

XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS

PLATAFORMA INTELIGENTE PARA ESTIMATIVA DE VAZÃO EM BACIAS NÃO MONITORADAS

Reinaldo Rios Ossuna Filho ;Paulo Tarso S. de Oliveira¹

Abstract: We present a web platform that transforms scarce data into accessible hydrological insights: through deep learning models, users of any background can obtain reliable streamflow estimates in river basins lacking direct measurements. Building on the historical challenges identified by the PUB initiative, we overcome the limitations of traditional methods and bring academic advances closer to everyday practice, offering a fast and effective tool for water resource planning, flood prevention, and sustainable decision-making. Ultimately, the platform is not merely a repository of models, but a gateway to critical information for managing water resources, mitigating flood and drought risks, and supporting strategic decisions. This work reflects the convergence of academic innovation and real-world applicability, paving the way for a more accessible, data-driven hydrology in Brazil.

Resumo: Apresentamos uma plataforma web pioneira que transforma dados escassos em insights hidrográficos acessíveis: por meio de modelos de aprendizado profundo, usuários de qualquer perfil podem obter estimativas confiáveis de vazão em bacias onde não há medição direta. Partindo dos desafios históricos identificados pelo programa PUB, superamos as limitações dos métodos tradicionais e aproximamos avanços acadêmicos da prática cotidiana, oferecendo uma ferramenta ágil para planejamento de recursos hídricos, prevenção de enchentes e tomada de decisões sustentáveis. Ao final, a plataforma não é apenas um repositório de modelos, mas um portal de acesso imediato a informações críticas para planejar recursos hídricos, mitigar riscos de cheias e secas, e fundamentar decisões estratégicas. Este trabalho reflete a convergência entre inovação acadêmica e aplicabilidade real, abrindo caminho para uma hidrologia mais acessível e data-driven no Brasil.

Palavras-Chave – LSTM; deep learning; PUB

1 INTRODUÇÃO

As medições de vazão em bacias hidrográficas são de fundamental importância para diversas aplicações, tais como a gestão de recursos hídricos, o planejamento energético, a previsão de eventos extremos (enchentes e secas), estudos ambientais e o dimensionamento de estruturas hidráulicas, como vertedouros e aterros (Sivapalan et al., 2003; Parajka et al., 2013). Contudo, muitas regiões ao redor do mundo carecem de redes de monitoramento ou apresentam infraestrutura de medição insuficiente, especialmente em países em desenvolvimento (Hrachowitz et al., 2013).

Visando enfrentar esse desafio, a International Association of Hydrological Sciences (IAHS) lançou, em 2003, o programa Prediction in Ungauged Basins (PUB), voltado à redução das incertezas em previsões hidrológicas em bacias sem monitoramento (Sivapalan et al., 2003). Embora esse esforço tenha gerado avanços relevantes, ainda persistem desafios significativos na previsão de

¹) Afiliação: Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia, Universidade Federal de Mato Grosso do Sul, Campo Grande-MS, reinaldo.ossuna@gmail.com, paulotarsoms@gmail.com

vazões em bacias não monitoradas, o que reforça a necessidade de soluções complementares (Yang et al., 2023; Hrachowitz et al., 2013).

Os modelos hidrológicos podem ser classificados em três categorias principais: modelos físicos, modelos conceituais e modelos baseados em dados. Enquanto os dois primeiros demandam conhecimento detalhado ou simplificações do ciclo hidrológico, os modelos baseados em dados utilizam técnicas de aprendizado de máquina para extrair padrões diretamente a partir de dados históricos. Com a crescente disponibilidade de dados em larga escala e o avanço do poder computacional, observa-se um aumento no interesse pelas abordagens baseadas em dados para a previsão de vazões em bacias não monitoradas.

Dentre os modelos orientados por dados, destacam-se as redes neurais recorrentes (Recurrent Neural Networks – RNNs), com ênfase nas redes Long Short-Term Memory (LSTM), propostas por Hochreiter e Schmidhuber (1997). As LSTMs são projetadas para capturar dependências de longo prazo em séries temporais, superando limitações das RNNs tradicionais, o que as torna particularmente promissoras para modelagem hidrológica, onde há forte influência de eventos pretéritos de precipitação. Estudos recentes demonstram que modelos LSTM podem representar com precisão a dinâmica de sistemas hidrológicos, muitas vezes superando o desempenho de modelos conceituais (Zhang et al., 2018; Yin et al., 2021; Gao et al., 2020; He et al., 2024; Kratzert et al., 2018).

Apesar desses avanços, a aplicação de modelos baseados em aprendizado profundo ainda depende de implementação de técnicas complexas. Frequentemente, tais soluções permanecem restritas ao meio acadêmico, carecendo de interfaces acessíveis a gestores e técnicos não especializados. Isso evidencia uma lacuna entre o desenvolvimento tecnológico e sua aplicabilidade prática em contextos reais. Diante desse cenário, este trabalho pretende conceber, implementar e validar uma plataforma de predição de vazão em bacias não monitoradas baseada em aprendizado profundo.

2 MATERIAL E MÉTODOS

2.1 Delimitação da bacia

A qualidade e a resolução do modelo digital de elevação (DEM) são cruciais nesta etapa, pois todas as informações subsequentes dependem da delimitação precisa da bacia para o cálculo das médias espaciais. Para garantir essa qualidade, utilizou-se o produto NASADEM (NASA JPL, 2020).

A delimitação das bacias foi realizada por meio da biblioteca PySheds, seguindo os procedimentos recomendados por Bartos (2020). Antes da delimitação, o DEM passou por um pré-processamento essencial, no qual foram removidas depressões e regiões planas, garantindo a continuidade do escoamento superficial. Em seguida, foram calculadas as direções de fluxo e os mapas de acúmulo de fluxo.

A partir de um ponto de exutório fornecido, o sistema busca a célula mais próxima no mapa de acúmulo de fluxo cuja contribuição seja superior a 100 células. Esse critério foi adotado conforme o procedimento descrito por Almagro et al. (2021). O ponto ajustado é então utilizado como exutório definitivo para a delimitação da bacia. Além dos contornos da bacia, também foram extraídos o perímetro, a área total para desnormalizar a vazão estimada e a inclinação média.

2.2 Dados Climáticos

Para representar de forma adequada as variáveis meteorológicas utilizadas nos modelos, foi calculada uma média ponderada dos dados dentro dos limites de cada bacia hidrográfica. Esse procedimento visa considerar as variações espaciais das variáveis e garantir uma representação mais fiel das condições climáticas locais.

A precipitação foi obtida a partir do produto CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data), que oferece estimativas quase globais de chuva com resolução espacial de 0,05° e cobertura temporal contínua desde 1981. Esse conjunto de dados integra informações de sensores de satélite com observações de estações pluviométricas, resultando em séries históricas consistentes e de alta resolução espacial. Sua ampla aceitação em estudos hidrológicos se deve, principalmente, à capacidade de suprir lacunas em regiões com escassez de medições terrestres (Funk et al., 2015).

As demais variáveis meteorológicas, temperatura do ar, umidade relativa, radiação solar e velocidade do vento, foram obtidas a partir do conjunto de dados desenvolvido por Xavier et al. (2016), que disponibiliza informações diárias espacialmente distribuídas para todo o território brasileiro no período de 1980 a 2013. Com base nesses dados, foi calculado o índice de aridez, definido como a razão entre a evapotranspiração potencial (PET) e a precipitação total. Essa variável tem papel relevante na caracterização do clima das bacias, especialmente em análises hidrológicas regionais.

Ambos os produtos adotados neste estudo destacam-se pela qualidade das estimativas e pela ampla disponibilidade, sendo utilizados na literatura para aplicações em modelagem hidrológica (Brocca et al., 2014).

2.3 Dados para o treinamento

Modelos de inteligência artificial (IA), especialmente aqueles aplicados em hidrologia, exigem grandes volumes de dados observacionais para treinamento e validação. A disponibilidade e a qualidade desses dados são fatores determinantes para a capacidade dos modelos de realizar previsões precisas e confiáveis sobre o comportamento de sistemas ambientais complexos.

Neste trabalho, foi utilizado o conjunto de dados CABra (Catchments Attributes in Brazil), visando treinar e testar os modelos de IA em diferentes condições hidrometeorológicas e fisiográficas. O CABra é um banco de dados de larga escala que abrange 735 bacias hidrográficas brasileiras e reúne séries temporais longas (30 anos) organizadas em oito grupos de atributos: vazão, água subterrânea, geologia, solo, topografia, clima, uso e cobertura da terra, e grau de antropização (Almagro et al., 2021).

A abrangência espacial e temporal do CABra, aliada à diversidade de variáveis disponíveis, torna esse conjunto especialmente útil para o desenvolvimento de modelos generalizáveis e robustos, além de permitir a avaliação do desempenho em diferentes regiões e regimes climáticos.

2.4 Modelos

Os modelos de predição foram desenvolvidos em Python, utilizando o framework PyTorch (Imambi et al., 2021), que oferece flexibilidade para definir arquiteturas sequenciais e gerenciar dinamicamente o grafo de computação. O processo de treinamento adotou o otimizador Adam, para minimizar a função de perda baseada no Nash–Sutcliffe Efficiency (NSE). Para garantir maior robustez e capacidade de generalização, a busca de hiper parâmetros, como taxa de aprendizado,

número de unidades ocultas, comprimento da sequência de entrada e taxa de dropout, foi automatizada com Optuna (Akiba et al., 2019), permitindo explorar eficientemente o espaço de configurações e selecionar aquelas que maximizam o desempenho nos dados de validação.

2.4.1 LSTM

Para capturar dependências de longo prazo nas séries de vazão, foi empregada a arquitetura LSTM (Long Short-Term Memory), um tipo de rede neural recorrente projetada para preservar informações relevantes por meio de seu estado de célula C_t e de portas de controle. Cada módulo LSTM possui três portões principais, de esquecimento f_t , de entrada i_t e de saída o_t , que regulam o fluxo de informações, permitindo que a rede decida quais valores antigos descartar, quais novos incorporar e o que transmitir adiante (Figura 1).

Durante o treinamento, sequências de precipitação e índices de aridez foram alimentadas na LSTM de forma a aprender padrões temporais associados a eventos extremos de vazão. A modelagem considerou janelas deslizantes de comprimento definido como hiper parâmetro, balanceando a capacidade de memória com o custo computacional. A aplicação de early stopping sobre a perda de validação evitou o overfitting, interrompendo o treinamento quando não houve melhora significativa por um número pré-definido de épocas.

Em resumo, a combinação de PyTorch, Otimização com Optuna e controle de overfitting via early stopping proporcionou configurações de LSTM capazes de produzir previsões de vazão precisas e com boa generalização espacial, atendendo aos objetivos de estimativa em bacias não monitoradas.

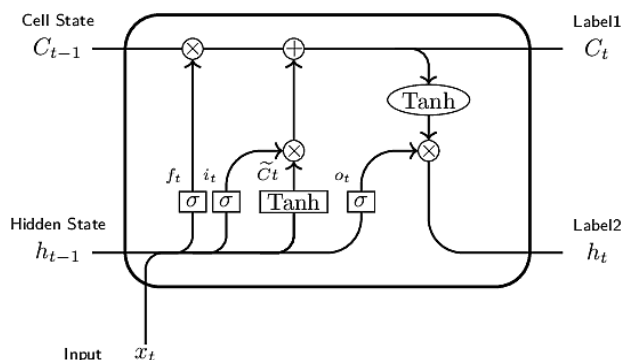


Figura 1: Representação esquemática de um módulo LSTM

2.5 Métricas

A avaliação da qualidade das previsões hidrológicas baseadas em redes neurais fundamenta-se na comparação entre as séries simuladas e os registros observados, buscando quantificar os desvios de forma robusta. Seguindo as recomendações de Moriasi et al. (2007), adota-se o Erro Quadrático Médio (MSE) como medida primária de dispersão, pois penaliza fortemente grandes discrepâncias entre valores previstos e observados. Além disso, Gupta et al. (2009) destacam a importância de decompor o MSE em componentes que esclareçam tendências sistemáticas e variabilidade residual, de modo a oferecer diagnóstico mais refinado do desempenho do modelo.

Para capturar diferentes aspectos da qualidade preditiva, os seguintes indicadores foram calculados:

- Coeficiente de Eficiência de Nash–Sutcliffe (NSE) Mede a habilidade do modelo em reproduzir a variabilidade da série observada, variando de $-\infty$ a 1, sendo valores próximos de 1 indicativos de excelente concordância (Moriasi et al., 2007).
- Índice de Desempenho de Kling–Gupta (KGE) Integra correlação, viés e razão de variabilidade entre séries simulada e observada em um único indicador, oferecendo visão balanceada dos erros de tendência e dispersão (Gupta et al., 2009).

2.6 Descrição do Software

A plataforma web foi desenvolvida segundo uma arquitetura de microsserviços containerizados, garantindo escalabilidade, portabilidade e facilidade de manutenção. Cada componente principal, frontend, backend, workers de processamento e sistema de mensageria, roda em containers Docker, o que assegura reprodutibilidade e simplifica o deploy em diferentes ambientes (Newman, 2021).

O frontend foi implementado em React, permitindo a criação de componentes interativos, reutilizáveis e de carregamento dinâmico, favorecendo a responsividade da interface (Lebedev, 2015). Já o backend utiliza o framework FastAPI, escolhido por sua performance elevada, tipagem estática via Python type hints e suporte nativo a chamadas assíncronas, o que reduz a latência das respostas e maximiza o uso dos recursos do servidor (Ramírez, 2021).

Para orquestrar tarefas pesadas, como inferência dos modelos de aprendizado profundo, adotou-se Celery em conjunto com Redis como broker e backend de resultados. Essa combinação viabiliza o processamento assíncrono de requisições, mantendo a API principal livre para atender novas solicitações de usuários sem bloqueio (Palach, 2014). Dessa forma, a separação clara entre interface, lógica de negócio e processamento assíncrono resulta em uma solução robusta, capaz de lidar com picos de demanda e facilitar futuras expansões.

3 RESULTADOS

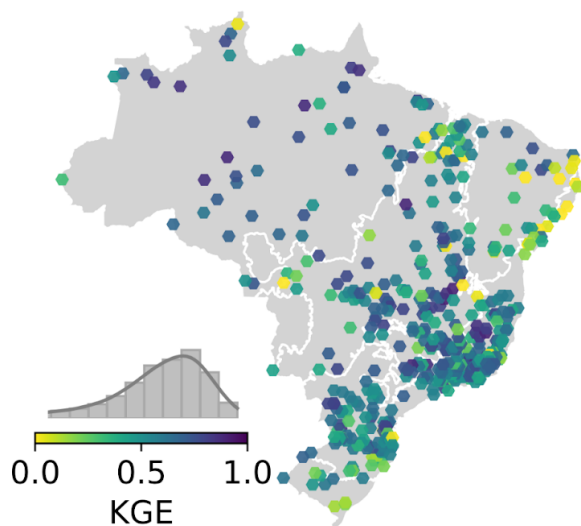
3.1 Modelagem hidrológica

Os resultados do modelo regional, em termos KGE, estão apresentados na Figura 2 para dois cenários: (i) apenas as bacias do conjunto de treino e (ii) apenas as bacias no período de teste. É importante destacar que as bacias do conjunto de teste não participaram do treinamento e, portanto, representam um conjunto verdadeiramente independente para avaliação (representando as bacias não monitoradas). Nesse conjunto, os valores medianos foram de 0,58 para o treino e 0,51 para o teste, com apenas 2% e 6% das bacias apresentando valores menores ou iguais a zero, respectivamente. Além disso, verifica-se que as métricas medianas de KGE são semelhantes entre os dois conjuntos de dados, demonstrando consistência nas simulações e afastando a hipótese de overfitting.

Importante notar que as bacias localizadas no extremo sul e no Nordeste do Brasil tendem a apresentar menores valores de KGE em todas as fases do estudo, sendo as bacias nordestinas, em

especial, as que apresentam maior dificuldade de modelagem. Ainda assim, de forma geral, o modelo demonstrou bom desempenho na simulação das vazões ao longo do território nacional.

a) Treino



b) Teste

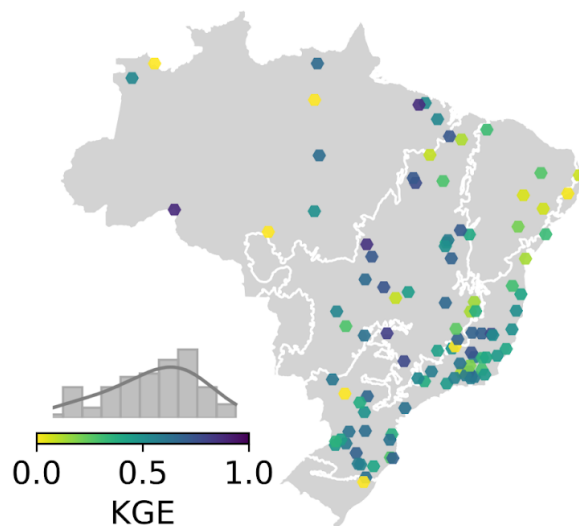


Figura 2: Distribuição espacial da performance do algoritmo LSTM em simular a vazão nas bacias hidrográficas brasileiras, em termos de valores medianos de KGE. O painel à esquerda representa as bacias de treinamento no período de treinamento; e o painel à direita representa as bacias de teste no período de teste

3.2 Plataforma

A plataforma desenvolvida é capaz de realizar automaticamente todas as etapas necessárias para a estimativa de vazão em bacias não monitoradas, incluindo a delimitação da bacia hidrográfica, a extração e o processamento dos dados meteorológicos e a inferência dos modelos de aprendizado profundo, sem a necessidade de intervenção do usuário. Seu funcionamento integrado permite que usuários obtenham estimativas de vazão de forma rápida e prática, a partir de modelos previamente treinados e validados, oferecendo uma solução acessível e eficiente para apoiar estudos hidrológicos e a gestão de recursos hídricos.

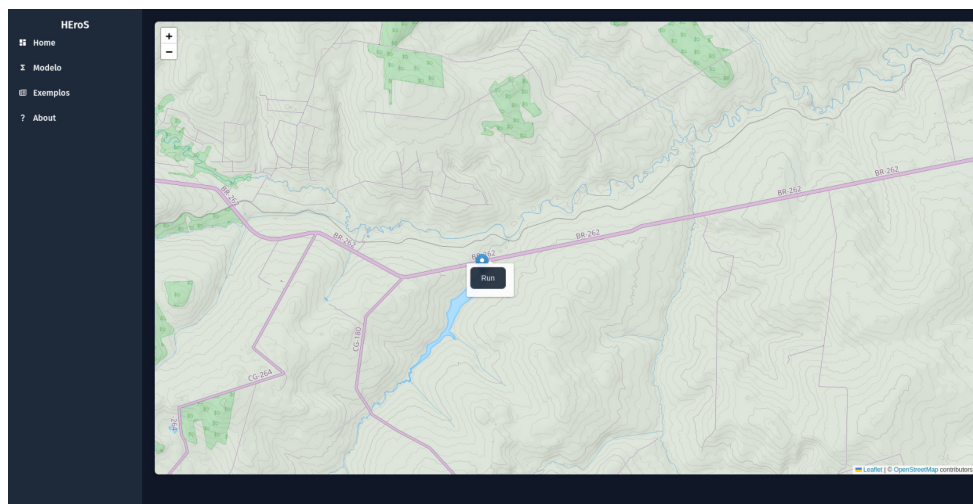


Figura 3: Usuário selecionou um ponto de exutório

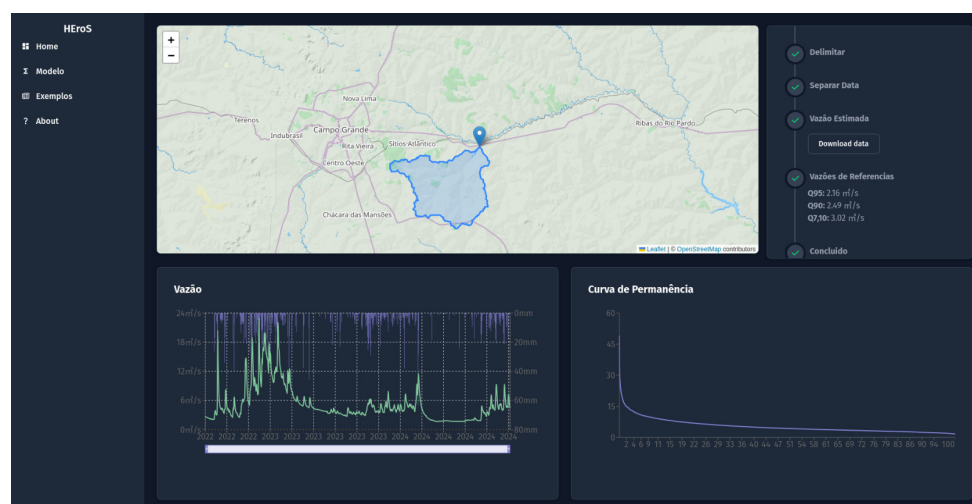


Figura 4: Resultado após simulação

4 CONCLUSÕES

Neste trabalho, exploramos a aplicação de modelos LSTM no contexto da previsão de vazão em bacias hidrográficas não monitoradas, por meio do desenvolvimento de uma plataforma web acessível e automatizada. A proposta visa democratizar o uso de técnicas de aprendizado profundo, permitindo que usuários com pouca ou nenhuma familiaridade com essas metodologias possam obter estimativas confiáveis de vazão em diferentes regiões do Brasil.

A plataforma foi projetada para executar todo o fluxo de trabalho necessário, desde a delimitação da bacia até a inferência do modelo, de forma autônoma, utilizando modelos previamente treinados e validados. Com isso, busca-se reduzir as barreiras técnicas no uso de inteligência artificial aplicada à hidrologia, ampliando o acesso a ferramentas avançadas de análise e suporte à decisão.

Como trabalho futuro, pretende-se ampliar o conjunto de modelos disponíveis na plataforma, incluindo arquiteturas alternativas que possam melhorar o desempenho, especialmente em regiões onde os resultados ainda apresentam limitações, como no Nordeste e no Sul do país.

5 REFERÊNCIAS

- AKIBA, Takuya *et al.* Optuna: A next-generation hyperparameter optimization framework. *In*: 2019.
- ALMAGRO, André *et al.* CABra: a novel large-sample dataset for Brazilian catchments. **Hydrology and Earth System Sciences**, v. 25, n. 6, p. 3105–3135, 9 jun. 2021.
- BARTOS, Matt. **pysheds: simple and fast watershed delineation in python**. , 2020. Disponível em: <<https://github.com/mdbartos/pysheds>>
- Fedosejev, A. (2015). React. Js Essentials. Packt Publishing Ltd.
- GAO, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-Term Runoff Prediction with GRU and LSTM Networks without Requiring Time Step Optimization during Sample Generation. *Journal of Hydrology*, 589, 125188. <https://doi.org/10.1016/j.jhydrol.2020.125188>
- GUPTA, Hoshin V. *et al.* Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. **Journal of hydrology**, v. 377, n. 1–2, p. 80–91, 2009.
- HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, nov. 1997.
- HE, M., Jiang, S., Ren, L., Cui, H., Qin, T., Du, S., Zhu, Y., Fang, X., & Xu, C.-Y. (2024). Streamflow Prediction in Ungauged Catchments through Use of Catchment Classification and Deep Learning. *Journal of Hydrology*, 639, 131638. <https://doi.org/10.1016/j.jhydrol.2024.131638>
- HRACHOWITZ, Markus *et al.* A decade of predictions in ungauged basins (PUB)—a review. **Hydrological sciences journal**, v. 58, n. 6, p. 1198–1255, 2013.
- IMAMBI, Sagar; PRAKASH, Kolla Bhanu; KANAGACHIDAMBARESAN, GR. PyTorch. **Programming with TensorFlow: solution for edge computing applications**, p. 87–104, 2021.
- KRATZERT, Frederik *et al.* Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. **Water Resources Research**, v. 55, n. 12, p. 11344–11354, dez. 2019.
- MORIASI, Daniel N. *et al.* Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **Transactions of the ASABE**, v. 50, n. 3, p. 885–900, 2007.
- NASA JPL. **NASADEM Merged DEM Global 1 arc second V001**. NASA Land Processes Distributed Active Archive Center, , 2020. Disponível em: <https://lpdaac.usgs.gov/products/nasadem_hgtv001>. Acesso em: 15 jun. 2025
- NEWMAN, S. (2021). Building Microservices: Designing Fine-Grained Systems. " O'Reilly Media, Inc."
- LUBANOVIC, B. (2023). FastAPI. " O'Reilly Media, Inc."
- OLAH, C. Understanding LSTM Networks. <https://colah.github.io/posts/2015-08-Understanding-LSTMs>

PALACH, J. (2014). *Parallel Programming with Python*. Packt Publishing Ltd.

PARAJKA, J. *et al.* Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies. **Hydrology and Earth System Sciences**, v. 17, n. 5, p. 1783–1795, 7 maio 2013.

SIVAPALAN, Murugesu *et al.* IAHS decade on predictions in ungauged basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. **Hydrological sciences journal**, v. 48, n. 6, p. 857–880, 2003.

XAVIER, Alexandre C.; KING, Carey W.; SCANLON, Bridget R. Daily gridded meteorological variables in Brazil (1980–2013). **International Journal of Climatology**, v. 36, n. 6, p. 2644–2659, maio 2016.

YANG, Xue *et al.* Regionalization methods for PUB: a comprehensive review of progress after the PUB decade. **Hydrology Research**, v. 54, n. 7, p. 885–900, 13 jul. 2023.

YIN, Hanlin *et al.* Runoff predictions in ungauged basins using sequence-to-sequence models. **Journal of Hydrology**, v. 603, p. 126975, 1 dez. 2021.

ZHANG, Yikui *et al.* Generalization of an Encoder-Decoder LSTM model for flood prediction in ungauged catchments. **Journal of Hydrology**, v. 614, p. 128577, 1 nov. 2022.