

**COMPARAÇÃO ENTRE
MODELOS PREDITIVOS DE
MACHINE LEARNING PARA
REGIONALIZAÇÃO DE
VAZÃO NA BACIA
HIDROGRÁFICA DO RIO
IVAÍ**

**COMPARISON BETWEEN PREDICTIVE MACHINE LEARNING MODELS FOR
STREAMFLOW REGIONALIZATION IN THE IVAÍ RIVER WATERSHED**

Ciências Exatas e da Terra, Engenharias • 29/06/2026

REGISTRO DOI: [10.70773/revistatopicos/782640352](https://doi.org/10.70773/revistatopicos/782640352)

Andrew Magri Martins¹

Eudes José Arantes²

Cristiane Kreutz

Natália Rezende Pereira

Camila Silva de Melo

RESUMO

A regionalização de vazões é fundamental para a gestão de recursos hídricos em bacias com baixa densidade de estações hidrométricas. Este estudo comparou seis modelos de Aprendizado de Máquina (Machine Learning – ML) para a predição de vazões na bacia hidrográfica do rio Ivaí, no estado do Paraná. Foram utilizados dados de 17 estações fluviométricas e 56 estações pluviométricas para o período de setembro de 1985 a julho de 2000, totalizando 16.386 dias após o pré-processamento. Os modelos avaliados foram: LightGBM, Extra Trees, Regressão Linear, XGBoost, Gradient Boosting e Random Forest. O desempenho foi aferido pelas métricas NSE (Eficiência de Nash–Sutcliffe), R^2 , MAE (Erro Médio Absoluto) e PBIAS (Viés Percentual). Os resultados evidenciaram que o LightGBM obteve o melhor desempenho pelo critério NSE (0,6895), seguido pelo Extra Trees (0,6880). Em relação ao MAE, o Extra Trees apresentou o menor erro absoluto (5,69 m³/s). Conclui-se que o LightGBM é o modelo mais adequado para a regionalização de vazões na bacia do rio Ivaí, podendo subsidiar tecnicamente os comitês de bacia e a gestão hídrica regional.

Palavras-chave: Aprendizado de Máquina; Regionalização de vazões; Bacia hidrográfica do rio Ivaí; LightGBM; Recursos hídricos.

ABSTRACT

Streamflow regionalization is essential for water resources management in basins with low hydrometric station density. This study compared six Machine Learning (ML) models for streamflow prediction in the Ivaí River watershed, in the state of Paraná, Brazil. Data from 17 streamflow stations and 56 rainfall stations were used for the period from September 1985 to July 2000, totaling 16,386 days after preprocessing. The evaluated models were: LightGBM, Extra Trees, Linear Regression, XGBoost, Gradient Boosting and Random

Forest. Performance was assessed using the NSE (Nash–Sutcliffe Efficiency), R^2 , MAE (Mean Absolute Error) and PBIAS (Percent Bias) metrics. Results showed that LightGBM achieved the best performance by the NSE criterion (0.6895), followed by Extra Trees (0.6880). Regarding MAE, Extra Trees presented the lowest absolute error (5.69 m³/s). It is concluded that LightGBM is the most suitable model for streamflow regionalization in the Ivaí River basin, and can technically support basin committees and regional water management.

Keywords: Machine Learning; Streamflow regionalization; Ivaí River watershed; LightGBM; Water resources.

1. INTRODUÇÃO

A gestão de recursos hídricos no Brasil era realizada, historicamente, a partir do paradigma da estacionariedade hidrológica, sendo possível em razão da previsibilidade das variáveis hidrológicas como precipitação e vazão. Esses parâmetros operavam dentro de uma sazonalidade definida e séries históricas estáveis; de modo que, nesse cenário, o planejamento hídrico subentendia a perenidade e inesgotabilidade do recurso. No entanto, em razão do agravamento dos efeitos das mudanças climáticas, tem-se perdido de vista essa estacionariedade, dando lugar a alterações nos padrões de comportamento meteorológico.

No Paraná, as estiagens foram alarmantes e obrigaram o poder executivo a declarar situação de emergência por meio do Decreto Estadual nº 7.258/2024. A baixa umidade relativa do ar, aliada às altas temperaturas, provocou cerca de 10.000 incêndios em todo o estado, segundo o Corpo de Bombeiros (Ministério Público do Paraná, 2024). Pesquisa divulgada pela Agência Nacional das Águas (ANA) indica

que as principais bacias hidrográficas do país podem ter uma redução da disponibilidade hídrica de até 40% até o ano de 2040, afetando setores agropecuários, energéticos e de abastecimento humano (CNN, 2024).

Diante da ruptura dos padrões históricos, a previsibilidade do comportamento hidrológico em bacias hidrográficas tornou-se um desafio central para a gestão de recursos hídricos. Devido ao intenso grau de estocasticidade inerente aos dados hidrológicos atuais, métodos com base em modelos estatísticos lineares e modelos conceituais de base física encontram dificuldades para captar as relações não lineares complexas. A alternativa que apresenta promissora evolução é o Aprendizado de Máquina (Machine Learning – ML), cujos algoritmos são capazes de organizar e sistematizar grande volume de dados heterogêneos e fornecer relações paramétricas complexas.

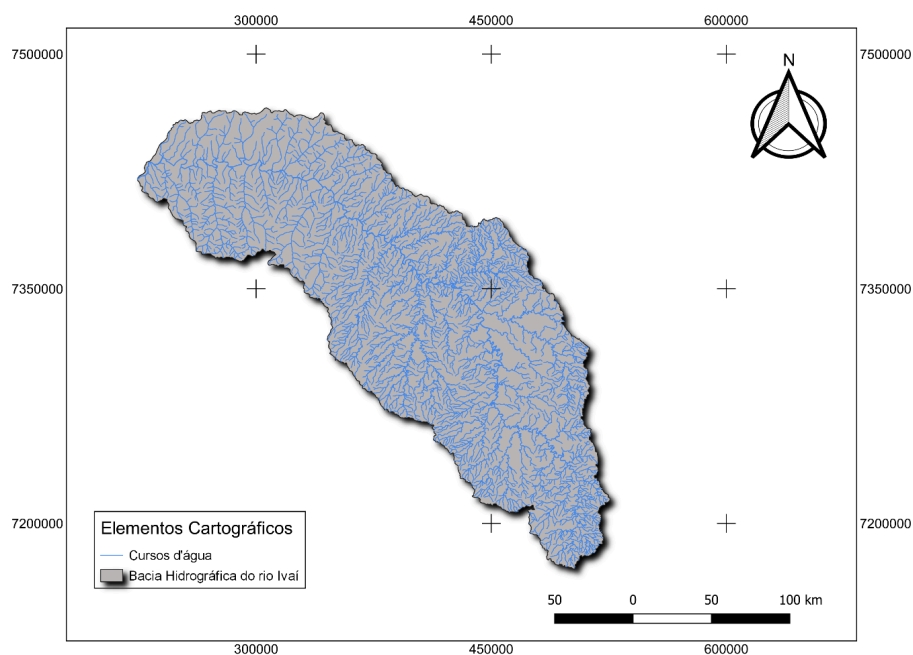
A bacia hidrográfica do rio Ivaí é a segunda maior do estado do Paraná, abrangendo 36.540 km² e 105 municípios. Seus recursos hídricos são utilizados para fins agropecuários, industriais e abastecimento público, o que torna a qualidade e a disponibilidade de dados hidrológicos de importância estratégica. No entanto, há uma lacuna na análise comparativa de modelos de ML para predição de vazão nessa bacia. Desta forma, o presente estudo tem como objetivo comparar modelos de Aprendizado de Máquina para a predição de vazões na bacia hidrográfica do rio Ivaí, visando propor o modelo de maior desempenho como ferramenta de suporte ao planejamento hídrico regional, em consonância com a Política Nacional de Recursos Hídricos (Lei Federal nº 9.433/97).

2. REVISÃO DE LITERATURA

2.1. Bacia Hidrográfica do Rio Ivaí

A bacia hidrográfica do rio Ivaí é totalmente situada no estado do Paraná, sendo a segunda maior do estado em extensão (Figura 1). A área de contribuição da bacia é de 36.540 km². O rio Ivaí é formado pela confluência do rio dos Patos e do rio São João, no município de Prudentópolis, a uma altitude de 1.120 metros, percorrendo cerca de 680 km até sua foz, situada entre os municípios de Icaraíma e Querência do Norte, no rio Paraná. A vazão média do rio é de 689 m³ s⁻¹ (Lellis, Stevaux e Nóbrega, 2024).

Figura 1. Bacia hidrográfica do rio Ivaí e sua rede hidrográfica.

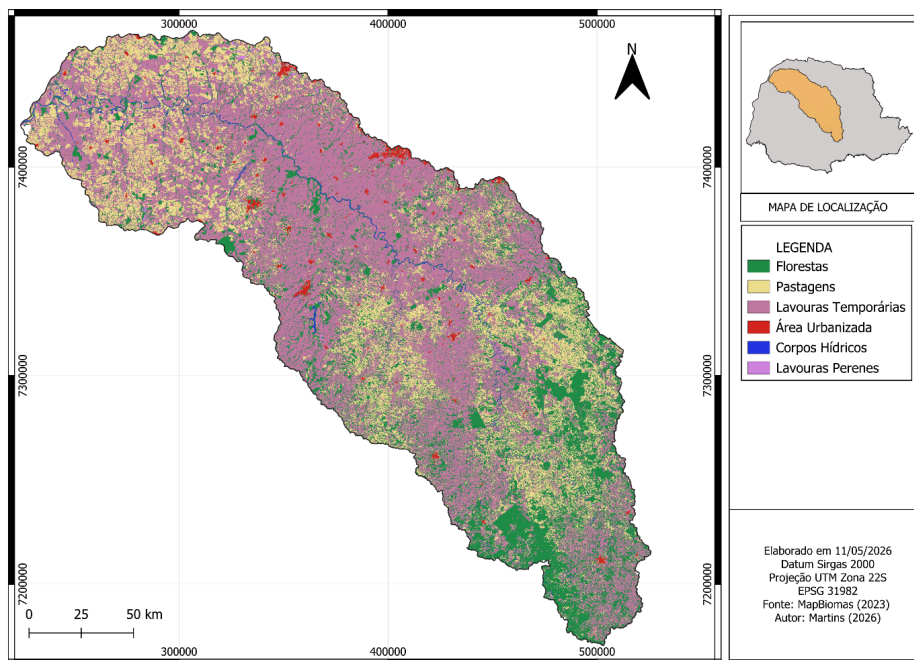


Fonte: Autoria Própria (2025).

A bacia apresenta grande variabilidade climática, com predomínio do clima subtropical com verão quente (Cfa) nos cursos médio e inferior e clima temperado (Cfb) no curso superior, segundo a classificação de Köppen. Os volumes de precipitação anuais giram em torno de 1.500 mm, caracterizando um clima úmido e sem estação seca bem definida (Andrade, 2003; Douhi, 2013). O uso e cobertura do solo é dominado por lavouras temporárias (80,90%),

com destaque para soja, milho e trigo, refletindo uma paisagem fortemente antropizada (MapBiomias, 2023) (Figura 2).

Figura 2. Uso e cobertura do solo da bacia hidrográfica do rio Ivaí.



Fonte: MapBiomias (2023).

2.2. Regionalização de Vazões e Machine Learning

Tucci (2017) define a regionalização de vazão como a transferência de informações de um local para outro dentro de uma área com comportamento hidrológico semelhante, sendo utilizada quando é necessário obter informações hidrológicas de locais com dados escassos ou ausentes. Os métodos tradicionais, contudo, apresentam limitações para lidar com as não linearidades dos sistemas hidrológicos modernos.

O uso de ML na hidrologia tem avançado consideravelmente. O trabalho pioneiro de Hsu, Gupta e Sorooshian (1995) iniciou a aplicação de Redes Neurais Artificiais (RNAs) na modelagem chuva-vazão. Posteriormente, surgiram métodos como o Support Vector Machine (Liong; Sivapragasam, 2002), Random Forest (Shortridge et

al., 2016), Gradient Boosting (Erdal; Karakurt, 2013) e suas variantes XGBoost e LightGBM. O Extra Trees foi introduzido por Geurts, Ernst e Wehenkel (2006) e aplicado à predição de vazões por Galelli e Castelletti (2013). Estudos recentes, como os de Szczepanek (2022), Kedam et al. (2024) e Malik et al. (2025), compararam esses métodos, evidenciando vantagens do LightGBM quanto ao desempenho preditivo e à interpretabilidade dos resultados.

3. METODOLOGIA

3.1. Dados e Área de Estudo

Os dados hidrológicos utilizados foram obtidos no portal *Hidroweb*, mantido pela Agência Nacional das Águas (ANA). A seleção das estações fluviométricas seguiu os critérios de consistência propostos por Tucci (2002): mínimo de 15 anos de dados contínuos e disponibilidade de curva-chave. Foram selecionadas 17 estações fluviométricas internas à bacia do rio Ivaí. Para as estações pluviométricas, adotou-se um raio de 20 km em torno de cada estação fluviométrica, resultando na seleção de 56 estações pluviométricas, dentre as 113 existentes na bacia.

A base de dados abrangeu o período de 21 de setembro de 1985 a 15 de julho de 2000. Após o pré-processamento, a base final contou com 16.386 dias, dos quais 80% (13.108 dias) foram destinados ao treinamento dos modelos e 20% (3.278 dias) ao conjunto de teste.

3.2. Pré-processamento e Engenharia de Atributos

O tratamento dos dados foi realizado em Python (v. 3.12), com uso do ambiente Google Colab e das bibliotecas Pandas, NumPy e Scikit-Learn. A integração dos dados de chuva e vazão foi feita por meio de

inner join, garantindo simultaneidade entre as variáveis. Os seguintes atributos foram gerados na etapa de engenharia de variáveis: (a) defasagens temporais (*lags*) de precipitação dos 7 dias anteriores; (b) média móvel de precipitação acumulada nos 7 dias anteriores (MA7); e (c) variáveis de sazonalidade (mês e dia do ano).

3.3. Modelos de Machine Learning

Foram comparados seis modelos de ML: Gradient Boosting (GB), XGBoost, LightGBM, Extremely Randomized Trees (Extra Trees), Random Forest (RF) e Regressão Linear (RL). Todos os modelos foram implementados via Scikit-Learn. A otimização de hiperparâmetros foi conduzida por meio de *Grid Search* com validação cruzada.

O LightGBM (Ke et al., 2017) é uma implementação otimizada do GB que utiliza a técnica GOSS (Gradient-based One-Side Sampling), priorizando instâncias com maiores erros residuais, e crescimento de árvore por folha (Leaf-Wise), resultando em maior eficiência computacional. O XGBoost aplica regularização L1 e L2 à função de perda, protegendo contra overfitting (Chen; Guestrin, 2016). O Extra Trees distingue-se pela randomização extrema dos pontos de corte e uso do conjunto completo de treinamento (Geurts, Ernst e Wehenkel, 2006). O Random Forest baseia-se no Bagging (Bootstrap Aggregating), com agregação dos resultados das árvores por média aritmética (Breiman, 2001).

3.4. Métricas de Desempenho

O desempenho dos modelos foi avaliado pelas seguintes métricas, calculadas exclusivamente sobre o conjunto de teste:

Eficiência de Nash–Sutcliffe (NSE): varia de $-\infty$ a 1, sendo que valores superiores a 0,75 são indicadores de alta performance em modelos hidrológicos.

Coeficiente de Determinação (R^2): mede o grau em que as variáveis independentes explicam a variância na variável dependente (0 a 1).

Erro Médio Absoluto (MAE): média das diferenças absolutas entre valores preditos e observados, expresso em m^3/s .

Viés Percentual (PBIAS): representa a tendência média do modelo em superestimar ou subestimar os dados observados.

4. RESULTADOS E DISCUSSÕES

4.1. Caracterização da Base de Dados

A planilha unificada gerada após a integração das bases pluviométricas e fluviométricas resultou em 73 colunas e 34.524 linhas, exportada em formato CSV. Após a engenharia de atributos (lags, médias móveis) e a remoção de valores nulos, a base final compreendeu 16.386 registros diários correspondentes ao período de setembro de 1985 a julho de 2000. Desse total, 80% dos registros (13.108 dias) foram destinados ao treinamento e 20% (3.278 dias) ao teste.

4.2. Desempenho dos Modelos

Os resultados das métricas de desempenho obtidas pelos seis modelos no conjunto de teste estão apresentados na Tabela 1.

Tabela 1. Desempenho dos modelos de Machine Learning na predição de vazões na bacia do rio Ivaí

Modelo	NSE	PBIAS (%)	R²	MAE (m³/s)
LightGBM	0,6895	0,49	0,6895	6,25
Extra Trees	0,6880	0,91	0,6880	5,69
Regressão Linear	0,6545	1,99	0,6545	6,82
XGBoost	0,6050	-0,94	0,6050	6,29
Gradient Boosting	0,5970	0,71	0,5970	6,33
Random Forest	0,5843	-0,34	0,5843	6,17

Fonte: Autoria própria (2025).

Em relação ao NSE, observa-se a proeminência do LightGBM (0,6895) sobre os demais modelos, seguido pelo Extra Trees (0,6880). Esses valores são coerentes com os encontrados por Szczepanek (2022), que demonstrou superioridade do LightGBM frente ao XGBoost e CatBoost na predição de vazões na bacia do rio Skawa (Polônia), com NSE de 0,886. Kumar et al. (2023) também identificaram melhor desempenho do LightGBM, com R² de 0,71, valor próximo ao obtido neste estudo. Machado (2025) corroborou esses achados ao identificar o LightGBM como o modelo de maior desempenho na predição do nível d'água do rio Guaíba.

A superioridade do LightGBM pode ser atribuída à sua arquitetura de crescimento Leaf-Wise, que prioriza a redução de erros nas folhas com maior resíduo, capturando melhor a forma do hidrograma e os picos de vazão. Ke et al. (2017) destacam que essa abordagem pode acelerar o treinamento em até 20 vezes em comparação com

métodos tradicionais. Além disso, ao contrário de modelos de Deep Learning, o LightGBM pertence à categoria de Inteligência Artificial Explicável (XAI), permitindo que hidrólogos identifiquem a importância relativa das variáveis preditoras (Szczepek, 2022).

Quanto à métrica MAE, o Extra Trees apresentou o menor erro absoluto (5,69 m³/s), superando o LightGBM (6,25 m³/s). Resultado semelhante foi observado por Hameed et al. (2021), em que o Extra Trees sobressaiu-se em relação ao Random Forest no erro absoluto. Essa vantagem está associada ao fato de o Extra Trees utilizar o conjunto completo de treinamento, sem reamostragem bootstrap, minimizando o viés e garantindo previsões mais estáveis no médio prazo (Geurts, Ernst e Wehenkel, 2006). Autores como Malik et al. (2025) e Kedam et al. (2024) também observaram que o LightGBM, apesar de apresentar o maior NSE, pode ser superado em MAE por outros modelos.

Em relação ao PBIAS, todos os modelos apresentaram valores inferiores a 10%, classificados como "muito bom". O Random Forest obteve o menor PBIAS (-0,34%), provavelmente em razão do processo de Bagging e agregação por média, que funciona como um filtro de estabilidade, produzindo estimativas com tendência de erro quase nula.

Ressalta-se que o NSE e o R² apresentaram os mesmos valores para todos os modelos testados, fenômeno explicado pelo baixo PBIAS alcançado. Dawson e Wilby (2001) esclarecem que o NSE é frequentemente referido por alguns autores como "coeficiente de determinação", pois numericamente equivale ao R² quando o viés é mínimo.

5. CONCLUSÃO

O presente estudo comparou seis modelos de Machine Learning para a predição de vazões na bacia hidrográfica do rio Ivaí, com base em dados hidrológicos de 17 estações fluviométricas e 56 estações pluviométricas.

O LightGBM demonstrou ser o modelo de maior desempenho pelo critério NSE (0,6895), sendo o mais indicado para a regionalização de vazões na bacia do rio Ivaí. O Extra Trees destacou-se pelo menor erro absoluto (MAE = 5,69 m³/s), evidenciando sua adequação para aplicações que demandam precisão diária nas estimativas. Os demais modelos apresentaram desempenho satisfatório, mas inferior ao LightGBM e ao Extra Trees.

Os resultados obtidos estão em linha com a literatura especializada, confirmando a superioridade do LightGBM em tarefas de predição de vazões. Os modelos desenvolvidos podem subsidiar tecnicamente os comitês de bacia e os órgãos gestores de recursos hídricos, contribuindo para o planejamento hídrico regional em consonância com a Política Nacional de Recursos Hídricos (Lei Federal nº 9.433/97) e com os Objetivos de Desenvolvimento Sustentável da ONU (ODS 6, 13 e 14).

Recomenda-se, para estudos futuros, a incorporação de variáveis físicas da bacia (declividade, uso do solo, índice de cobertura vegetal) como atributos preditores, bem como a aplicação de técnicas de explainability (SHAP values) para aprofundar a interpretação física dos modelos.

REFERÊNCIAS BIBLIOGRÁFICAS

AKBULUT, Uğur; CIFCI, Mehmet Akif; ASLAN, Zafer. Hybrid Modeling for Stream Flow Estimation: Integrating Machine Learning and Federated Learning. *Applied Sciences*, v. 13, n. 10203, 2023.

BALDO, M. C. Variabilidade pluviométrica e a dinâmica atmosférica na bacia hidrográfica do rio Ivaí-PR. 2006. Tese (Doutorado em Geografia) – Universidade Estadual Paulista, Presidente Prudente, 2006.

BREIMAN, Leo. Random forests. *Machine learning*, v. 45, n. 1, p. 5–32, 2001.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

DAWSON, C. W.; WILBY, R. L. An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal*, v. 43, p. 47–66, 1998.

DESTEFANI, E. V. Regime hidrológico do rio Ivaí-PR. 2005. 93 f. Dissertação (Mestrado em Geografia) – Departamento de Geografia, Universidade Estadual de Maringá, Maringá, 2005.

ERDAL, Halil Ibrahim; KARAKURT, Onur. Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms. *Journal of Hydrology*, v. 477, p. 119–128, 2013.

GALELLI, S.; CASTELLETTI, A. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. *Hydrology and Earth System Sciences*, v. 17, p. 2669–2684, 2013.

GEURTS, Pierre; ERNST, Damien; WEHENKEL, Louis. Extremely randomized trees. *Machine learning*, v. 63, n. 1, p. 3–42, 2006.

HAMEED, Mohammed Majeed et al. An Extra Tree Regression Model for Discharge Coefficient Prediction. *Mathematical Problems in Engineering*, 2021.

HSU, Kuo-lin; GUPTA, Hoshin Vijai; SOROOSHIAN, Soroosh. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, v. 31, n. 10, p. 2517–2530, 1995.

KE, G. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 2017.

KEDAM, N. et al. River stream flow prediction through advanced machine learning models for enhanced accuracy. *Results in Engineering*, v. 22, p. 102215, 2024.

KUMAR, et al. Streamflow prediction using LightGBM model. 2023.

LELLIS, Isabel Terezinha; STEVAUX, José Cândido; NOBREGA, M. T. Dinâmica espacial da hidrologia da bacia do Rio Ivaí. *Boletim de Geografia*, v. 28, n. 2, p. 41–47, 2010.

LIONG, S. Y.; SIVAPRAGASAM, C. Flood Stage Forecasting With Support Vector Machines. *Journal of the American Water Resources Association*, v. 38, n. 1, p. 173–196, 2002.

MACHADO, Autor. Predição do nível d'água do Rio Guaíba com LightGBM. 2025.

MALIK, M. H.; ANNAYAT, W.; MIR, R. R. Streamflow Prediction in the Jhelum River Basin Using Ensemble Machine Learning Models and Long-Term Hydro-Meteorological Data. 2025.

MAPBIOMAS. Anual Land Use and Land Cover Maps, Coleção 9. 2023. Disponível em: <https://mapbiomas.org>. Acesso em: mai. 2025.

MINISTÉRIO PÚBLICO DO PARANÁ (MPPR). Governador decreta situação de emergência em razão de estiagem. 11 set. 2024. Disponível em: <https://site.mppr.mp.br>. Acesso em: 09 fev. 2025.

MOURA, Denis Anderson Vieira et al. Tendências Espaço-Temporais das Precipitações na Bacia do Rio Ivaí (PR). Geografia (Londrina), v. 29, n. 2, p. 123–142, 2020.

SANTOS, Leonardo José Cordeiro et al. Mapeamento geomorfológico do Estado do Paraná. Revista Brasileira de Geomorfologia, v. 7, n. 2, 2006.

SZCZEPANEK, R. Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost. Hydrology, v. 9, n. 226, 2022.

TUCCI, Carlos E. M. Regionalização de Vazões. Porto Alegre: Associação Brasileira de Recursos Hídricos (ABRH), 2002.

¹ Mestrando em Gestão e Regulação de Recursos Hídricos – ProfGÁgua/UTFPR, Campus Campo Mourão. E-mail: [acesse o artigo original para visualizar o e-mail](#)

² Professor Orientador – UTFPR, Campus Campo Mourão. E-mail:
[acesse o artigo original para visualizar o e-mail](#)