TRANSFERIBILIDADE DE MODELOS DE APRENDIZADO DE MÁQUINA PARA MODELAGEM DE VAZÃO ENTRE SUB-BACIAS DO SEMIÁRIDO: ESTUDO DE CASO NA BACIA ATLÂNTICO LESTE

Taís Fonte Boa de Campos Maia[1]; Bruno Melo Brentan[2] & André Ferreira Rodriges[3]

Palavras-Chave – Modelagem hidrológica, Aprendizado por transferência, Semiárido.

INTRODUÇÃO

O monitoramento hidrológico brasileiro apresenta desigualdade espacial, com limitações de disponibilidade e qualidade dos dados em diversas bacias (Collares et al., 2021), cenário crítico no semiárido, marcado por elevada variabilidade pluviométrica e recorrência de eventos extremos e que abrange cerca de 12% do território nacional (INSA, 2024).

A modelagem de vazões é essencial à gestão hídrica, mas modelos físicos dependem de extensas séries para calibração e validação (Silva et al., 2020; Cea et al., 2010). Como alternativa, redes neurais artificiais, especialmente as redes Long Short-Term Memory (LSTM), têm demonstrado elevado potencial para representar relações não lineares e dependências temporais em séries hidrológicas (Freire, 2012), mas seu desempenho também pode ser limitado pela escassez de dados (Yang *et al.*, 2023; Pan; Yang, 2009). Nesse contexto, o aprendizado por transferência surge como uma abordagem promissora para transferir conhecimento de bacias bem monitoradas para bacias com dados limitados. Estudos recentes demonstram ganhos expressivos de desempenho com essa estratégia, superando tanto modelos físicos quanto modelos treinados localmente em cenários de escassez de dados (Ma *et al.*, 2024; Ouyang *et al.*, 2025).

METODOLOGIA

Para a realização deste estudo, foram avaliadas três abordagens de modelagem: Modelo Fonte, uma LSTM treinada com dados de múltiplas sub-bacias para capturar padrões hidrológicos regionais; Modelo Local, treinado individualmente para cada sub-bacia alvo e utilizado como referência; e Modelo de Transferibilidade, obtido por ajuste fino (fine-tuning) do Modelo Fonte nas sub-bacias alvo. A Figura 1 apresenta a área de estudo, as sub-bacias selecionadas e sua classificação como fonte ou alvo.

Importante destacar, que as sub-bacias consideradas neste estudo correspondem às áreas de drenagem das estações fluviométricas selecionadas, delimitadas a partir de suas respectivas áreas contribuintes disponibilizadas pela base de dados CAMELS-BR. Para fins de identificação e avaliação dos resultados, cada bacia foi denominada pelo código da sua respectiva estação fluviométrica.

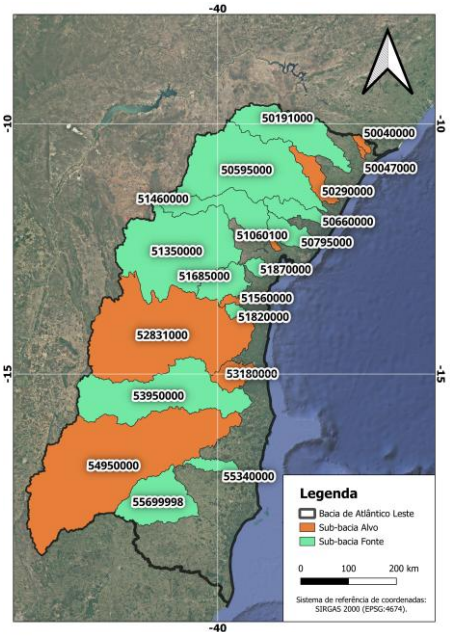


Figura 1. Sub-bacias selecionadas.

1) SMARH - Universidade Federal de Minas Gerais (UFMG), Avenida Presidente Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG. 31270-901, taisfb10@ufmg.br
2) SMARH - Universidade Federal de Minas Gerais (UFMG), Avenida Presidente Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG. 31270-901, brentan@ehr.ufmg.br
3) SMARH - Universidade Federal de Minas Gerais (UFMG), Avenida Presidente Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG. 31270-901, afrodrigues@ehr.ufmg.br

Como entradas, utilizaram-se variáveis climáticas diárias (precipitação, umidade do solo, evapotranspiração e temperatura média) e variáveis derivadas, incluindo índice de precipitação antecedente e precipitação acumulada em 15 dias. A vazão foi adotada como variável alvo. Os modelos empregaram janela temporal de 365 dias, função de perda Huber, otimizador Adam e técnicas de regularização. A arquitetura foi definida por busca em grade, avaliando diferentes combinações de neurônios, camadas e tamanho de batch, selecionadas a partir de um índice combinado composto por NSE, LogNSE e KGE.

No modelo de transferibilidade, foram mantidos os pesos e a arquitetura do Modelo Fonte, ajustando-se apenas a última camada LSTM e a camada totalmente conectada, enquanto as demais permaneceram congeladas. Foram avaliados cinco cenários de treinamento (1, 2, 5 e 10 anos, além da série completa) para determinar a quantidade mínima de dados necessária para uma transferência eficaz. Para todos os modelos, os quatro últimos anos da série histórica foram reservados para teste, sem participação nas etapas de treinamento e validação.

RESULTADOS

Modelo Alvo

Para a apresentação e discussão dos resultados, adotou-se o critério de escolha da bacia que apresentou melhor e da bacia que apresentou pior NSE para a fase de treino (Tabela 1). Para enriquecer a discussão, a Tabela 2 apresenta as características destas mesmas bacias.

Tabela 1 – Métricas individuais – Modelo Local.

Nome	Código	Período	NSE	logNSE	KGE	Correlation	PBIAS
Santa Luzia	51870000	Validação	0.65	0.69	0.75	0.81	6.13
		Teste	0.66	0.67	0.70	0.82	9.68
Itapebi	54950000	Validação	0.45	0.51	0.65	0.69	-5.81
		Teste	-0.68	-0.74	-0.02	0.59	83.82

Tabela 2 – Características das Bacias Alvo.

Código	Q médio (m^3/s)	Q mín. (m^3/s)	Q máx. (m^3/s)	Razão de esc.	Elast. da vazão	Inclin. da curva	Índice de esc. base	AD (km^2)	Aridez
51870000	10.53	0.38	91.40	0.38	0.78	1.64	0.75	832	1.02
54950000	329.58	0.00	6480.55	0.16	1.24	2.32	-	68283	1.13

As bacias Santa Luzia e Itapebi apresentam comportamentos hidrológicos bastante distintos, refletindo diferenças marcantes em escala e resposta hidrológica, o que pode ser observado através das métricas que cada modelo local obteve. A bacia 51870000, com área de drenagem menor, portanto, possui menores magnitudes de vazão, além disso ela não apresenta vazão mínima igual a zero. Suas características incluem os maiores valores de razão de escoamento e um elevado valor de índice de escoamento de base entre as bacias avaliadas, indicando uma maior conversão da precipitação em vazão e uma contribuição significativa do escoamento subterrâneo para a manutenção das vazões ao longo do tempo.

Já a bacia 54950000 apresenta uma variabilidade maior, com valores mínimos nulos e máximos que atingem 6480,55 m^3/s, fato bem representado pelo maior valor da inclinação da curva. Sua menor razão de escoamento (0,16) sugere uma menor parcela da precipitação convertida em escoamento superficial quando comparada à bacia 51870000. Além disso, o maior valor de elasticidade da vazão indica uma resposta mais sensível às variações de precipitação. Apesar dessas diferenças hidrológicas, ambas as bacias apresentam índices de aridez próximos, apontando condições climáticas semelhantes.

Acredita-se que a principal diferença nos resultados esteja relacionada à magnitude das vazões e à presença de valores nulos. A bacia Santa Luzia apresentou a menor vazão máxima entre todas as analisadas, enquanto a bacia Itapebi apresentou a maior. Além disso, a segunda bacia é caracterizada por uma variabilidade muito mais brusca e uma resposta mais rápida e intensa aos eventos de precipitação. Esse conjunto de fatores torna o comportamento hidrológico mais difícil de ser modelado, especialmente pela presença de vazões iguais a zero e pela ocorrência de picos extremos, o que contribui para a pior performance do modelo nessa bacia em comparação com a primeira.

Modelo Transferibilidade

A Figura 2 apresenta as variações percentuais de cada métrica, para todos os cenários avaliados no período de teste, em comparação com os resultados obtidos pelo modelo local para todas bacias alvo da Bacia Atlantico Leste. Importante ressaltar, que continuou-se apresentando os resultados das mesmas bacias selecionadas no item anterior.

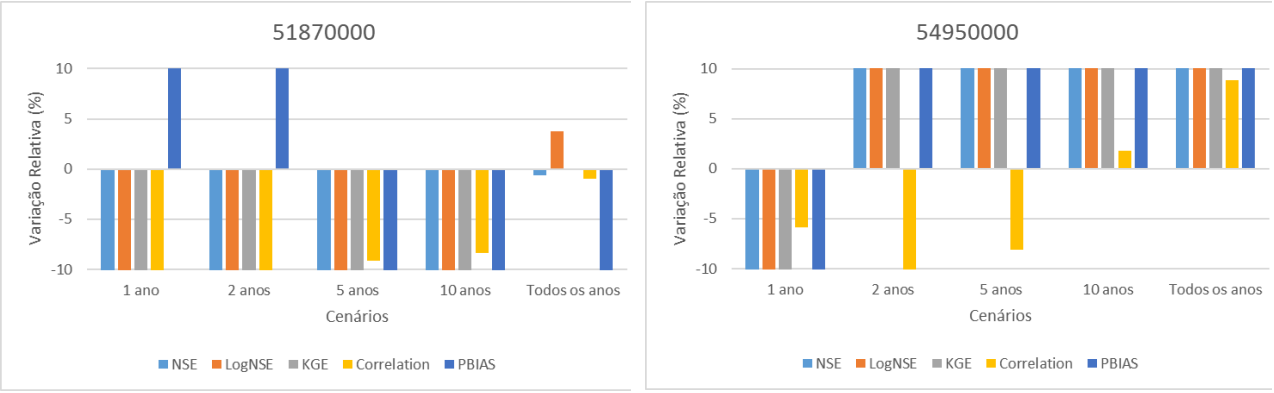


Figura 2. Métricas individuais – Modelo Transferibilidade.

A análise dos resultados indica comportamentos distintos entre as bacias avaliadas. Para a estação 51870000, o cenário de transferibilidade que apresentou o melhor desempenho foi aquele que utilizou todos os anos (Figura 3) de dados disponíveis para o ajuste do modelo. Já para a estação 54950000, os cenários com 2, 5, 10 (Figura 4) e todos os anos disponíveis de dados resultaram em melhorias superiores a

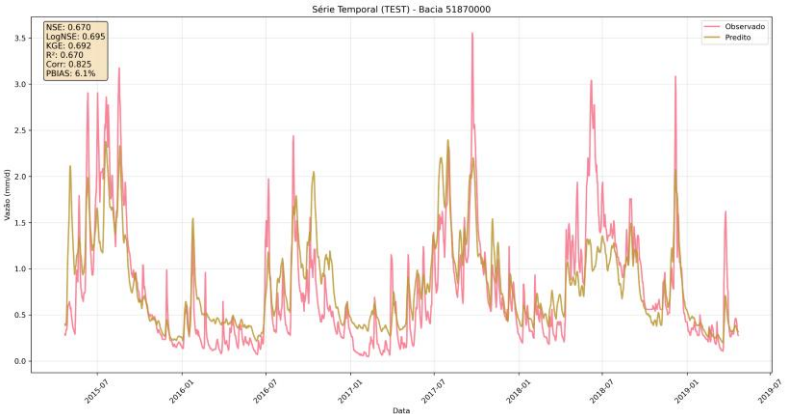


Figura 3 – Série Histórica – Observado x Simulado – 51870000 – Todos os anos.

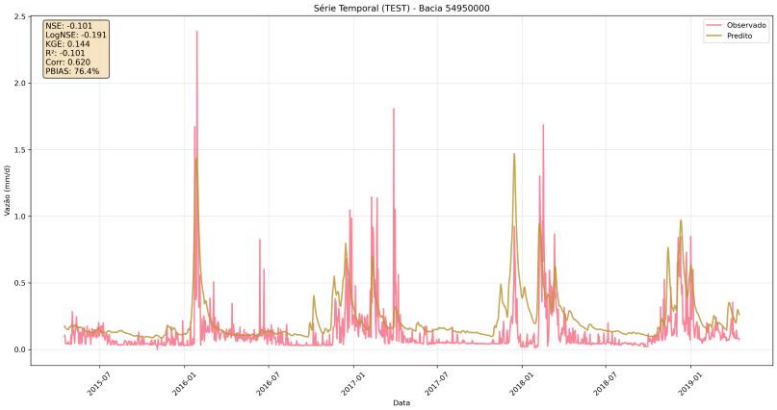


Figura 4 – Série Histórica – Observado x Simulado – 54950000 – 10 anos.

10% nas métricas NSE, LogNSE e KGE quando comparados ao treinamento local.

De forma geral, observa-se que bacias que já apresentam desempenho satisfatório quando modeladas apenas com dados locais tendem a obter ganhos marginais com a aplicação da transferibilidade. Em contrapartida, bacias cujo desempenho inicial é

limitado demonstram maior potencial de benefício.

CONCLUSÕES

Conclui-se que o aprendizado por transferência pode manter ou até melhorar, ainda que de forma limitada, a acurácia das previsões em comparação com modelos locais, especialmente em cenários de escassez de dados. No entanto, percebe-se o desempenho está fortemente condicionado às características hidrológicas das bacias. Observou-se, inicialmente, melhor desempenho em sub-bacias com relações chuva–vazão bem definidas, enquanto resultados inferiores ocorreram em bacias dominadas com maior variabilidade hidrológica, presença de vazões nulas e ocorrência de picos extremos. Durante a elaboração deste manuscrito, a autora utilizou o ChatGPT apenas para revisão linguística, incluindo correções gramaticais e melhorias de clareza textual. A ferramenta não foi utilizada para gerar conteúdo científico, os autores revisaram o manuscrito e assumem total responsabilidade por seu conteúdo.

REFERÊNCIAS

a) Livro

FREIRE, P. K. M. M. (2012). *O uso da transformada Wavelet na previsão de vazão*. Dissertação de Mestrado em Engenharia Urbana e Ambiental.Universidade Federal da Paraíba, João Pessoa/PB, 130.

INSTITUTO NACIONAL DO SEMIÁRIDO – INSA (2024). *Delimitação do Semiárido Brasileiro 2024*. Campina Grande/PB.

b) Artigo em revista

CEA, L.; GARRIDO, M.; PUERTAS, J. (2010). *Experimental validation of two-dimensional depth-averaged models for forecasting rainfall–runoff from precipitation data in urban areas*. Journal of Hydrology, v. 382, n. 1–4, p. 88–102.

COLLARES, E. G.; et al. (2022). *Análise comparativa entre vazão medida e vazão regionalizada em três pares de bacias hidrográficas morfometricamente semelhantes do estado de Minas Gerais*. Revista Mineira de Recursos Hídricos, v. 3, n. 1, p. e022007.

HUBER, P. J. (1964). *Robust estimation of a location parameter*. Annals of Mathematical Statistics, v. 35, n. 1, p. 73–101.

MA, K.; et al. (2024). *Transfer learning framework for streamflow prediction in large-scale transboundary catchments*. Journal of Geographical Sciences, v. 34, p. 963–984.

OUYANG, W.; et al. (2025). *Dive into transfer-learning for daily rainfall-runoff modeling in data-limited basins*. Journal of Hydrology.

PAN, S. J.; YANG, Q. (2009). *A survey on transfer learning*. IEEE Transactions on Knowledge and Data Engineering, v. 22, n. 10, p. 1345–1359.

SILVA, F. B.; et al. (2020). *Pluviometric and fluviometric trends in association with future projections in areas of conflict for water use*. Journal of Environmental Management, v. 271.

YANG, M.; et al. (2023). *A new few-shot learning model for runoff prediction*. Environmental Modelling & Software, v. 162, p. 105659.

c) Artigo em anais de congresso ou simpósio

KINGMA, D. P.; BA, J. (2015). *Adam: A Method for Stochastic Optimization*. Proceedings of the International Conference on Learning Representations (ICLR 2015).

AGRADECIMENTOS

À Universidade Federal de Minas Gerais (UFMG), pela infraestrutura acadêmica e suporte institucional oferecidos para a realização deste estudo.

Ao grupo de pesquisa Hydro-Uai, pelo apoio técnico-científico e pelas contribuições ao longo do desenvolvimento desta pesquisa.