

MODELOS DE APRENDIZADO DE MÁQUINA APLICADOS A PREDIÇÃO DE PARÂMETROS DA QUALIDADE DA ÁGUA NA BACIA DO RIO PARAÓPEBA

Rafael Luís Silva Dias¹; Ricardo Santos Silva Amorim²; Demetrius David da Silva³; Elpidio Inácio Fernandes-Filho⁴; Gustavo Vieira Veloso⁵ & Ronam Henrique Fonseca Macedo⁶

Palavras-Chave – PlanetScope, Águas continentais, Eventos extremos.

INTRODUÇÃO

O monitoramento da qualidade da água (QA) é essencial para conservação dos recursos hídricos e detecção de alterações em seus parâmetros. Métodos tradicionais, baseados em amostragens pontuais, apresentam limitações espaciais e temporais, principalmente durante eventos extremos como rompimentos de barragens (Arango e Nairn, 2020).

O sensoriamento remoto surge como ferramenta complementar para ampliar a cobertura e frequência de observações (Ogashawara et al., 2017). Contudo, limitações nas resoluções espectral, espacial, temporal e radiométrica dos sensores ainda impõem desafios para estimativas de parâmetros não opticamente ativos (Tian et al., 2023).

Diante disso, este estudo desenvolveu uma metodologia para predição de parâmetros de QA, opticamente ativos e não ativos, em ambientes lóticos e lênticos, utilizando imagens de satélite combinadas com técnicas de aprendizado de máquina.

METODOLOGIA

A área de estudo compreende a bacia do rio Paraopeba e o reservatório de Três Marias no estado de Minas Gerais, impactada pelo rompimento da barragem de Brumadinho em 2019. Foram utilizados dados de 24 estações de monitoramento entre 2016 e 2023, para oito parâmetros de qualidade da água: turbidez, sólidos suspensos totais (SST), clorofila-a (chl-a), fósforo (P), nitrogênio (N), demanda química de oxigênio (DQO), oxigênio dissolvido (OD) e ferro (Fe).

Visando identificar o produto de sensoriamento remoto mais adequado para a predição dos parâmetros de QA, a modelagem foi conduzida com três bases de dados: (i) imagens S2; (ii) imagens PS; e (iii) imagens PS normalizadas, conforme metodologia de Dias et al. (2024).

Como covariáveis preditores foram utilizadas: bandas espectrais, razões de banda, índices espectrais, dados acumulados de precipitação (2 a 15 dias antes da coleta) e informações temporais (pré/pós-rompimento da barragem e período hidrológico).

Foram utilizados quatro algoritmos de aprendizado de máquina: Random Forest, k-Nearest Neighbors, Support Vector Machines com Kernel RBF e Cubist. Inicialmente realizou-se a seleção de variáveis, em seguida para garantir estimativas confiáveis, a base de dados foi subdividida em dois conjuntos: 75% para treinamento e 25% para teste. Essa subdivisão obedeceu a dois critérios: i) proporção equilibrada de dados pré e pós-rompimento da barragem; e ii) distribuição proporcional entre os períodos chuvoso e seco (75/25% em ambos). O desempenho foi avaliado por RMSE, MAE, CCC, R² e RMSE e MAE para o modelo nulo (NULL).

1) Eng. Ambiental, Doutorando em Eng. Agrícola, Depto. de Eng. Agrícola, UFV, Viçosa – MG, rafael.luis@ufv.br

2) Eng. Agrônomo, Prof. Associado. Doutor, Depto. de Eng. Agrícola, UFV, Viçosa – MG, rsamorim@ufv.br

3) Eng. Agrônomo, Prof. Titular. Doutor, Depto. de Eng. Agrícola, UFV, Viçosa – MG, demetrius@ufv.br

4) Eng. Agrônomo, Prof. Adjunto. Doutor, Depto. de Solos, UFV, Viçosa – MG, elpidio@ufv.br

5) Eng. Agrônomo, Doutor, Depto. de Solos, UFV, Viçosa – MG, gustavo.veloso@ufv.br

6) Eng. de Agrimensura e Cartográfica, Depto. de Eng. Civil, UFV, Viçosa – MG, ronam.macedo@ufv.br

RESULTADOS

Com base nos resultados das estatísticas dos modelos de aprendizado de máquina, observa-se que os modelos desenvolvidos com base nos dados das imagens PS normalizadas apresentaram os melhores desempenhos 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, com valores de MAE e RMSE inferiores aos limites estabelecidos pelos modelos nulos. Quanto aos parâmetros DQO, N e Chl-a, embora tenham demonstrado valores superiores aos dos outros dois bancos de dados, seus valores de CCC permaneceram abaixo de 0,50, variando entre 0,45 e 0,26; para o R^2 , os valores 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. Esse resultado decorre das características do sensor PS, como sua alta resolução espacial (3.7m), que permite a detecção de feições em escala mais fina; e à sua resolução temporal diária que permitem uma melhor caracterização da variabilidade aquática. O fato das imagens normalizadas apresentarem melhor desempenho que as brutas reforçam a importância do processo de normalização radiométrica, conforme proposto por Dias et al. (2024). Esses resultados sugerem que o PS oferece vantagens expressiva para a modelagem de parâmetros de QA, particularmente em sistemas aquáticos complexos onde a variabilidade espacial e temporal é crítica.

CONCLUSÕES

Com base nos resultados, concluímos que a metodologia desenvolvida para a predição dos parâmetros de QA, utilizando dados de sensoriamento remoto orbital combinados com técnicas de aprendizado de máquina, mostrou-se adequada para estimativa dos parâmetros T, SST, Fe, P e OD. Entretanto, para os parâmetros N, DQO e Chl-a, ainda há limitação na generalização dos modelos, requerendo estudo de outras estratégias de modelagem.

Os modelos desenvolvidos com dados do sensor PS normalizado apresentaram o melhor desempenho na estimativa dos parâmetros de QA. Os algoritmos baseados em árvores, especialmente Random Forest (RF) e Cubist, foram mais robustos em comparação ao KNN e SVM-RBF.

REFERÊNCIAS

- 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.
- DIAS, R. L. S. et al. Relative Radiometric Normalization for the PlanetScope Nanosatellite Constellation Based on Sentinel-2 Images. **Remote Sensing**, v. 16, n. 21, p. 4047, 2024.
- 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.
- 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.

AGRADECIMENTOS

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Código Financeiro 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG (bolsa APQ-01957-22). Agradecemos a Universidade Federal de Viçosa pelo apoio institucional e ao Instituto Mineiro de Gestão das Águas – IGAM.