72,223	35,917	0,761	0,856		
	RF	66,144	34,668	0,821	0,873	68,845	34,304	0,784	0,869		
	SVM-RBF	67,443	34,888	0,816	0,854	71,007	34,627	0,771	0,843		
SST	<i>Cubist</i>	78,835	39,065	0,722	0,793	93,638	38,546	0,614	0,740	136,236	86,864
	KNN	76,613	39,054	0,713	0,794	90,293	41,962	0,608	0,745		
	RF	69,697	36,685	0,750	0,850	70,189	37,233	0,747	0,848		
	SVM-RBF	75,733	39,432	0,712	0,774	91,637	43,119	0,585	0,701		
Fe	<i>Cubist</i>	1,197	0,489	0,522	0,577	1,345	0,487	0,550	0,653	1,866	0,792
	KNN	1,211	0,478	0,528	0,589	1,084	0,424	0,656	0,764		
	RF	1,375	0,588	0,409	0,463	1,608	0,609	0,369	0,476		
	SVM-RBF	1,223	0,525	0,506	0,530	1,458	0,537	0,488	0,531		
P	<i>Cubist</i>	0,089	0,053	0,309	0,469	0,090	0,058	0,264	0,456	0,091	0,071
	KNN	0,078	0,066	0,307	0,472	0,087	0,055	0,258	0,457		
	RF	0,076	0,051	0,335	0,477	0,078	0,054	0,293	0,468		
	SVM-RBF	0,076	0,048	0,389	0,557	0,073	0,046	0,390	0,553		
OD	Cubist	1,150	0,798	0,238	0,353	1,017	0,704	0,392	0,557	1,304	0,852
	KNN	1,258	0,852	0,175	0,264	1,179	0,780	0,195	0,370		
	RF	1,099	0,739	0,269	0,381	1,114	0,708	0,271	0,402		
	SVM-RBF	1,209	0,815	0,156	0,265	1,259	0,829	0,127	0,257		
DQO	<i>Cubist</i>	10,085	7,906	0,195	0,345	9,599	7,470	0,214	0,381	10,838	8,487
	KNN	10,354	8,172	0,184	0,347	9,392	7,344	0,247	0,422		
	RF	9,252	7,289	0,280	0,414	9,041	7,140	0,302	0,445		
	SVM	9,867	7,827	0,209	0,349	9,234	7,265	0,273	0,401		
N	<i>Cubist</i>	0,226	0,175	0,258	0,404	0,228	0,174	0,228	0,390	0,257	0,202
	KNN	0,240	0,187	0,193	0,351	0,233	0,180	0,201	0,373		
	RF	0,220	0,171	0,285	0,425	0,221	0,172	0,272	0,414		
	SVM-RBF	0,229	0,178	0,240	0,378	0,232	0,178	0,212	0,383		
Chla	<i>Cubist</i>	2,690	1,461	0,129	0,211	1,923	1,221	0,024	0,150	2,650	1,485
	KNN	2,621	1,472	0,133	0,221	1,704	1,162	0,070	0,251		
	RF	2,392	1,373	0,150	0,255	1,604	1,169	0,094	0,271		
	SVM-RBF	2,336	1,310	0,140	0,239	1,490	1,064	0,115	0,264		

Onde: RMSE - raiz do erro quadrático médio, MAE - erro médio absoluto, CCC - coeficiente de correlação de concordância de Lin, R² - coeficiente de determinação e NULL RMSE – raiz do erro quadrático médio do modelo nulo e NULL MAE - erro médio absoluto do modelo nulo.

Nota: As linhas em negrito representam os modelos com melhor desempenho para cada parâmetro.

Ao avaliar os índices RMSE e MAE observa-se que os valores obtidos são baixos, com boa concordância entre os dados. Analisando o percentual de ganho dos modelos desenvolvidos em relação aos valores de RMSE NULL e MAE NULL, verifica-

se em todos os parâmetros avaliados que os modelos desenvolvidos apresentaram ganhos reais variando de 59,95 a 13,98% em relação ao RMSE NULL. No caso do MAE, os modelos desenvolvidos mostraram vantagem na faixa de 68,77 a 15,04%.

De forma geral, com exceção do parâmetro Chla nos dados provenientes dos sensores S2 e PS, todos os modelos desenvolvidos (Tabelas 6, 7 e 8) exibiram valores de MAE e RMSE inferiores aos do modelo nulo, o qual representa o limite mínimo de desempenho a ser superado pelos modelos (KUHN; JOHNSON, 2013; MELLO et al., 2022). Assim, a adoção dos modelos propostos é mais recomendada em comparação ao uso dos valores médios dos dados.

O parâmetro Chla apresentou os menores valores de CCC e R^2 , evidenciando uma grande dificuldade na generalização dos algoritmos nos três bancos de dados analisados. Ao avaliar as propriedades ópticas desse parâmetro, observa-se que a Chla é frequentemente afetada pela presença de outros Componentes Ópticamente Ativos (COAs), como SST e Matéria Orgânica Dissolvida Colorida (CDOM) (BERTONE et al., 2024; DALL'OLMO et al., 2005; KUTSER et al., 2005). Nesse contexto, destaca-se que a área de estudo foi impactada pelo rompimento de uma barragem de rejeitos de mineração, o que proporcionou elevados valores de SST nos conjuntos de dados utilizados na modelagem. Isso colaborou expressivamente para o baixo desempenho dos modelos de aprendizado de máquina para a predição do parâmetro Chla. Além disso, deve-se destacar que as características dos solos predominantes na região que, segundo a Embrapa (2006), são classificados como Latossolos Vermelhos, Latossolos Vermelho-Amarelos, Cambissolos Háplicos e Cambissolos Húmicos, impactam diretamente na coloração da água e, conseqüentemente, na sua resposta espectral.

A Tabela 9 apresenta os resultados das estatísticas dos modelos de aprendizado de máquina com melhor desempenho para cada parâmetro avaliado nas três abordagens distintas. Complementarmente, a Figura 4 exhibe os gráficos de dispersão dos valores preditos e observados em uma reta 1:1, permitindo uma avaliação visual da precisão.

Tabela 9: Valores das métricas estatísticas utilizadas para a avaliação do desempenho dos modelos com melhor desempenho para os parâmetros de qualidade da água (T – Turbidez, SST – Sólidos Suspensos Totais, Fe – Ferro, P – Fósforo, OD – Oxigênio Dissolvido, DQO – Demanda Bioquímica de Oxigênio, N – Nitrogênio e Chla – Clorofila-a) considerando os sensores Sentinel-2, PlanetScope e PlanetScope Normalizado.

	Dados.	Modelos	RMSE	MAE	R ²	CCC	NULL RMSE	NULL MAE
T	S2	RF	39,176	14,626	0,75	0,82	74,451	37,619
	PS	<i>Cubist</i>	66,664	32,609	0,80	0,88	140,817	96,021
	PS Norm	<i>Cubist</i>	56,389	29,982	0,85	0,92	140,817	96,021
SST	S2	KKNN	37,482	16,528	0,60	0,73	57,012	29,762
	PS	RF	83,945	40,389	0,66	0,77	136,236	86,864
	PS Norm	RF	70,189	37,233	0,75	0,85	136,236	86,864
Fe	S2	<i>Cubist</i>	0,474	0,213	0,28	0,45	0,519	0,269
	PS	<i>Cubist</i>	1,282	0,456	0,59	0,67	1,866	0,792
	PS Norm	KKNN	1,084	0,424	0,66	0,76	1,866	0,792
P	S2	SVM-RBF	0,0690	0,0430	0,25	0,43	0,077	0,055
	PS	SVM-RBF	0,076	0,049	0,34	0,51	0,091	0,071
	PS Norm	SVM-RBF	0,073	0,046	0,39	0,55	0,091	0,071
OD	S2	RF	1,325	0,864	0,20	0,36	1,457	1,006
	PS	<i>Cubist</i>	1,139	0,781	0,29	0,48	1,304	0,852
	PS Norm	<i>Cubist</i>	1,017	0,704	0,39	0,56	1,304	0,852
DQO	S2	RF	8,026	6,367	0,21	0,37	8,978	7,471
	PS	RF	9,681	7,565	0,22	0,39	10,838	8,487
	PS Norm	RF	9,041	7,140	0,30	0,45	10,838	8,487
N	S2	RF	0,221	0,177	0,16	0,31	0,237	0,191
	PS	RF	0,227	0,174	0,233	0,391	0,257	0,202
	PS Norm	RF	0,221	0,172	0,27	0,41	0,257	0,202
Chla	S2	SVM-RBF	3,079	1,253	0,05	0,12	3,031	1,364
	PS	KKNN	2,911	1,378	0,12	0,26	2,650	1,485
	PS Norm	SVM-RBF	1,490	1,064	0,11	0,26	2,650	1,485

Onde: RMSE - raiz do erro quadrático médio, MAE - erro médio absoluto, CCC - coeficiente de correlação de concordância de Lin, R² - coeficiente de determinação e NULL RMSE – raiz do erro quadrático médio do modelo nulo e NULL MAE - erro médio absoluto do modelo nulo.

Nota: As linhas em negrito representam os modelos com melhor desempenho para cada parâmetro.

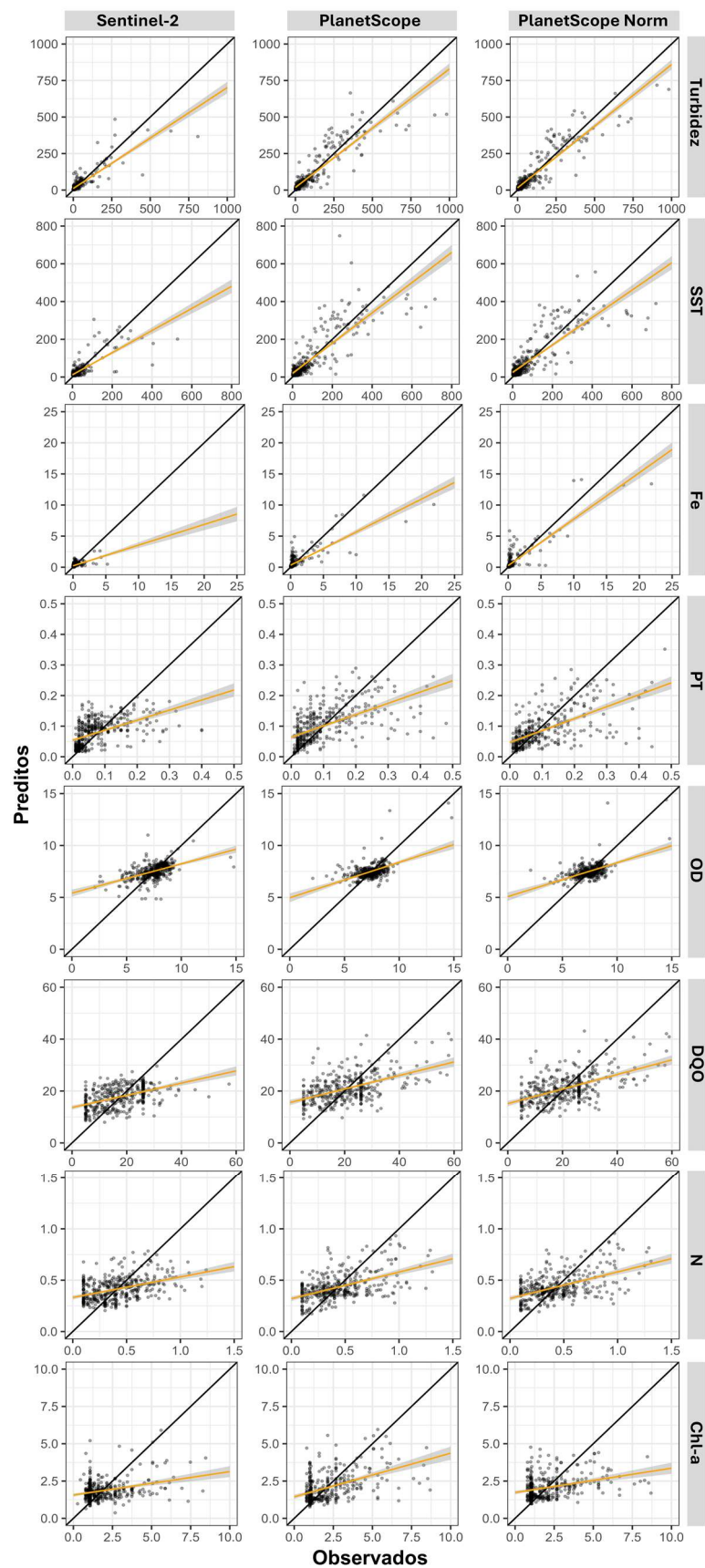


Figura 4: Comparação entre os valores de refletância preditos e observados dos modelos de aprendizado de máquina com melhor desempenho para os parâmetros T, SST, Fe, P, OD, DQO, N e Chl-a nos três bancos de dados analisado (S2, PS e PS Norm).

Ao analisar a Tabela 9 e a Figura 4, observa-se que o banco de dados proveniente das imagens PS normalizadas alcançou os melhores resultados nos modelos desenvolvidos para todos os parâmetros avaliados. Os parâmetros T, SST, Fe, P e OD apresentaram valores de CCC entre 0,92 e 0,55, enquanto para o R^2 os modelos exibiram valores de 0,85 a 0,39. Os valores de MAE e RMSE foram inferiores aos limites estabelecidos pelos modelos nulos. Quanto aos parâmetros DQO, N e Chla, embora tenham demonstrado valores superiores aos dos outros dois bancos de dados, os valores de CCC permaneceram abaixo de 0,50, variando entre 0,45 e 0,26; e os de R^2 oscilaram entre 0,30 e 0,11.

Os resultados demonstram a superioridade no desempenho dos modelos desenvolvidos com base no banco de dados PS, tanto nas imagens brutas quanto nas normalizadas, em relação aos dados S2 para todos os parâmetros analisados. Essa vantagem decorre diretamente das características do sensor PS, como a sua alta resolução espacial (3.7m) que permite a detecção de feições em escala mais detalhadas e a sua resolução temporal diária que permite uma melhor caracterização da variabilidade aquática. Esses resultados sugerem que o PS oferece vantagens expressivas para a modelagem de parâmetros de QA, particularmente em sistemas aquáticos complexos onde a variabilidade espacial e temporal é crítica.

O sensor MSI/S2, apesar de amplamente utilizado, apresenta limitações críticas em corpos hídricos estreitos (<30 m de largura), onde sua resolução espacial (10 m) induz a erros de mistura espectral. Além disso, sua refletância é sensível a interferências externas (leito fluvial, vegetação ripária e sedimentos), conforme comprovado por Barbosa et al. (2019), Greb et al. (2018) e Isidro et al. (2018), reduzindo sua precisão em sistemas aquáticos mais complexos.

Em suma, ao analisar as características de cada sensor nas três bases de dados avaliadas, os resultados indicam que os modelos desenvolvidos a partir dos dados do sensor PS normalizados apresentaram o melhor desempenho, superando os modelos baseados nos dados do PS bruto e do sensor MSI/Sentinel-2. Esse desempenho superior sugere que os dados PS normalizados são mais adequados para a predição dos parâmetros de QA, especialmente T, SST, Fe e P.

Nas Figura 5 e 6 são apresentados os dados preditos e observados dos parâmetros T, SST, Fe, P, OD, DQO, N e Chla, modelados a partir do sensor PS normalizado para ambientes lênticos e lóticos. De forma geral, observa-se que, devido às características do ambiente, os dados do ambiente lêntico apresentam baixa

dispersão, com valores próximos entre si. Por outro lado, no ambiente lótico, verifica-se maior dispersão nos dados de todos os parâmetros.

Ao analisar os índices estatísticos, observa-se que os parâmetros T, SST, DQO e N apresentam pequenas variações de desempenho do modelo entre os ambientes lênticos e lóticos. Por outro lado, os parâmetros Fe, P, OD e Chla demonstram variações mais expressivas de desempenho dos modelos entre esses dois ambientes. No ambiente lêntico, os parâmetros T, SST e OD apresentaram os melhores desempenhos, com valores de CCC na faixa de 0,97 a 0,72 e R^2 entre 0,95 e 0,65. Já Fe, P, DQO, N e Chla exibiram valores de CCC entre 0,43 e 0,12 e R^2 de 0,28 e 0,04. No ambiente lótico, T, SST e Fe destacaram-se com CCC, variando entre 0,91 e 0,68, e R^2 entre 0,83 e 0,52. Já os parâmetros OD, P, DQO, N e Chla apresentaram valores de CCC e R^2 mais baixos, variando de 0,48 a 0,19 e de 0,31 a 0,06, respectivamente.

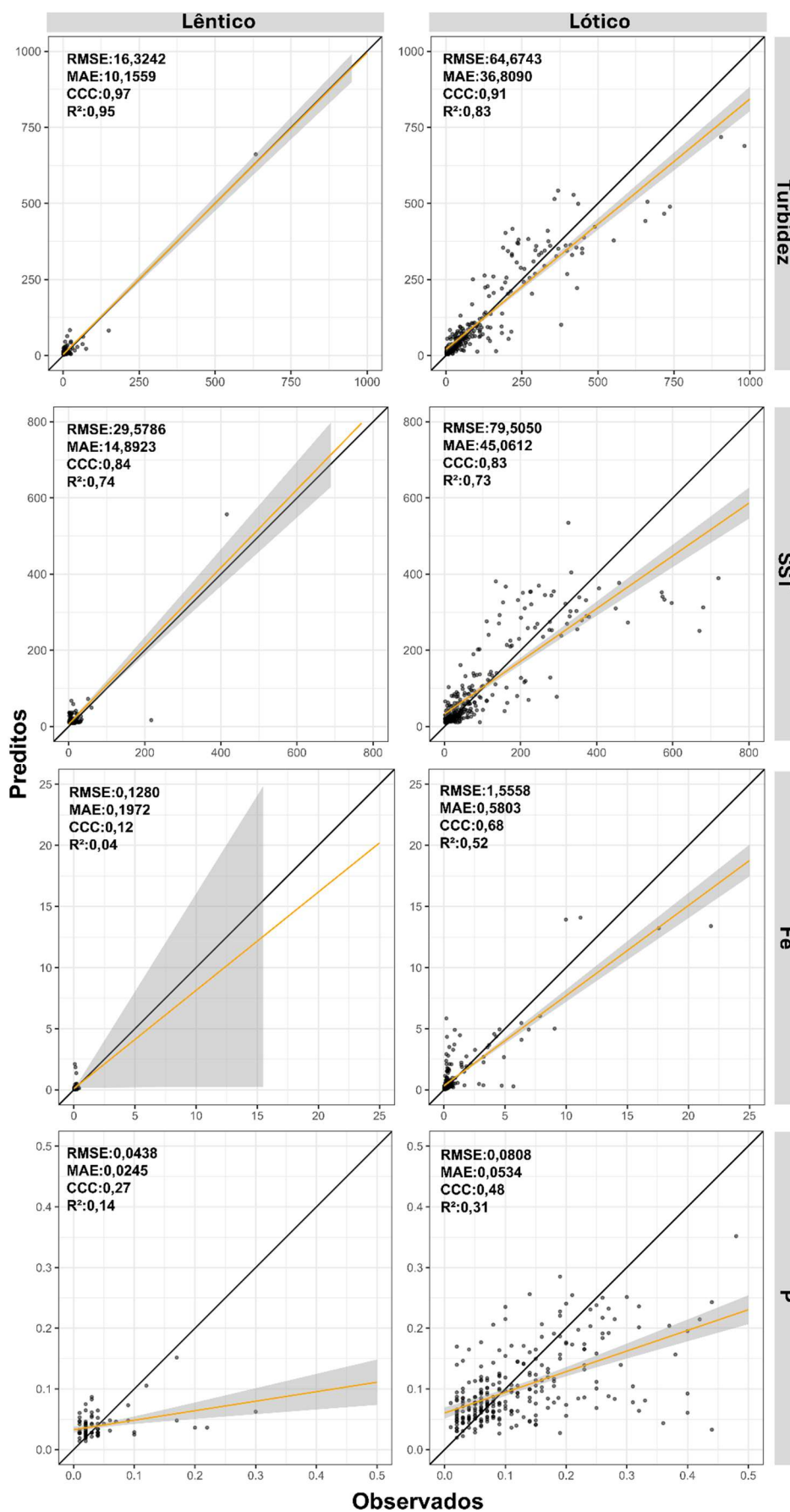


Figura 5: Dispersão dos dados preditos e observados parâmetros T, SST, Fe e P, modelados a partir do sensor PS normalizado para ambientes lênticos e lóticos.

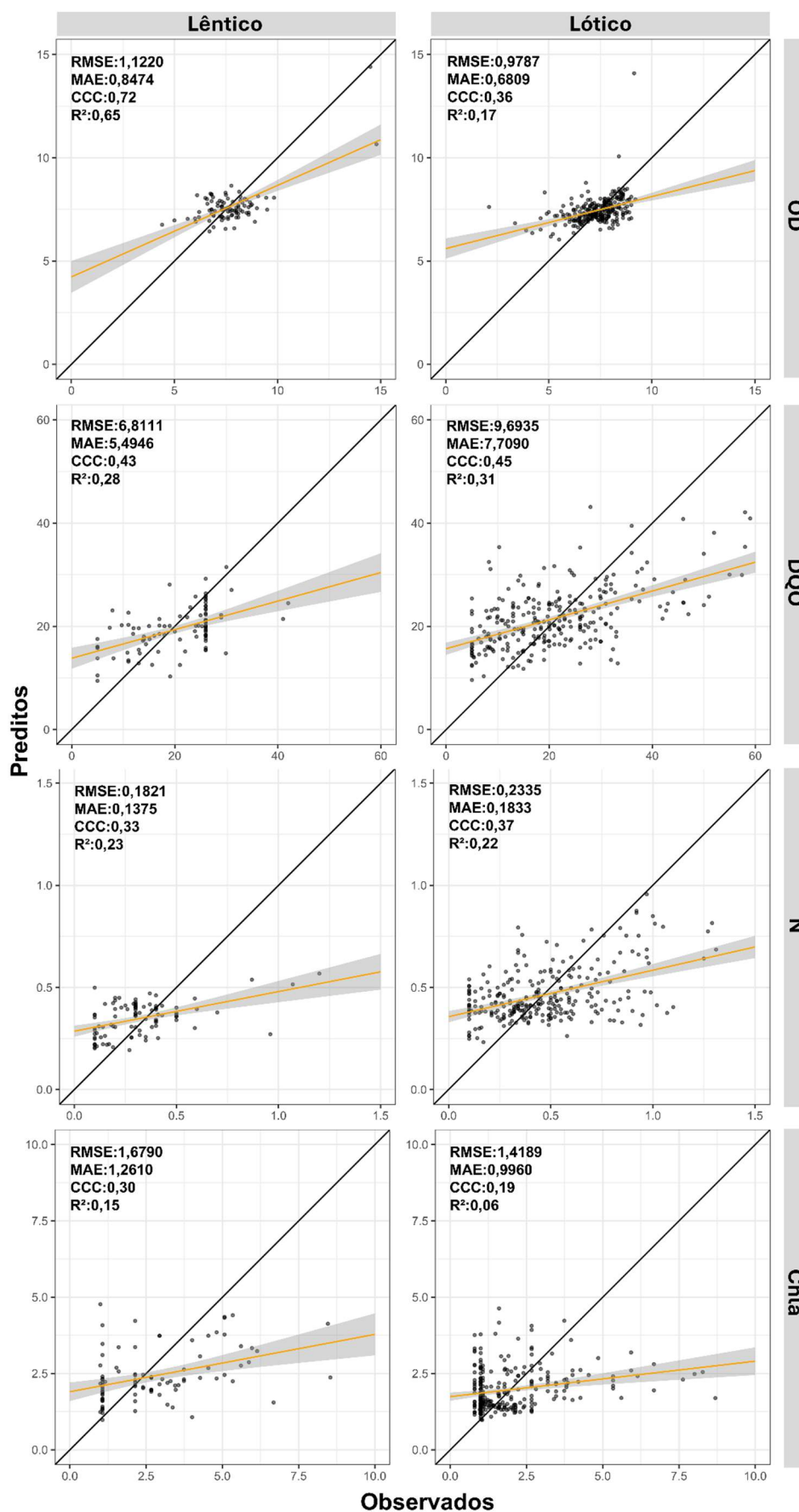


Figura 6: Dispersão dos dados preditos e observados parâmetros OD, DQO, N e Chla, modelados a partir do sensor PS normalizado para ambientes lênticos e lóticos.

A metodologia proposta neste estudo se mostrou robusta para a avaliação da QA, utilizando uma série histórica de imagens obtidas pelo sensor PS. A análise comparativa entre os dados do PS e os de uma constelação consolidada, como o S2, revelou a capacidade do PS de fornecer informações valiosas, mesmo diante das limitações impostas por sua menor resolução espectral e pelas diferenças radiométricas entre os sensores. Os resultados indicam que, embora o PS apresente restrições, ele demonstra elevado potencial para o monitoramento de parâmetros de QA em águas continentais, especialmente após o processo de normalização radiométrica.

A metodologia proposta não apenas valida o uso dos dados do sensor PS na predição de parâmetros de QA, como também fornece contribuições relevantes para a integração de diferentes fontes de dados, com o objetivo de aprimorar a precisão das predições. Com base nas evidências apresentadas, é possível afirmar que os nanossatélites Planet são uma ferramenta promissora para o monitoramento ambiental, especialmente em contextos que exigem monitoramento contínuo e de grande escala, o que abre novas possibilidades para a gestão de recursos hídricos e o entendimento dos impactos ambientais.

Ressalta-se que, embora os dados do sensor PS se destaquem como um dos produtos comerciais de imagem de alta resolução mais acessíveis e amplamente disponíveis no mercado, é importante observar que essas imagens não são gratuitas, ao contrário das imagens fornecidas pelo sensor MSI/S2. Assim, a utilização dos dados PS deve ser cuidadosamente avaliada por cada usuário, considerando a relação custo-benefício. Essa análise é fundamental para determinar se o investimento financeiro necessário para a aquisição das imagens do PS é justificável, especialmente diante da disponibilidade gratuita das imagens do S2, que pode representar uma alternativa viável a depender das demandas específicas do monitoramento.

No entanto, cabe destacar que as imagens do sensor MSI/S2 apresentam limitações, como menor resolução espacial em comparação ao PS, o que pode comprometer a detecção de alvos ou fenômenos em pequena escala. Além disso, fatores como a presença frequente de cobertura de nuvens e a menor frequência temporal de aquisição em determinadas regiões podem limitar a aplicabilidade dos dados do S2 em estudos que exigem alta resolução espacial e temporal

2.2.4 CONCLUSÕES

Com base nos resultados, conclui-se que:

- A metodologia desenvolvida para a predição dos parâmetros de qualidade da água (QA), utilizando dados de sensoriamento remoto orbital combinados com técnicas de aprendizado de máquina, mostrou-se adequada para predição dos parâmetros T, SST, Fe, P e OD. Entretanto, apresentou limitações para predição dos parâmetros N, DQO e Chla.

- A banda do infravermelho próximo (NIR) destacou-se como covariável fundamental nos modelos de predição, sendo consistentemente selecionada em todas as abordagens analisadas (B8 para o Sentinel-2 e B4 para o PS). Além disso, os índices espectrais, as razões entre bandas e os produtos CHIRPS apresentaram elevado potencial preditivo na estimativa dos parâmetros de qualidade da água (QA).

- Para melhorar o desempenho dos modelos na predição dos parâmetros de QA, a partir de dados do sensor PlanetScope (PS), é fundamental que as imagens do Sensor PlanetScope sejam normalizadas.

- Os modelos baseados em árvores de decisão, especificamente o *Random Forest* (RF) e o *Cubist*, foram mais robustos e com maior desempenho na predição dos parâmetros de QA do que os modelos KNN e SVM-RBF.

REFERÊNCIAS

- ABDULRAHMAN, S. A. et al. Comparative study for 8 computational intelligence algorithms for human identification. **Computer Science Review**, v. 36, p. 100237, 2020.
- ADAMS, A.; STERLING, L. **AI '92**. WORLD SCIENTIFIC, , dez. 1992. Disponível em: <<http://www.worldscientific.com/doi/abs/10.1142/9789814536271>>
- ADJOVU, G. E. et al. Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters. **Remote Sensing**, v. 15, n. 7, 2023.
- AGUIRRE-GUTIÉRREZ, J. et al. Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. **Remote Sensing of Environment**, v. 252, n. May 2020, p. 112122, 2021.
- ALTMAN, D. G. **Practical statistics for medical research**. [s.l.] Chapman and Hall/CRC, 1990.
- ALVARES, C. A. et al. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2014.
- ANA. **Agência Nacional de Águas. Conjuntura dos recursos hídricos no Brasil 2017: relatório pleno / Agência Nacional de Águas**. Brasília - DF, 2017.
- ANSPER, A.; ALIKAS, K. Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. **Remote Sensing**, v. 11, n. 1, p. 64, 2018.
- ARANGO, J. G.; NAIRN, R. W. Prediction of Optical and Non-Optical Water Quality Parameters in Oligotrophic and Eutrophic Aquatic Systems Using a Small Unmanned Aerial System. **Drone**, 2020.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v. 4, p. 40–79, 2010.
- ASNER, G. P. Cloud cover in Landsat observations of the Brazilian Amazon. **International Journal of Remote Sensing**, v. 22, p. 3855–3862, 2010.
- BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos**. 1º ed. São José dos Campos: INPE, 2019.
- BARNES, J. D. et al. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. **Environmental and Experimental botany**, v. 32, n. 2, p. 85–100, 1992.
- BERTONE, E. et al. Satellite-based estimation of total suspended solids and chlorophyll-a concentrations for the Gold Coast Broadwater , Australia. **Marine Pollution Bulletin**, v. 201, n. December 2023, p. 116217, 2024.

BEVEN, K. J.; CLOKE, H. L. Comment on “hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al. **Water Resources Research**, v. 48, n. 1, p. 1–10, 2012.

BEYCA, O. F. et al. Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. **Energy Economics**, v. 80, p. 937–949, 2019.

BILGIN, A.; KONANÇ, M. U. Evaluation of surface water quality and heavy metal pollution of Coruh River Basin (Turkey) by multivariate statistical methods. **Environmental Earth Sciences**, v. 75, p. 1–18, 2016.

BO, Y. et al. Additional surface-water deficit to meet global universal water accessibility by 2030. **Journal of Cleaner Production**, v. 320, n. August, p. 128829, 2021.

BONANSEA, M. et al. Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in Río Tercero reservoir (Argentina). **Remote Sensing of Environment**, v. 158, p. 28–41, 2015.

BOUSBIH, S. et al. Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2. **Remote Sensing**, v. 11, n. 13, 2019.

BREIMAN, L. Random forests. **Machine learning**, v. 45.1, p. 27, 2001.

BRUCE, A.; BRUCE, P. **Estatística prática para cientistas de dados**. [s.l.] Alta Books, 2019.

CALAZANS, G. M. et al. The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. **Environmental Monitoring and Assessment**, v. 190, n. 8, 2018.

CARVALHO, M. L. S.; CABÚS, R. C. Eficiência da luz solar refletida e desempenho de dispositivos de sombreamento. **Ambiente Construído**, v. 20, n. 2, p. 191–209, 2020.

CEMIG. **CEMIG Website**. , 2021. Disponível em: <<https://www.cemig.com.br>>. Acesso em: 22 set. 2021

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. **Geoscientific model development**, v. 7, n. 3, p. 1247–1250, 2014.

CHAWLA, I.; KARTHIKEYAN, L.; MISHRA, A. K. A review of remote sensing applications for water security: Quantity, quality, and extremes. **Journal of Hydrology**, v. 585, n. March, p. 124826, 2020.

CHIRPS. **CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara**. Disponível em: <<https://www.chc.ucsb.edu/data/chirps>>. Acesso em: 23 jul. 2023.

CLAVERIE, M. et al. The Harmonized Landsat and Sentinel-2 surface reflectance data set. **Remote Sensing of Environment**, v. 219, n. August 2017, p. 145–161, 2018.

- CLEVERS, J. et al. Derivation of the red edge index using the MERIS standard band setting. **International Journal of Remote Sensing**, v. 23, n. 16, p. 3169–3184, 2002.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, v. 20.3, p. 25, 1995.
- COSTA, L.; NUNES, L.; AMPATZIDIS, Y. A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. **Computers and Electronics in Agriculture**, v. 172, n. November 2019, p. 105334, 2020.
- COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE transactions on information theory**, v. 13, n. 1, p. 21–27, 1967.
- CRUSAN, J.; GALICA, C. NASA's CubeSat Launch Initiative: Enabling broad access to space. **Acta Astronautica**, v. 157, p. 51–60, 2019.
- CSILLIK, O. et al. Monitoring tropical forest carbon stocks and emissions using Planet satellite data. **Scientific Reports**, v. 9, n. 1, p. 1–12, 2019.
- CUI, Y. et al. Deep learning–based remote sensing estimation of water transparency in shallow lakes by combining Landsat 8 and Sentinel 2 images. **Environmental Science and Pollution Research**, 2021.
- DA SILVEIRA, V. A. et al. Modeling and mapping of Inselberg habitats for environmental conservation in the Atlantic Forest and Caatinga domains, Brazil. **Environmental Advances**, v. 8, n. October 2021, 2022.
- DALL'OLMO, G. et al. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. **Remote Sensing of Environment**, v. 96, n. 2, p. 176–187, 2005.
- DE CARVALHO JÚNIOR, O. A. et al. Radiometric normalization of temporal images combining automatic detection of pseudo-invariant features from the distance and similarity spectral measures, density scatterplot analysis, and robust regression. **Remote Sensing**, v. 5, n. 6, p. 2763–2794, 2013.
- DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing**. (P. V. U. Amsterdam, Ed.) **Proefschrift Vrije Universiteit Amsterdam**, 2001.
- DEVENTER, A. P. VAN et al. Using thematic mapper data to identify contrasting soil plains and tillage practices. **Photogrammetric Engineering and Remote Sensing**, v. 63, p. 87–93, 1997.
- DIAS, R. L. S. et al. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatics**, n. August, p. 101414, 2021.
- DONG, J. et al. A survey of smart water quality monitoring system. **Environmental Science and Pollution Research**, v. 22, p. 4893–4906, 2015.

DOS SANTOS, E. P. et al. 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. February, 2022.

DU, C. et al. Estimation of total phosphorus concentration using a water classification method in inland water. **International Journal of Applied Earth Observation and Geoinformation**, v. 71, n. April, p. 29–42, 2018.

DU, Y. et al. Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band. **Remote Sensing**, v. 8, n. 4, 2016.

DU, Y.; TEILLET, P. M.; CIHLAR, J. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. **Remote Sensing of Environment**, v. 82, n. 1, p. 123–134, 2002.

DUAN, W. et al. Spatiotemporal evaluation of water quality incidents in Japan between 1996 and 2007. **Chemosphere**, v. 93, n. 6, p. 946–953, 2013a.

DUAN, W. et al. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River , Japan , 1985 to 2010. **Science of the Total Environment**, The, v. 461–462, p. 499–508, 1 set. 2013b.

DUDANI, S. A. The distance-weighted k-nearest-neighbor rule. **IEEE Transactions on Systems, Man, and Cybernetics**, n. 4, p. 325–327, 1976.

DUNN, O. J. Multiple comparisons using rank sums. **Technometrics**, v. 6, n. 3, p. 241–252, 1964.

EMBRAPA. **Sistema brasileiro de classificação de solos**. (EMBRAPA, Ed.)Rio de Janeiro, 2006.

ENGMAN, E. T.; GURNEY, R. J. **Remote sensing in hydrology**. Londres: Chapman and Hall Ltd, 1991.

ESA. **Guia de Missão Sentinel 2** , 2023. Disponível em: <<https://sentinel.esa.int/web/sentinel/missions/sentinel-2>>

EVERARD, M. Meeting global drinking water needs. **Nature Sustainability**, v. 2, n. 5, p. 360–361, 2019.

FERDOUS, J.; RAHMAN, M. T. U. Developing an empirical model from Landsat data series for monitoring water salinity in coastal Bangladesh. **Journal of Environmental Management**, v. 255, n. October 2019, p. 109861, 2020.

FERREIRA, R. G. et al. Machine learning models for streamflow regionalization in a tropical watershed. **Journal of Environmental Management**, v. 280, n. August 2020, 2021.

FEYISA, G. L. et al. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. **Remote Sensing of Environment**, v. 140, p.

23–35, 2014.

FILGUEIRAS, R. et al. Crop NDVI monitoring based on sentinel 1. **Remote Sensing**, v. 11, n. 12, 2019.

FISHER, A.; FLOOD, N.; DANAHER, T. Comparing Landsat water index methods for automated water classification in eastern Australia. **Remote Sensing of Environment**, v. 175, p. 167–182, 2016.

FRANCINI, S. et al. Near-real time forest change detection using PlanetScope imagery. **European Journal of Remote Sensing**, v. 53, n. 1, p. 233–244, 2020.

FRAZIER, A. E.; HEMINGWAY, B. L. A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. **Remote Sensing 2021, Vol. 13, Page 3930**, v. 13, n. 19, p. 3930, 2021.

FUNK, C. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific data**, v. 2, n. 1, p. 1–21, 2015.

GAO, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote sensing of environment**, v. 58, n. 3, p. 257–266, 1996.

GAO, B. et al. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. **Remote sensing of environment**, v. 113, p. S17–S24, 2009.

GAO, Y. et al. Remote sensing estimation of the total phosphorus concentration in a large lake using band combinations and regional multivariate statistical modeling techniques. **Journal of Environmental Management**, v. 151, p. 33–43, 2015.

GASCON, F. et al. Copernicus Sentinel-2A calibration and products validation status. **Remote Sensing**, v. 9, n. 6, 2017.

GHOLIZADEH, H. M.; ASSEFA, M. M.; REDDI, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. **sensors**, 2016.

GHOLIZADEH, M. H.; MELESSE, A. M.; REDDI, L. **A comprehensive review on water quality parameters estimation using remote sensing techniques. Sensors (Switzerland)**MDPI AG, , 16 ago. 2016.

GHOSH, A.; JOSHI, P. K. A comparison of selected classification algorithms for mappingbamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, n. 1, p. 298–311, 2014.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289–298, 1996.

GOMES, L. C. et al. Modelling and mapping soil organic carbon stocks in Brazil.

Geoderma, v. 340, p. 337–350, 2019.

GREB, S. et al. Earth observations in support of global water quality monitoring. 2018.

HAFAEEZ, S. et al. Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: A case study of Hong Kong. **Remote sensing**, v. 11, n. 6, p. 617, 2019.

HAJIGHOLIZADEH, M. et al. Land–lake linkage and remote sensing application in water quality monitoring in lake okeechobee, florida, usa. **Land**, v. 10, n. 2, p. 147, 2021.

HELMAN, D. et al. Using time series of high-resolution planet satellite images to monitor grapevine stem water potential in commercial vineyards. **Remote Sensing**, v. 10, n. 10, p. 1–22, 2018.

HENRICH, V. et al. **Development of an online indices database: Motivation, concept and implementation**. Proceedings of the 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications, Tel Aviv, Israel. **Anais...**2009.

HEWSON, R. D.; CUDAHY, T. J.; HUNTINGTON, J. F. **Geologic and alteration mapping at Mt Fitton, South Australia, using ASTER satellite-borne data**. IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217). **Anais...**2001.

HIJMANS, R. J. et al. Package ‘terra’. **Maintainer: Vienna, Austria**, 2022.

HOUBORG, R.; MCCABE, M. F. High-Resolution NDVI from planet’s constellation of earth observing nano-satellites: A new data source for precision agriculture. **Remote Sensing**, v. 8, n. 9, 2016.

HOUBORG, R.; MCCABE, M. F. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. **Remote Sensing**, v. 10, n. 6, 2018a.

HOUBORG, R.; MCCABE, M. F. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. **Remote Sensing of Environment**, v. 209, n. February, p. 211–226, 2018b.

HOUBORG, R.; MCCABE, M. F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 135, p. 173–188, jan. 2018c.

HUETE, A. R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295–309, 1988.

IGAM. **Instituto Mineiro de Gestão das Águas. Água Superficial.** , 2021. Disponível em: <<http://portalinfohidro.igam.mg.gov.br/sem-categoria/318-agua-superficial>>

IORDANIS, I.; KOUKOUVINOS, C.; SILOU, I. On the efficacy of conditioned and

progressive Latin hypercube sampling in supervised machine learning. **Applied Numerical Mathematics**, n. December 2023, 2024.

ISIDRO, C. M. et al. Quantifying suspended solids in small rivers using satellite data. **Science of the Total Environment**, v. 634, p. 1554–1562, 2018.

JIANG, J. et al. Multi-sensor and multi-platform consistency and interoperability between UAV, Planet CubeSat, Sentinel-2, and Landsat reflectance data. **GIScience and Remote Sensing**, v. 59, n. 1, p. 936–958, 2022.

KAUFMAN, Y. J.; TANRE, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. **IEEE transactions on Geoscience and Remote Sensing**, v. 30, n. 2, p. 261–270, 1992.

KEITH, D. et al. Monitoring algal blooms in drinking water reservoirs using the landsat-8 operational land imager. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2818–2846, 2018.

KENNEDY, J. B.; NEVILLE, A. M. **Basic Statistical Methods for Engineers and Scientists**. Nova IorqueHarper and Row, , 1986.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. **Journal of the American statistical Association**, v. 47, n. 260, p. 583–621, 1952.

KUHN, M. **caret: Classification and Regression Training**. , 2020. Disponível em: <<https://cran.r-project.org/package=caret>>

KUHN, M.; JOHNSON, K. **Applied Predictive Modeling**. [s.l.] Springer, 2013.

KUTSER, T. et al. Mapping lake CDOM by satellite remote sensing. **Remote Sensing of Environment**, v. 94, n. 4, p. 535–540, 2005.

LACAUX, J. P. et al. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. **Remote Sensing of Environment**, v. 106, n. 1, p. 66–74, 2007.

LAI, W. et al. A portable algorithm to retrieve bottom depth of optically shallow waters from top-of-atmosphere measurements. **Journal of Remote Sensing**, 2022.

LAMQUIN, N. et al. An inter-comparison exercise of Sentinel-2 radiometric validations assessed by independent expert groups. **Remote Sensing of Environment**, v. 233, n. August, 2019.

LATTE, N.; LEJEUNE, P. PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. **Remote Sensing**, v. 12, n. 15, 2020.

LEACH, N.; COOPS, N. C.; OBRKNEZEV, N. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. **Computers and Electronics in Agriculture**, v. 164, n. January, p. 104893, 2019.

LEE RODGERS, J.; NICEWANDER, W. A. Thirteen Ways to Look at the Correlation Coefficient. **The American Statistician**, v. 42, n. 1, p. 59–66, fev. 1988.

LI, H.; LIU, Q. **Comparison of NDBI and NDVI as indicators of surface urban heat island effect in MODIS imagery**. Proc.SPIE. **Anais...**2008. Disponível em: <<https://doi.org/10.1117/12.815679>>

LI, J. et al. Recovering low quality MODIS-Terra data over highly turbid waters through noise reduction and regional vicarious calibration adjustment: A case study in Taihu Lake. **Remote Sensing of Environment**, v. 197, p. 72–84, 2017a.

LI, W. et al. Generating red-edge images at 3M spatial resolution by fusing sentinel-2 and planet satellite products. **Remote Sensing**, v. 11, n. 12, p. 1–18, 2019.

LI, Y. et al. Monitoring spatiotemporal variations in nutrients in a large drinking water reservoir and their relationships with hydrological and meteorological conditions based on Landsat 8 imagery. **Science of the Total Environment**, v. 599–600, p. 1705–1717, 1 dez. 2017b.

LILLESAND, T.; KIEFER, R. W.; CHIPMAN, J. **Remote sensing and image interpretation**. [s.l.] John Wiley & Sons, 2015.

LIN, L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. **Biometrics**, v. 45, n. 1, p. 255, mar. 1989.

LOUIS, J. et al. Sentinel-2 SEN2COR: L2A processor for users. **European Space Agency**, v. SP-740, n. August, p. 9–13, 2016.

MARTINS, V. S. et al. Assessment of atmospheric correction methods for sentinel-2 MSI images applied to Amazon floodplain lakes. **Remote Sensing**, v. 9, n. 4, 2017.

MATHEW, M. M.; SRINIVASA RAO, N.; MANDLA, V. R. Development of regression equation to study the Total Nitrogen, Total Phosphorus and Suspended Sediment using remote sensing data in Gujarat and Maharashtra coast of India. **Journal of Coastal Conservation**, v. 21, n. 6, p. 917–927, 1 dez. 2017.

MATTHEWS, M. W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. **International Journal of Remote Sensing**, v. 32, n. 21, p. 6855–6899, 2011.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996a.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 27 maio 1996b.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code.

Technometrics, v. 42, n. 1, p. 55–61, 1979.

MCKIGHT, P. E.; NAJAB, J. Kruskal-wallis test. **The corsini encyclopedia of psychology**, p. 1, 2010.

MELLO, D. C. D. et al. A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. **Geoscientific Model Development**, v. 15, n. 3, p. 1219–1246, 2022.

MEYER, H. et al. From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: Do we need hyperspectral information? **International Journal of Applied Earth Observation and Geoinformation**, v. 55, p. 21–31, 2017.

MEYER, H. et al. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. **Environmental Modelling and Software**, v. 101, p. 1–9, 2018.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers and Geosciences**, v. 32, n. 9, p. 1378–1388, 2006.

MORSE-MCNABB, E. M.; HASAN, M. F.; KARUNARATNE, S. A Multi-Variable Sentinel-2 Random Forest Machine Learning Model Approach to Predicting Perennial Ryegrass Biomass in Commercial Dairy Farms in Southeast Australia. **Remote Sensing**, v. 15, n. 11, 2023.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 3, p. 247–259, 2011.

MUKHERJEE, N. R.; SAMUEL, C. Assessment of the temporal variations of surface water bodies in and around Chennai using landsat imagery. **Indian Journal of Science and Technology**, v. 9, n. 18, 2016.

MÜLLER-WILM, U. Sentinel-2 MSI – Level-2A Prototype Processor Installation and User Manual. **European Space Agency, (Special Publication) ESA SP**, v. 49, n. 2.2, p. 1–51, 2016.

MUÑOZ-ROMERO, S. et al. Informative variable identifier: Expanding interpretability in feature selection. **Pattern Recognition**, v. 98, 2020.

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective. Chance Encounters: Probability in Education** London, England The MIT Press, , 2013.

NASA. **Detalhes da missão do Landsat 8**. , 2023. Disponível em: <<https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-mission-details>>

NEOGI, S.; DAUWELS, J. Factored Latent-Dynamic Conditional Random Fields for single and multi-label sequence modeling. **Pattern Recognition**, v. 122, p. 108236,

2022.

NGUYEN, U. N. T.; PHAM, L. T. H.; DANG, T. D. An automatic water detection approach using Landsat 8 OLI and Google Earth Engine cloud computing to map lakes and reservoirs in New Zealand. **Environmental monitoring and assessment**, v. 191, p. 1–12, 2019.

NOI, P. T.; DEGENER, J.; KAPPAS, M. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. **Remote Sensing**, v. 9, n. 5, 2017.

NOVO, E. M. L. M. **Sensoriamento remoto: princípios e aplicações**. 4. ed. São Paulo: Blucher, 2010.

OGASHAWARA, I.; MISHRA, D. R.; GITELSON, A. A. **Remote Sensing of Inland Waters: Background and Current State-of-the-Art**. [s.l.] Elsevier Inc., 2017.

ONU. **Objetivos de Desenvolvimento Sustentável**. , 2015.

PAHLEVAN, N. et al. Sentinel-2/Landsat-8 product consistency and implications for monitoring aquatic systems. **Remote Sensing of Environment**, v. 220, n. October 2018, p. 19–29, 2019.

PAPA, F. et al. **Water Resources in Africa under Global Change: Monitoring Surface Waters from Space**. [s.l.: s.n.]. v. 44

PIZANI, F. M. C.; MAILLARD, P. Um Índice De Turbidez Para Águas Relativamente Claras. **Anais do XX Simpósio Brasileiro de Sensoriamento Remoto**, p. 668–671, 2023.

PLANET TEAM. Planet Surface Reflectance Product v2. Planet Labs, Inc. Accessed 30.11.2020.

https://assets.planet.com/marketing/PDF/Planet_Surface_Reflectance_Technical_White_Paper.pdf. p. 1–18, 2020.

PLANET TEAM. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA, USA. **Planet Labs Inc**, p. 1–96, 2021.

POLIDORIO, A. M.; IMAI, N. N.; TOMMASELLI, A. M. G. Índice indicador de corpos d'água para imagens multiespectrais. I **Simpósio de Ciências Geodésicas e Tecnologias da Geoinformação**, v. 9, 2004.

POLITIS, D. N. et al. **Subsampling in the IID Case**. [s.l.] Springer, 1999.

R CORE TEAM. **R: A language and environment for statistical computing, version 3.3. 1. Vienna, Austria: R Foundation for Statistical Computing**. , 2020. Disponível em: <<https://www.r-project.org>>

RAMARAJ, M.; SIVAKUMAR, R. Remote Sensing and Nonlinear Auto-regressive Neural Network (NARNET) Based Surface Water Chemical Quality Study: A Spatio-

Temporal Hybrid Novel Technique (STHNT). **Bulletin of Environmental Contamination and Toxicology**, v. 110, n. 1, p. 1–10, 2023.

REUNANEN, J. Overfitting in making comparisons between variable selection methods. **Journal of Machine Learning Research**, v. 3, p. 1371–1382, 2003.

RIPLEY, B. D. **Pattern recognition and neural networks**. [s.l.] Cambridge university press, 1996.

ROUSE, J. W. et al. Monitoring vegetation systems in the Great Plains with ERTS. **NASA special publication**, v. 351, n. 1974, p. 309, 1974.

ROWAN, L. C.; MARS, J. C. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. **Remote Sensing of Environment**, v. 84, n. 3, p. 350–366, 2003.

ROY, D. P. et al. Landsat-8: Science and product vision for terrestrial global change research. **Remote Sensing of Environment**, v. 145, p. 154–172, 2014.

SABERIOON, M. et al. Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. **Ecological Indicators**, v. 113, n. March, 2020.

SADEH, Y. et al. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3 m surface reflectance and wheat LAI monitoring. **International Journal of Applied Earth Observation and Geoinformation**, v. 96, p. 102260, 2021a.

SADEH, Y. et al. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3 m surface reflectance and wheat LAI monitoring. **International Journal of Applied Earth Observation and Geoinformation**, v. 96, p. 102260, 2021b.

SAGAN, V. et al. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. **Earth-Science Reviews**, v. 205, p. 103187, 2020.

SCHEFFLER, D. et al. AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data. **Remote Sensing**, v. 9, n. 7, 2017.

SCHOTT, J. R.; SALVAGGIO, C.; VOLCHOK, W. J. Radiometric Scene Normalization Using Pseudoinvariant Features. **Remote Sensing of Environment**, v. 16, p. 1–16, 1988.

SCHROEDER, T. A. et al. Radiometric correction of multi-temporal Landsat data for characterization of early successional forest patterns in western Oregon. **Remote Sensing of Environment**, v. 103, n. 1, p. 16–26, 2006.

SENA, N. C. et al. Soil sampling strategy in areas of difficult access using the cLHS method. **Geoderma Regional**, v. 24, 2021.

SESTINI, M. F. **Variáveis Geomorfológicas no estudo de deslizamentos em Caraguatatuba-SP utilizando imagens TM-Landsat e SIG**. São José dos

Campos/SP: Instituto Nacional de Pesquisas Espaciais, 1999.

SHEPARD, D. **A two-dimensional interpolation function for irregularly-spaced data**. Proceedings of the 1968 23rd ACM national conference. **Anais...**1968.

SHIN, J. et al. High spatial-resolution red tide detection in the southern coast of Korea using u-net from planetscope imagery. **Sensors**, v. 21, n. 13, 2021.

SLOGGETT, D. et al. Operational uses of ocean colour data-Perspectives for the OCTOPUS Programme. **Sensors and Environmental Applications of Remote Sensing**, 1995.

SONG, K. et al. Quantification of lake clarity in China using Landsat OLI imagery data. **Remote Sensing of Environment**, v. 243, p. 111800, 2020.

STEVENS, A. et al. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. **PLoS ONE**, v. 8, n. 6, 2013.

SWAIN, R.; SAHOO, B. Mapping of heavy metal pollution in river water at daily time-scale using spatio-temporal fusion of MODIS-aqua and Landsat satellite images. **Journal of Environmental Management**, v. 192, p. 1–14, 2017.

SYARIZ, M. A. et al. Spectral-consistent relative radiometric normalization for multitemporal Landsat 8 imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 147, n. November 2018, p. 56–64, 2019.

TAVAKOL, M. et al. Application of multivariate statistical methods to optimize water quality monitoring network with emphasis on the pollution caused by fish farms. **Iranian Journal of Public Health**, v. 46, n. 1, p. 83, 2017.

TIAN, S. et al. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. **Environmental Science and Pollution Research**, v. 30, n. 7, p. 18617–18630, 2023.

TIYASHA, T. et al. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. **Marine Pollution Bulletin**, v. 170, n. May, p. 112639, 2021.

TOMING, K. et al. First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery. **Remote Sensing**, v. 8, n. 8, 2016.

TUCKER, C. J. **Red and Photographic Infrared Linear Combinations for Monitoring Vegetation**. **REMOTE SENSING OF ENVIRONMENT**, 1979.

VANHELLEMONT, Q.; RUDDICK, K. **Acolite for Sentinel-2: Aquatic applications of MSI imagery**. Proceedings of the 2016 ESA Living Planet Symposium, Prague, Czech Republic. **Anais...**2016.

WALDNER, F. et al. High temporal resolution of leaf area data improves empirical estimation of grain yield. **Scientific Reports**, v. 9, n. 1, p. 1–14, 2019.

WANG, H. et al. Technological Forecasting & Social Change Tension in big data using machine learning : Analysis and applications ☆. **Technological Forecasting & Social Change**, v. 158, n. May 2019, p. 120175, 2020.

WANG, J. et al. Automatic cloud and cloud shadow detection in tropical areas for PlanetScope satellite images. **Remote Sensing of Environment**, v. 264, n. May, p. 112604, 2021.

WEGMUELLER, S. A.; LEACH, N. R.; TOWNSEND, P. A. LOESS radiometric correction for contiguous scenes (LORACCS): Improving the consistency of radiometry in high-resolution satellite image mosaics. **International Journal of Applied Earth Observation and Geoinformation**, v. 97, p. 102290, 2021.

WILKS, D. S. **Statistical methods in the atmospheric sciences**. [s.l.] Academic press, 2011.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate research**, v. 30, n. 1, p. 79–82, 2005.

WILSON, E. H.; SADER, S. A. Detection of forest harvest type using multiple dates of Landsat TM imagery. **Remote Sensing of Environment**, v. 80, n. 3, p. 385–396, 2002.

WINSTON, R. J.; DORSEY, J. D.; HUNT, W. F. Quantifying volume reduction and peak flow mitigation for three bioretention cells in clay soils in northeast Ohio. **Science of The Total Environment**, v. 553, p. 83–95, maio 2016.

WU, S. et al. Monitoring tree-crown scale autumn leaf phenology in a temperate forest with an integration of PlanetScope and drone remote sensing observations. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 171, n. June 2020, p. 36–48, 2021.

XIONG, Y. et al. Remotely assessing and monitoring coastal and inland water quality in China: Progress, challenges and outlook. **Critical Reviews in Environmental Science and Technology**, v. 50, n. 12, p. 1266–1302, 2020.

XU, H. et al. A novel automatic method on pseudo-invariant features extraction for enhancing the relative radiometric normalization of high-resolution images. **International Journal of Remote Sensing**, v. 42, n. 16, p. 6155–6186, 2021.

YANG, H. et al. A review of remote sensing for water quality retrieval: progress and challenges. **Remote Sensing**, v. 14, n. 8, p. 1770, 2022.

YANG, X.; LO, C. P. Relative radiometric normalization performance for change detection from multi-date satellite images. **Photogrammetric Engineering and Remote Sensing**, v. 66, n. 8, p. 967–980, 2000.

YIGIT AVDAN, Z. et al. Monitoring the water quality of small water bodies using high-resolution remote sensing data. **ISPRS International Journal of Geo-Information**, v. 8, n. 12, p. 553, 2019.

YOUNG, S. L. et al. Perspective: The Importance of Water Security for Ensuring Food Security, Good Nutrition, and Well-being. **Advances in Nutrition**, v. 12, n. 4, p. 1058–1073, 2021.

ZHANG, K. et al. Remote sensing of seasonal changes and disturbances in mangrove forest: A case study from South Florida. **Ecosphere**, v. 7, n. 6, 2016.

ZHANG, L.; WU, C.; DU, B. Automatic Radiometric Normalization for Multitemporal Remote Sensing Imagery With Iterative Slow Feature Analysis. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 10, p. 6141–6155, 2014.

ZHANG, M. L.; ZHOU, Z. H. ML-KNN: A lazy learning approach to multi-label learning. **Pattern Recognition**, v. 40, n. 7, p. 2038–2048, 2007.

ZHANG, Y.; LI, A.; DENG, B. Data-driven predictive models for chemical durability of oxide glass under different chemical conditions. **npj Materials Degradation**, p. 1–11, 2020.

ZHU, Z. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 370–384, 2017.

3 CONCLUSÕES GERAIS

A normalização radiométrica desenvolvida para o sensor PlanetScope (PS), com base nos dados Sentinel-2, mostrou-se necessária e adequada para atenuar as inconsistências da constelação de nanossatélites da Planet.

A modelagem utilizando os dados provenientes dos sensores PS normalizados apresentou melhor desempenho na predição dos parâmetros de qualidade da água em comparação aos dados PS brutos e do Sentinel-2.

O modelo desenvolvido com base no algoritmo *Random Forest* foi mais eficiente e robusto na predição dos parâmetros de qualidade da água em comparação aos modelos baseados nos algoritmos *Cubist*, KNN e SVM-RBF.

A metodologia proposta, a qual utilizada dados normalizados do sensor PS aliada a técnicas de aprendizado de máquina, foi eficaz para a predição dos parâmetros T, SST, Fe, P e OD. No entanto, para os parâmetros N, DQO e Chla, a metodologia apresenta limitações de generalização dos modelos, ou seja, baixa capacidade preditiva.

APÊNDICE A – COVARIÁVEIS UTILIZADAS PARA IMPLEMENTAÇÃO DOS MODELOS

Sensor MSI/Sentinel-2

COVARIÁVEIS		
Mês	B5/B11	B12/B4
Ano	B5/B12	B12/B5
Código	B6/B2	B12/B6
P_clima	B6/B3	B12/B7
Amostragem	B6/B4	B12/B8
B2	B6/B5	B12/B8A
B3	B6/B7	B12/B11
B4	B6/B8	Aweish
B5	B6/B8A	Wri
B6	B6/B11	SCI
B7	B6/B12	SAVI
B8	B7/B2	FE_SI
B8A	B7/B3	FE_OX
B11	B7/B4	FE2
B12	B7/B5	FE3
B2/B3	B7/B6	NDTI
B2/B4	B7/B8	NDTI
B2/B5	B7/B8A	NDMI
B2/B6	B7/B11	NDMI2
B2/B7	B7/B12	Iron
B2/B8	B8/B2	Clay
B2/B8A	B8/B3	Toming
B2/B11	B8/B4	NDVI
B2/B12	B8/B5	NDWI
B3/B2	B8/B6	HA
B3/B4	B8/B7	Irededge
B3/B5	B8/B8A	GLI
B3/B6	B8/B11	ARVI
B3/B7	B8/B12	NPQI
B3/B8	B8A/B2	GNDVI
B3/B8A	B8A/B3	NDVIVIS
B3/B11	B8A/B4	AWEI
B3/B12	B8A/B5	ITBDN
B4/B2	B8A/B6	Rompimento
B4/B3	B8A/B7	Chuva01
B4/B5	B8A/B8	Chuva02
B4/B6	B8A/B11	Chuva03
B4/B7	B8A/B12	Chuva04
B4/B8	B11/B2	Chuva05
B4/B8A	B11/B3	Chuva06
B4/B11	B11/B4	Chuva07
B4/B12	B11/B5	Chuva08
B5/B2	B11/B6	Chuva09
B5/B3	B11/B7	Chuva10
B5/B4	B11/B8	Chuva11
B5/B6	B11/B8A	Chuva12
B5/B7	B11/B12	Chuva13
B5/B8	B12/B2	Chuva14
B5/B8A	B12/B3	Chuva15

Sensor PlanetScope

COVARIÁVEIS		
Amostragem	B4/B1	Rompimento
Mês	B4/B2	Chuva01
Ano	B4/B3	Chuva02
P_clima	SAVI	Chuva03
Código	NDTI	Chuva04
B1	NDMI2	Chuva05
B2	Iron	Chuva06
B3	Toming	Chuva07
B4	NDVI	Chuva08
B1/B2	NDWI	Chuva09
B1/B3	HA	Chuva10
B1/B4	Irededge	Chuva11
B2/B1	GLI	Chuva12
B2/B3	ARVI	Chuva13
B2/B4	NPQI	Chuva14
B3/B1	GNDVI	Chuva15
B3/B2	NDVIVIS	
B3/B4	ITBDN	