



# Modelagem de vazões sazonais no semiárido brasileiro utilizando inteligência artificial e dados climáticos gradeados

Seasonal streamflow modelling in brazilian semiarid using artificial intelligence and gridded climate data

N. M. S. R. Saturnino<sup>1\*</sup>; I. R. S. Neto<sup>1</sup>; E.H.M. Schneider<sup>1</sup>; R. D. Gonçalves<sup>1,2</sup>

<sup>1</sup>Programa de Pós-Graduação em Recursos Hídricos, Universidade Federal de Sergipe, 49107-230, São Cristóvão-Sergipe, Brasil

<sup>2</sup>Departamento de Geologia, Universidade Federal de Sergipe, 49107-230, São Cristóvão-Sergipe, Brasil

\* nyviasaturnino@academico.ufs.br

(Recebido em 30 de julho de 2025; aceito em 15 de setembro de 2025)

A integração de dados climáticos gradeados na modelagem hidrológica tem se consolidado como uma alternativa eficaz em regiões com rede escassa de monitoramento hidrometeorológico, como ocorre em grande parte do semiárido nordestino. Essa abordagem pode fortalecer a gestão dos recursos hídricos, promovendo maior segurança no planejamento dos usos múltiplos da água. Neste contexto, os modelos baseados em dados, ou *Data-Driven Models* (DDMs), têm ganhado destaque por dispensarem a complexidade estrutural dos modelos baseados em processos físicos, utilizando algoritmos numéricos para capturar padrões a partir de dados históricos. No entanto, a seleção de variáveis de entrada adequadas permanece como um desafio, especialmente devido à heterogeneidade espacial e temporal dos processos hidrológicos. Este trabalho visa avaliar a aplicação de diferentes métodos de DDM, como regressão linear múltipla (RLM), redes neurais artificiais (RNA) e o algoritmo dos k-vizinhos mais próximos (KNN), na modelagem da relação chuva-vazão na bacia hidrográfica do rio Vaza-Barris, empregando dados climáticos espaciais como variáveis explicativas. Os resultados indicam que, apesar das dificuldades na simulação precisa das vazões máximas, os modelos foram capazes de representar satisfatoriamente as vazões mínimas e a sazonalidade hidrológica local. O desempenho global foi classificado como apenas moderado, com valores do coeficiente de eficiência de Nash-Sutcliffe (NSE) variando entre os algoritmos testados, sugerindo a necessidade de refinamento na seleção de variáveis e parametrização dos modelos.

**Palavras-chave:** aprendizado de máquina, monitoramento hidrológico, modelos hidrológicos.

The integration of gridded climate data into hydrological modelling has become an effective alternative in regions with a sparse network of hydrometeorological monitoring stations, as is the case in much of the Brazilian semi-arid region. This approach can enhance water resource management, providing greater security in planning for multiple water uses. In this context, data-driven models (DDMs) have gained prominence for bypassing the structural complexity of physically based models, relying instead on numerical algorithms to capture patterns from historical data. However, selecting appropriate input variables remains a challenge, particularly due to the spatial and temporal heterogeneity of hydrological processes. This study aims to evaluate the application of different DDM approaches, such as multiple linear regression (MLR), artificial neural networks (ANN), and the k-nearest neighbours algorithm (KNN), for modelling the rainfall-runoff relationship in the Vaza-Barris River watershed, using spatial climate data as explanatory variables. The results indicate that, despite limitations in accurately simulating peak flows, the models were able to satisfactorily represent minimum flows and the watershed's hydrological seasonality. Overall performance was classified as only moderate, based on the Nash-Sutcliffe Efficiency (NSE) coefficient, with values varying across the tested algorithms. These findings highlight the need for improvements in variable selection and model parameterization.

**Keywords:** machine learning, hydrological monitoring, hydrological models.

## 1. INTRODUÇÃO

A simulação de vazões é uma tarefa complexa, muito devido à alta variabilidade espacial e temporal, afetada por diversos fatores ao longo da bacia hidrográfica, particularmente pelos componentes do balanço hídrico [1, 2]. Ao longo da história, diversas metodologias foram desenvolvidas para esse fim, incluindo o método racional e o hidrograma unitário, além de

modelos conceituais, distribuídos, físicos e baseados em inteligência artificial [3-5]. A evolução computacional desde a década de 1950 impulsionou a criação e otimização desses modelos, com um esforço crescente entre os anos de 1960 e 2010 para otimizar sua precisão e reduzir a dependência de parâmetros [6]. Contudo, a heterogeneidade das bacias hidrográficas e a variação das condições meteorológicas locais apresentam desafios significativos para a modelagem hidrológica, especialmente em regiões com dados limitados, onde a aplicação desses modelos pode resultar em significativas incertezas [7].

Os modelos conceituais são comumente utilizados para simular os processos físicos de bacias hidrográficas com precisão moderada, embora enfrentem desafios como a interpretação algorítmica, incerteza nos parâmetros e a exigência de um extenso conhecimento hidrológico e coleta de dados de campo [8]. Em situações em que a complexidade dos sistemas hidrológicos demanda uma abordagem mais flexível e adaptativa, os modelos baseados em dados (DDMs) têm se destacado. Entre eles, as redes neurais artificiais (RNA) são amplamente reconhecidas por sua capacidade de modelar sistemas não-lineares complexos e lidar com qualquer volume de dados, incluindo informações ruidosas [9-11]. Ademais, o modelo do algoritmo dos k-vizinhos mais próximos (KNN), classificado como modelo de aprendizado baseado em instâncias (IBL: *instance-based learning*), e a regressão linear múltipla (RLM) têm se mostrado promissores em estudos recentes [12-15].

Os Modelos Baseados em Dados (DDMs) são abordagens empíricas que se concentram na análise e correlação dos dados de entrada e saída através de funções matemáticas, sem a necessidade de uma conexão direta com os processos físicos subjacentes. Esses modelos utilizam técnicas estatísticas e de aprendizado de máquina para identificar padrões e relações entre variáveis, permitindo previsões e simulações sem a necessidade de detalhadas equações que descrevam os processos físicos envolvidos [16, 17]. No aprendizado baseado em instâncias (IBL), todos os exemplos históricos (instâncias) são armazenados na memória sem qualquer generalização dos dados fornecidos durante o processo de aprendizado e o processamento do algoritmo ocorre durante a previsão, onde uma métrica é utilizada para identificar os exemplos passados mais semelhantes à situação atual, permitindo prever eventos futuros com base no que ocorreu anteriormente [18].

Estudos desse tipo utilizam dados de pluviômetros ou estações meteorológicas como entrada nos modelos [18], uma vez que a precipitação é fundamental no ciclo hidrológico, sendo a variável principal para a modelagem hidrológica e para representar os processos de transformação chuva-vazão [19]. Porém, nos países em desenvolvimento, a qualidade desses dados é frequentemente comprometida devido a uma má distribuição espacial e temporal desses equipamentos. Atualmente, diversas organizações disponibilizam produtos de sensoriamento remoto, como dados de precipitação, temperatura e umidade do ar, essenciais para a modelagem hidrológica [20, 21]. Exemplos incluem o *Climate Forecast System Reanalysis* (CFSR), a *Tropical Rainfall Measuring Mission* (TRMM), a *Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record* (PERSIANN-CDR) e o *Climate Hazards Group Infrared Precipitation with Stations* (CHIRPS), que oferecem conjuntos de dados em escala global ou quase global nos últimos anos [22-25]. No Brasil, surge como uma alternativa viável Rasera et al. (2023) [26], por exemplo, a base de dados gradeada desenvolvida por Xavier et al. (2022) [27], conhecida como *Brazilian Daily Weather Gridded Data* (BRDWGD).

Embásado neste contexto, este trabalho busca avaliar o desempenho da aplicação da BRDWGD em modelos chuva-vazão utilizando metodologias de DDMs, como regressão linear múltipla (RLM), redes neurais artificiais (RNA) e algoritmo dos k-vizinhos mais próximos (KNN), em escala mensal, na bacia hidrográfica do rio Vaza-Barris, na região Nordeste.

## 2. MATERIAL E MÉTODOS

A bacia hidrográfica do rio Vaza-Barris é de dominância federal, uma vez que abrange os estados da Bahia e Sergipe (Figura 1), e revela uma diversidade de aspectos socioambientais distintos. A região na Bahia compreende 82% do território da bacia, é caracterizada pela

intermitência significativa dos rios e por uma menor densidade populacional em comparação com Sergipe, que abrange apenas 18% [28]. Com uma extensão territorial de 16.183,25 km<sup>2</sup>, a bacia abrange 30 municípios e possui uma população estimada em 891.426 habitantes [29]. Sua nascente localiza-se na Serra da Canabrava, no município de Uauá/BA, e sua foz no estado de Sergipe, entre os municípios de Aracaju/SE e São Cristóvão/SE.

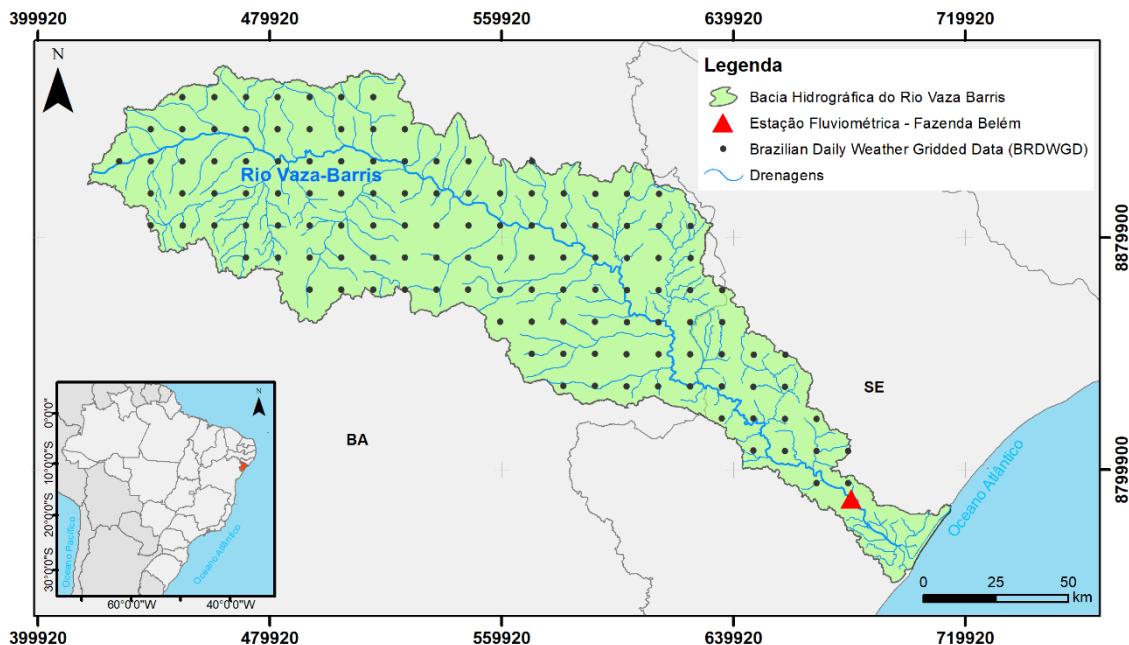


Figura 1: Mapa de localização da bacia hidrográfica do rio Vaza-Barris.

Na bacia, o clima é predominantemente semiárido, classificado como Bsh e Bwh segundo a classificação de Köppen, marcado por períodos prolongados e frequentes de seca, onde a precipitação média anual varia de 500 a 800 mm. Em contrapartida, na região costeira, o clima é úmido (Aw), com uma média anual de precipitação que varia entre 1200 e 1600 mm [30]. O regime hidrológico na área de captação apresenta uma clara sazonalidade, com médias mensais de 6 m<sup>3</sup>/s na porção inferior do curso. No curso superior, o canal do rio permanece seco durante a maior parte do ano. No entanto, ocorrem inundações significativas na parte inferior, com fluxos diários que podem exceder 500 m<sup>3</sup>/s, embora sejam eventos de baixa frequência, com uma recorrência estimada de 10 anos para inundações com um volume de 388 m<sup>3</sup>/s [31].

Neste estudo são utilizados dados de vazão (Q), precipitação (P) e evapotranspiração potencial de referência (ETo) para o período de 1961 a 2019. Os dados fluviométricos foram obtidos através do portal HidroWeb, da Agência Nacional de Águas e Saneamento Básico (ANA), especificamente da estação denominada FAZENDA BELÉM (código: 50191000), localizada no município de Itaporanga D'Ajuda, no estado de Sergipe. Os dados de precipitação e ETo foram extraídos da base de dados gradeada desenvolvida por Xavier et al. (2022) [27], a BRDWGD, com resolução espacial de 0,1° × 0,1° e resolução temporal diária. Foram utilizados 128 pontos da grade para a Bacia do Rio Vaza-Barris, como pode ser visto na Figura 1.

A coleta de dados de precipitação, vazão e evapotranspiração potencial de referência (ETo) é essencial para a modelagem chuva-vazão, pois eles constituem os atributos de entrada dos modelos. Sendo assim, no estudo, foram utilizados dados de vazão média mensal (Q, em m<sup>3</sup>/s), precipitação acumulada mensal (P, em mm) e evapotranspiração média mensal (ETo, em mm) para o período de 1961 a 2019, o qual foi estabelecido um critério para a seleção de dados de vazão sem falhas, escolhendo-se meses com dados disponíveis em mais de 28 dias. Dessa forma, 98% dos meses disponíveis foram considerados, resultando em 696 meses válidos de um total de 708 meses da série histórica completa.

O procedimento de modelagem adotado para prever as vazões estima o valor da vazão  $Q$  no período  $t+H$ , onde  $t$  é o mês atual e  $H$  é o horizonte de previsão. Essa estimativa é baseada nos valores de precipitação  $P$  e deflúvio em meses passados ( $t-1, t-2, \dots$ ) e nos valores de  $ETo$  do mês atual  $ETo(t)$ . Foi considerado  $y=Q(t+H)$  como a vazão a ser estimada e  $X=[P(t-1), P(t-2), \dots, P(t-\tau), Q(t-1), Q(t-2), \dots, Q(t-\tau), ETo(t)]$  como o vetor de atributos de entrada, ou seja, os valores utilizados de precipitação e vazão em meses anteriores e evapotranspiração potencial de referência que serão utilizados para estimar o deflúvio de meses futuros. O objetivo é encontrar uma função  $h$  que mapeie a entrada  $X$  com saída  $y$  da forma  $y=h(X)$  do modo da Equação 1.

$$Q(t+H) = h(P(t), P(t-1), \dots, P(t-\tau), Q(t), Q(t-1), \dots, Q(t-\tau), ETo(t)) \quad \text{Equação 1}$$

Esta pesquisa usou  $H=0$ . As funções  $h$  adotadas foram os DDMs de Regressão Linear Múltipla (RLM), Rede Neural Artificial (ANN) e modelo de K-Vizinhos Mais Próximos (KNN). As formulações dos DDMs foram escolhidas de maneira empírica, e estão apresentadas no Quadro 1.

Quadro 1 - Formulações adotadas para a modelagem.

1	$Q(t) = h(P(t), P(t-1), P(t-2), Q(t-1), ETo(t))$
2	$Q(t) = h(P(t), P(t-1), Q(t-1), Q(t-2), ETo(t))$
3	$Q(t) = h(P(t), Q(t-1), Q(t-2), Q(t-3), ETo(t))$
4	$Q(t) = h(P(t-1), P(t-2), Q(t-1), Q(t-2), ETo(t))$
5	$Q(t) = h(P(t-1), Q(t-1), Q(t-2), Q(t-3), ETo(t))$

As formulações dos DDMs foram escolhidas de forma a capturar não apenas a resposta imediata da bacia às chuvas do mês corrente, mas também o impacto do escoamento de base e dos eventos de precipitação dos meses anteriores. A inclusão de vazões e precipitações defasadas como covariáveis [32-35] é recomendada para bacias de grande extensão como a do Vaza-Barris, onde o tempo de resposta hidrológica é mais lento e o fluxo de base tem um papel significativo na vazão total. Além disso, a inclusão da  $ETo$  como uma variável de entrada apresenta uma contribuição importante, uma vez que a maioria dos estudos na área considera apenas a precipitação como principal entrada para os modelos hidrológicos.

As configurações das camadas ocultas inseridas nas RNA e a quantidade de k-vizinhos aplicados ao KNN para as 5 formulações estão dispostas na Tabela 1. Para realizar a modelagem, os dados coletados foram divididos em duas partes: calibração e validação. A etapa de calibração é utilizada para treinar o modelo, enquanto a etapa de validação verifica a eficácia do modelo em dados não utilizados durante o treinamento. Neste estudo, 70% dos dados foram alocados para calibração e 30% para validação, selecionados aleatoriamente, sem considerar a cronologia.

Tabela 1 - Formulações adotadas para a modelagem.

Redes Neurais Artificiais	K-Vizinhos Mais Próximos
RNA - A	KNN - 3
RNA - 5,2	KNN - 5
RNA - 10, 5, 2	KNN - 10
RNA - 20, 10, 5	KNN - 15

Foi empregado o *software* Weka (*Waikato Environment for Knowledge Analysis*), desenvolvido na Universidade de Waikato (Nova Zelândia), multiplataforma e escrito na

linguagem Java para a aplicação de RLM, RNA e KNN na modelagem chuva-vazão. Para avaliar o desempenho da simulação gerada pelo Weka, foi utilizada a eficiência de *Nash-Sutcliffe* (Equação 2). Esta equação foi aplicada aos dados de calibração, aos dados de validação e ao conjunto completo de dados.

$$NSE = 1 - \frac{\sum_{i=1}^m (y_i - \bar{Y})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad \text{Equação 2}$$

Onde  $y_i$  é o valor observado,  $\bar{Y}$  é o valor previsto,  $\bar{y}$  é a média dos valores observados e  $m$  é o número de amostras (instâncias). A eficiência de *Nash-Sutcliffe* ( $NSE$ ) pode variar de  $-\infty$  a 1 e quanto mais próximo do valor unitário, melhor o modelo.

Destacam-se especialmente os valores de  $NSE$  obtidos a partir dos dados de validação, uma vez que esses dados não são usados durante o treinamento do modelo, contribuindo assim para testar e confirmar a eficácia da modelagem desenvolvida. Os coeficientes de  $NSE$  calculados pela Equação 2 são então classificados conforme Moriasi et al. (2007) [36] como: Muito Bom:  $(0,75 < NSE \leq 1)$ , Bom  $(0,65 < NSE \leq 0,75)$ , Satisfatório  $(0,5 < NSE \leq 0,65)$ , Insatisfatório  $(NSE \leq 0,5)$ .

### 3. RESULTADOS E DISCUSSÃO

Os resultados mostraram variações no desempenho do  $NSE$  para as diferentes formulações adotadas, tanto na análise visual dos gráficos quanto entre as diferentes técnicas de DDM utilizadas neste estudo (Regressão Linear Múltipla, Redes Neurais Artificiais e modelo de K-Vizinhos Mais Próximos). A discussão a seguir detalha essas diferenças, destacando a capacidade dos modelos em prever os picos de vazão, vazões mínimas e sazonalidade. Para cada formulação foi selecionado o modelo com melhor resultado de  $NSE$ .

A Figura 2 apresenta o modelo da Formulação 1 (RNA-A), cujo  $NSE$  resultou em 0,55 (satisfatório). O que se observa na análise visual do gráfico, ao se comparar as vazões reais com as calculadas ao longo da série histórica, é que o modelo tende a subestimar os picos de vazão, com exceção dos picos de mar/85 e jul/88. Todos os picos acima de  $100 \text{ m}^3/\text{s}$  não foram captados pelo modelo, indicando uma dificuldade em prever a intensidade dos eventos extremos. No que se trata das vazões mínimas, o modelo teve um desempenho mais consistente, acompanhando a série de modo geral, com pequenas flutuações acima ou abaixo das vazões observadas. O RNA-A acompanha bem a sazonalidade da bacia, refletindo os padrões de aumento e diminuição do deflúvio, apesar da dificuldade com as máximas.

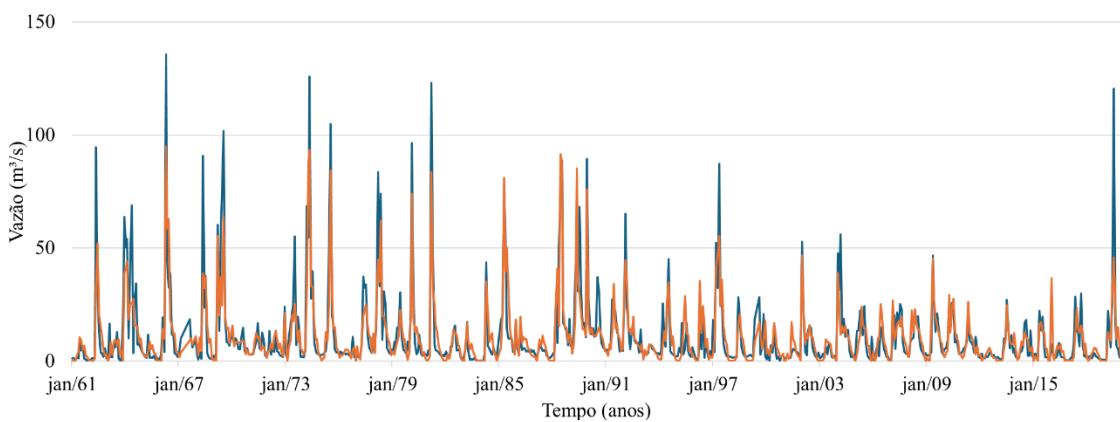


Figura 2: Vazões reais (azul) e vazões calculadas no modelo da formulação 1 (RNA-A).

A dificuldade na simulação precisa das vazões máximas, observada para o RNA-A e outros modelos, é um desafio comum em modelagem hidrológica. Estudos recentes apontam que,

embora as RNAs sejam amplamente reconhecidas por sua capacidade de modelar sistemas não-lineares complexos e lidar com dados ruidosos, elas podem apresentar limitações na captura de eventos extremos. Safari et al. (2020) [37], por exemplo, descobriram que a RBFNN foi o pior modelo para previsão de vazão em tempos de avanço maiores em comparação com outras técnicas. No entanto, outros trabalhos, como o de Turhan (2021) [38] e Aoulmi et al. (2021) [39], indicaram que as RNAs podem apresentar bons resultados estatísticos e até superar métodos convencionais, com o desempenho melhorando com a inclusão de mais variáveis de entrada.

Aoulmi et al. (2021) [39] demonstraram que quanto mais variáveis de entrada (precipitação, evapotranspiração, temperatura, umidade, velocidade do vento) eram fornecidas, mais eficientes se tornavam os modelos ANN. Isso sugere que o modelo RNA-A do presente trabalho tenha apresentado desempenho satisfatório para a sazonalidade e vazões mínimas. A inclusão de variáveis adicionais ou a otimização da estrutura da rede, como demonstrado por Morovati et al. (2024) [40], poderia potencialmente aprimorar a previsão de picos de vazão.

Na formulação 2, o modelo com melhor desempenho foi o RLM, com NSE de 0,57 (satisfatório). Ao se comparar os gráficos das vazões reais com as vazões calculadas na Figura 3, nota-se que esse modelo também apresentou dificuldade em prever os picos de vazão da bacia. A formulação tende a seguir as elevações de deflúvio, destacando sua capacidade de captar a sazonalidade, porém somente em picos entre 40 e 60 m<sup>3</sup>/s, são mais precisos, como em 2002 e 2004. Picos acima de 60 m<sup>3</sup>/s não foram bem estimados. O modelo, apesar de algumas discrepâncias, atende melhor a previsão das vazões mínimas, também captando bem as diminuições das vazões.

A dificuldade da RLM em prever picos de vazão acima de 60 m<sup>3</sup>/s é consistente com pesquisas que apontam limitações de modelos lineares para capturar a complexidade não linear da relação chuva-vazão. Singh et al. (2022) [41] concluíram que a RLM tinha precisão inferior em comparação com os outros métodos heurísticos mais complexos. Adicionalmente, Diththakit et al. (2021) [42] mostraram que o modelo *Random Forest* superou a RLM em termos de precisão na estimativa de parâmetros. Por outro lado, Shekar et al. (2023) [43] relataram que a RLM pode atingir desempenho aceitável na simulação de vazão mensal, com valores de NSE de 0,65 (treinamento) e 0,76 (teste), o que sugere que o desempenho pode variar significativamente dependendo das características da bacia e do pré-processamento dos dados. A capacidade da RLM de captar a sazonalidade e prever bem as vazões mínimas é um ponto positivo que se alinha com a utilidade da RLM para capturar padrões gerais em dados históricos.

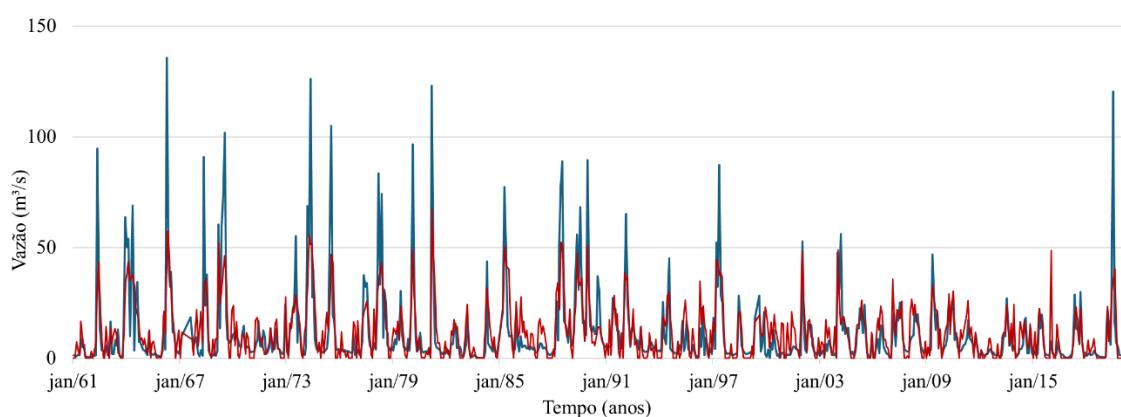


Figura 3: Vazões reais (azul) e vazões calculadas no modelo da formulação 2 (RLM).

Dois modelos na formulação 3 apresentaram o mesmo desempenho, com um NSE igual a 0,56 (satisfatório): os modelos KNN-5 e RLM. Na Figura 4A são exibidos os gráficos de vazão observada e calculada no modelo KNN-5. Assim como no modelo discutido anteriormente, o KNN-5 da formulação 3 teve dificuldades em captar as vazões de pico acima de 60 m<sup>3</sup>/s. Nas vazões mínimas, o modelo, apesar de algumas flutuações, seguiu o comportamento das demais

com uma maior consistência. Se destaca mais uma vez a capacidade de seguir as elevações e diminuições do deflúvio, ou seja, prevendo bem a sazonalidade das vazões da bacia.

O comportamento observado nos gráficos das vazões reais e das vazões calculadas pelo modelo de RLM na formulação 3, na Figura 4B, foi semelhante aos demais modelos em relação à previsão dos picos de vazão, mostrando dificuldade em captar vazões superiores a  $60 \text{ m}^3/\text{s}$ , exceto em 1988. Entretanto, ao analisar as baixas vazões, o modelo apresentou um desempenho melhor, apesar de algumas pequenas flutuações. A sazonalidade é bem detectada, com as elevações e diminuições de vazão sendo bem condizentes com o que é medido no posto fluviométrico.

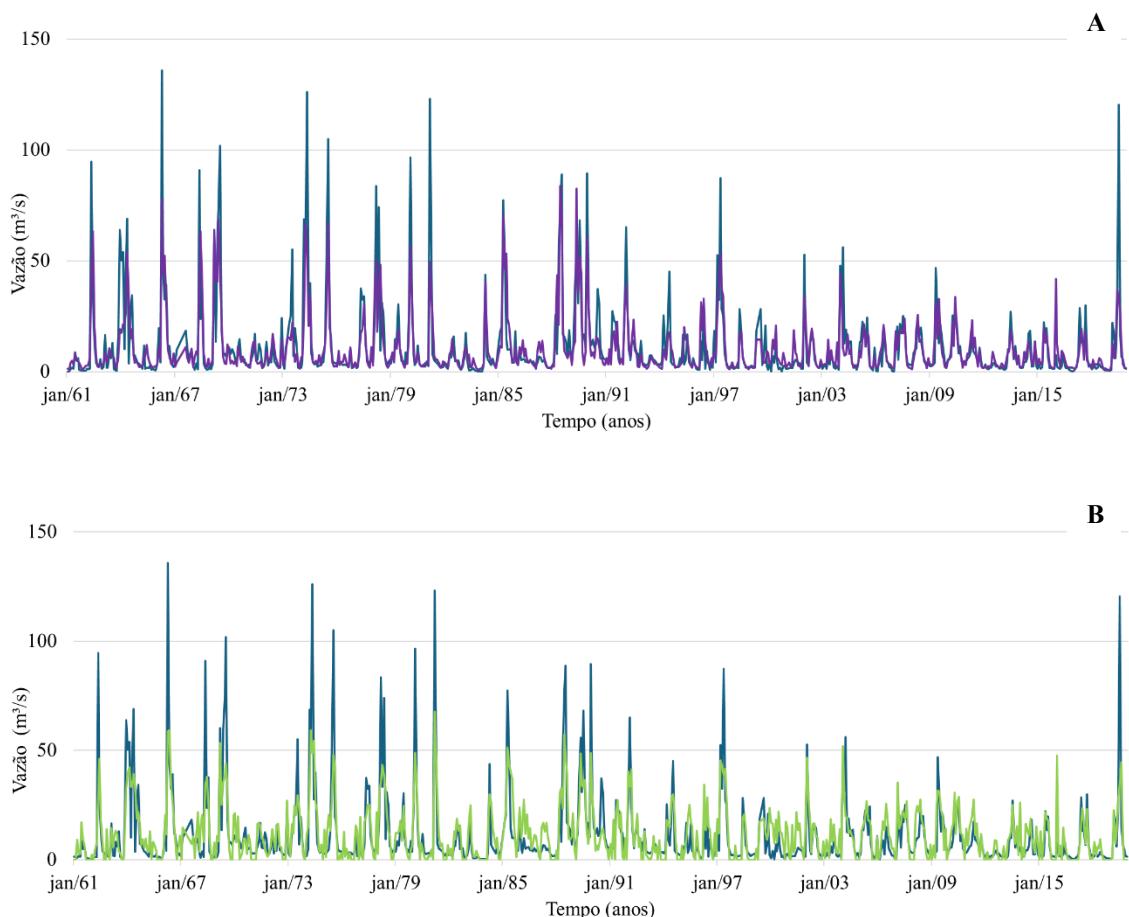


Figura 4: Vazões reais (azul) e vazões calculadas no modelo da formulação 3 (A) KNN-5 e (B) RLM.

McEachran et al. (2024) [44] demonstraram que o método KNN, apesar de simples, pode oferecer resultados promissores, alcançando um NSE de 0,73 em bacias dominadas pela chuva. Contudo, a capacidade do KNN de gerar extremos fora da faixa dos dados observados [45], pode ser uma explicação para a dificuldade observada. Apesar disso, o KNN pode ser eficaz, especialmente quando combinado com outras técnicas para correção de erros, como o modelo híbrido LSTM-KNN (*Long Short-Term Memory neural network - k-nearest neighbor*) proposto por Liu et al. (2020) [46], que demonstrou desempenho superior ao LSTM isolado.

Anaraki et al. (2023) [47] encontraram que o modelo KNN-GTO (*gorilla troop optimizer*) apresentou o melhor desempenho geral entre todos os modelos testados, mas ainda assim, ressaltam que o KNN, mesmo de forma autônoma, demonstrou desempenho razoável em comparação com outras abordagens. A semelhança no desempenho entre KNN-5 e RLM na formulação 3, ambos com NSE de 0,56, sugere que, para essa formulação específica, as relações capturadas pelos dois modelos apresentaram um nível de precisão comparável, embora com as mesmas limitações para picos extremos.

Os modelos da formulação 4 resultaram em desempenho insatisfatório, o melhor deles foi o KNN-5 que obteve um NSE de 0,27 (insatisfatório). O gráfico de comparação entre as vazões reais e as calculadas apresentado na Figura 5 mostra que o modelo, apesar de acompanhar as elevações e diminuições de vazão da bacia, captando a sazonalidade dela, também seguiu o padrão de ter mais dificuldade em prever os picos do que as baixas vazões. Porém, nos 4 maiores picos da série histórica, acima de  $120 \text{ m}^3/\text{s}$ , 1966, 1974, 1981 e 2019, esse modelo teve previsões abaixo de  $60 \text{ m}^3/\text{s}$ , o que pode estar impactando no coeficiente de NSE.

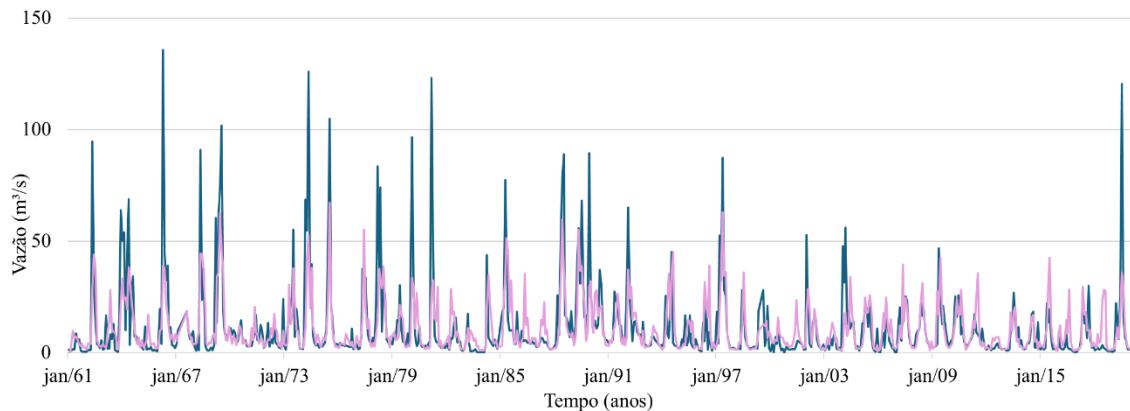


Figura 5: Vazões reais (azul) e vazões calculadas no modelo da formulação 4 (KNN-5).

Os modelos da formulação 5 também apresentaram desempenho insatisfatório, sendo o melhor modelo o KNN-10, com NSE 0,38. Assim como no modelo KNN-5 da formulação 4, as previsões dos maiores picos da série histórica ficaram bem abaixo quando comparado com os outros modelos, o que também pode estar impactando no NSE. Porém, o comportamento da sazonalidade e das baixas vazões é bem modelado, como visto na Figura 6.

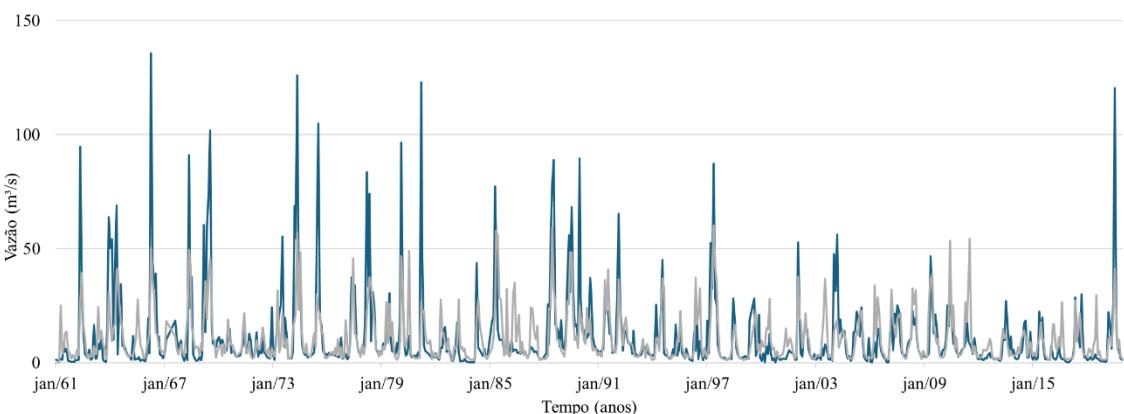


Figura 6: Vazões reais (azul) e vazões calculadas no modelo da formulação 5 (KNN-10).

Uma análise geral do comportamento do desempenho de NSE para cada formulação pode ser feita observando-se a Figura 7. Nota-se que nenhum modelo foi classificado como muito bom ou bom (NSE acima de 0,65). A formulação que obteve um maior número de NSE satisfatórios foi a número 2, com apenas dois modelos com desempenho insatisfatório (KNN-3 e KNN-5), logo em seguida foi a formulação número 3, com 3 modelos classificados com eficiência insatisfatória (RNA-A, RNA-5, 2, RNA-10, 5, 2). A formulação 1 obteve apenas um modelo satisfatório: RNA-A. As formulações com as menores eficiências foram a 4 e a 5, todos os seus modelos foram classificados como insatisfatórios, destacando inclusive que os modelos de RNA apresentaram NSE negativos na formulação 5 e no modelo RNA-A da formulação 4.

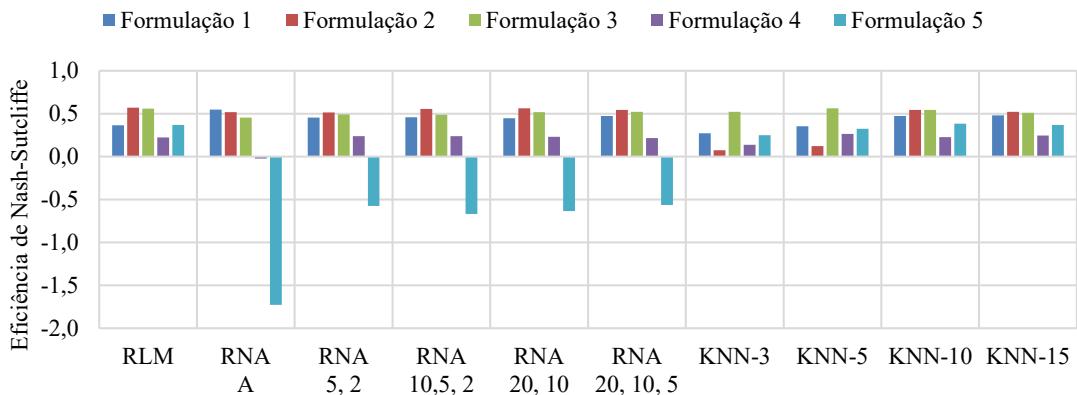


Figura 7: Desempenho dos modelos para cada formulação empregada.

As técnicas de DDM que se destacaram em eficiência foram a RLM, RNA-A, RNA-20,10, RNA-20,10,5, KNN-10, KNN-15, cada uma com duas formulações classificadas como satisfatórias, as demais obtiveram uma formulação satisfatória cada. Nas RNA, o comportamento variou à medida que o número de camadas e neurônios foram crescendo. Em algumas formulações isso melhorou a eficiência até um modelo ótimo, por exemplo, na formulação 2 (RNA-20,10); ou diminuiu a eficiência, como na formulação 1; na formulação 5, porém, todos os valores de NSE foram negativos, inclusive a RNA-A dessa formulação obteve a pior eficiência. Nos KNN o comportamento foi otimizado à medida que o número de vizinhos crescia até um modelo ótimo, exceto na formulação 4, onde esse padrão não foi observado.

A baixa eficiência e dificuldade dos modelos em captar as vazões máximas pode ser explicada pela diversidade climática da região, onde segundo Santos e Mendes (2020) [48] os trimestres úmidos da cabeceira da bacia e do litoral são FMA (fevereiro, março, abril) e AMJ (abril, maio e junho) respectivamente, e que as lâminas acumuladas anuais médias são 272,94 mm e 1356,89 mm, respectivamente. A estação fluviométrica Fazenda Belém se encontra próxima ao litoral e, portanto, as vazões máximas ali registradas podem ser mais sensíveis às contribuições das precipitações do clima local. Uma abordagem com vetores de precipitação média dessas duas regiões é um apropriado teste para verificar se haveria uma melhora na eficiência.

A análise dos vetores de entrada de cada formulação revela um comportamento interessante: apenas as formulações 4 e 5, que não incluíram a precipitação do mês atual, apresentaram NSE insatisfatórios em todos os modelos. Embora a inclusão de precipitações defasadas tenha como objetivo capturar a influência do escoamento de base, estes resultados sugerem que, para a estação fluviométrica Fazenda Belém, a sensibilidade do modelo é maior em relação às precipitações mais recentes. Este comportamento pode ser atribuído a características da bacia onde as maiores lâminas de precipitação, que levam menos tempo para escoar até a estação, exercem uma contribuição mais significativa para o defluvio.

Apesar do desempenho moderado do modelo na previsão de picos de vazão, os resultados obtidos apresentam implicações práticas para a gestão de recursos hídricos no semiárido. A consistência na reprodução da sazonalidade e, em especial, a capacidade de prever as vazões mínimas é importante para o planejamento de uso da água, a alocação de recursos em períodos de seca e o manejo de reservatórios. Assim, a aplicação dos modelos, mesmo com as limitações apontadas, contribui para o monitoramento e gestão da disponibilidade de água na bacia.

A complexidade da bacia do Vaza-Barris, com sua heterogeneidade climática e hidrológica, sugere que uma abordagem mais granular poderia ser benéfica. Recomenda-se a divisão da bacia em sub-regiões climatológicas, o que permitiria uma análise mais direcionada e poderia melhorar a capacidade dos modelos de capturar as variações hidrológicas em diferentes partes da bacia. Além disso, a aplicação de outras fontes de dados climáticos e hidrológicos podem oferecer uma representação mais precisa da precipitação e de outras variáveis.

Adicionalmente, apesar do esforço de otimização já realizado neste trabalho, a área de inteligência artificial oferece uma vasta gama de possibilidades. Futuros estudos poderiam explorar outras arquiteturas de redes neurais ou diferentes métodos de otimização de hiperparâmetros, visando um melhor desempenho, especialmente na previsão de eventos extremos de vazão.

#### 4. CONCLUSÃO

Os resultados obtidos denotam a dificuldade dos modelos em captar os picos de vazão na bacia, o que sugere uma limitação na representação dos processos hidrológicos em resposta às condições climáticas locais. Contudo, todos os modelos mostraram consistência na previsão das vazões mínimas e na reprodução da sazonalidade das vazões. A avaliação do desempenho dos modelos utilizando o coeficiente de eficiência de *Nash-Sutcliffe* demonstra que, embora nenhum modelo tenha alcançado um desempenho classificado como "muito bom" ou "bom", algumas formulações apresentaram resultados satisfatórios. Diante da análise comparativa entre as diferentes formulações utilizadas, é evidente que a formulação 2, no modelo de Regressão Linear Múltipla (RLM), apresentou uma maior eficiência em comparação com as demais (NSE de 0,57). No entanto, dentre todos os modelos, a formulação 1 foi a que obteve a menor dificuldade em prever as máximas, de acordo com a análise visual.

O desempenho da aplicação da BRDWGD em modelos chuva-vazão na bacia do rio Vaza-Barris, empregando metodologias de DDM, mostrou-se limitada com as formulações utilizadas para previsão de cheias, devido à dificuldade em prever as vazões de pico. Isso se deve a complexidade da bacia, marcada por sua diversidade climática e hidrológica, o que ressalta a necessidade de abordagens diferentes e adaptáveis para a simulação de vazões. Recomenda-se que estudos futuros explorem estratégias de modelagem mais integradas, levando em consideração não apenas os dados hidrológicos disponíveis, mas também informações climáticas regionais, bem como características particulares da bacia.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Kumar V, Sharma KV, Caloiero T, Mehta DJ, Singh K. Comprehensive overview of flood modeling approaches: A review of recent advances. *Hydrology*. 2023;10(7):141. doi: 10.3390/hydrology10070141
2. Yaseen ZM. A New Benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Eng Sci*. 2023;4(3):65-103. doi: 10.51526/kbes.2023.4.3.65-103
3. Al-Amri NS, Ewea HA, Elfeki AM. Revisit the rational method for flood estimation in the Saudi arid environment. *Arab J Geosci*. 2022;15(6):532. doi: 10.1007/s12517-021-09219-0
4. Guo J. Application of general unit hydrograph model for baseflow separation from rainfall and streamflow data. *J Hydrol Eng*. 2022;27(11):1-11.
5. Aghelpour P, Varshavian V. Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series. *Stoch Environ Res Risk Assess*. 2020;34(1):33-50. doi: 10.1007/s00477-019-01761-4
6. Peel MC, McMahon TA. Historical development of rainfall-runoff modeling. *WIREs Water*. 2020 Sep;7(5):e1471. doi: 10.1002/wat2.1471
7. Rahim BEEA, Yusoff I. A practical approach for numerical modeling of a complex and data-limited hydrological system. *Hydrol Res*. 2023;54(11):1368-86. doi: 10.2166/nh.2023.048
8. Herrera PA, Marazuela MA, Hofmann T. Parameter estimation and uncertainty analysis in hydrological modeling. *Wiley Interdiscip Rev Water*. 2022;9(1):1-23. doi: 10.1002/wat2.1569
9. Debastiani AB, Rafaeli Neto SL, da Silva RD. Árvore modelo frente a uma rede neural artificial para a modelagem chuva-vazão. *Nativa*. 2019;7(5):527-34. doi: 10.31413/nativa.v7i5.7089
10. Mendonça L, Gomide I, Sousa J, Blanco C. Modelagem chuva-vazão via redes neurais artificiais para simulação de vazões de uma bacia hidrográfica da Amazônia. *Rev Gestão Água da América Lat*. 2021;18(1):e2. doi: 10.21168/rega.v18e2
11. Sharma P, Singh S, Sharma SD. Artificial Neural Network approach for hydrologic river flow time series forecasting. *Agric Res*. 2022;11(3):465-76. doi: 10.1007/s40003-021-00585-5

12. de Araújo CBC, de Souza Filho FA, de Araújo Júnior LM, Silveira CS. Seasonal flow forecast for the orós dam (Ceará, brazil) using neural networks and the resampling technique of k-neighbors. *Rev Bras Meteorol.* 2020;35(2):197-207. doi: 10.1590/0102-7786351015
13. Malfatti MGL, Cardoso AO, Hamburger DS. Modelo Empírico Linear para previsão de vazão de rios na Usina Hidrelétrica de Itaipu – Bacia do rio Paraná. *Rev Bras Meteorol.* 2018;33(2):257-68. doi: 10.1590/0102-7786332005
14. Lucas EWM, de Sousa FAS, Silva FDS, da Rocha Júnior RL, Ataide KRP. Previsões de vazões mensais na Bacia Hidrográfica do Xingu - leste da Amazônia. *Rev Bras Meteorol.* 2020;35(spe):1045-56. doi: 10.1590/0102-77863550012
15. Monteiro RML, dos Santos FAA, Rolim PAM, Barreto CEC, da Rocha EJP. Hydrological Statistical Modeling in the Tocantins River Hydrographic Basin. *Rev Bras Geogr Fis.* 2024;17(2):1158-73. doi: 10.26848/rbgf.v17.2.p1158-1173
16. Celeste AB, Menezes TLA, Fontes VA. Modelagem do processo chuva-vazão pelo Tank Model e por redes neurais artificiais na bacia do rio Japaratuba, Sergipe. *Sci Plena.* 2014;10(7):1-13..
17. Hauswirth SM, Bierkens MFP, Beijk V, Wanders N. The potential of data driven approaches for quantifying hydrological extremes. *Adv Water Resour.* 2021;155:104017. doi: 10.1016/j.advwatres.2021.104017
18. Santos JRS, Celeste AB. Avaliação de estratégias de modelagem guiada por dados para previsão de vazão em rio sergipano. *Ambient e Agua - An Interdiscip J Appl Sci.* 2014 Sep 1;9(3):445-58. doi: 10.4136/ambi-agua.1356.
19. Loritz R, Hrachowitz M, Neuper M, Zehe E. The role and value of distributed precipitation data in hydrological models. *Hydrol Earth Syst Sci.* 2021;25(1):147-67. doi: 10.5194/hess-25-147-2021
20. Musie M, Sen S, Srivastava P. Comparison and evaluation of gridded precipitation datasets for streamflow simulation in data scarce watersheds of Ethiopia. *J Hydrol.* 2019;579:124168. doi: 10.1016/j.jhydrol.2019.124168
21. Xiang Y, Chen J, Li L, Peng T, Yin Z. Evaluation of eight global precipitation datasets in hydrological modeling. *Remote Sens.* 2021;13(14):1-20. doi: 10.3390/rs13142831
22. Bui MT, Lu J, Nie L. Evaluation of the climate forecast system reanalysis data for hydrological model in the Arctic watershed Målselv. *J Water Clim Chang.* 2021;12(8):3481-504. doi: 10.2166/wcc.2021.346
23. Che Wan Zanial WN, Malek MA, Md Reba MN, Zaini N, Ahmed AN, Sherif M, et al. Rainfall-runoff modelling based on global climate model and tropical rainfall measuring mission (GCM -TRMM): A case study in Hulu Terengganu catchment, Malaysia. *Heliyon.* 2023;9(5):e15740. doi: 10.1016/j.heliyon.2023.e15740
24. Salehi H, Sadeghi M, Golian S, Nguyen P, Murphy C, Sorooshian S. The application of PERSIANN family datasets for hydrological modeling. *Remote Sens.* 2022;14(15):3675. doi: 10.3390/rs14153675
25. Yu C, Hu D, Duan X, Zhang Y, Liu M, Wang S. Rainfall-runoff simulation and flood dynamic monitoring based on CHIRPS and MODIS-ET. *Int J Remote Sens.* 2020;41(11):4206-25. doi: 10.1080/01431161.2020.1714779
26. Rasera JB, da Silva RF, Piedade S, Mourão Filho FAA, Delbem ACB, Saraiva AM, et al. Do Gridded weather datasets provide high-quality data for agroclimatic research in Citrus production in Brazil? *Agri Engineering.* 2023;5(2):924-40. doi: 10.3390/agriengineering5020057
27. Xavier AC, Scanlon BR, King CW, Alves AI. New improved Brazilian daily weather gridded data (1961-2020). *Int J Climatol.* 2022;42(16):8390-404. doi: 10.1002/joc.7731
28. Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico Bahia [Internet]; 2022 [acesso em 02/04/2025]. Disponível em: <https://cidades.ibge.gov.br/brasil/ba/panorama>.
29. Nascimento RLX. Vaza-Barris [Internet]; Companhia de Desenvolvimento dos Vales do São Francisco e do Parnaíba (CODEVASF); 2020 [acesso em 02/04/2025]. Disponível em: <https://www.codevasf.gov.br/area-de-atuacao/bacia-hidrografica/vaza-barris>.
30. Santos LA, Latrubbles EM. The Vaza-Barris River: A late Quaternary decoupled accumulation-erosion model for rivers of the NE Brazil Atlantic passive margin. *J South Am Earth Sci.* 2022;120:104077. doi: 10.1016/j.jsames.2022.104077
31. Superintendência de Recursos Hídricos (SRH-Bahia). Plano diretor de recursos hídricos das Bacias dos Rios Vaza Barris e Real. Salvador: SRH; 1996..
32. Kwon M, Kwon HH, Han D. A hybrid approach combining conceptual hydrological models, support vector machines and remote sensing data for rainfall-runoff modeling. *Remote Sens.* 2020;12(11):1801.. doi: 10.3390/rs12111801
33. Melo L, Claudio DM, Cavalcante J, Frede B, Carvalho DO. Recurrent neural networks for rainfall - runoff modeling of small Amazon catchments. *Model Earth Syst Environ.* 2023;9(2):2517-31. doi: 10.1007/s40808-022-01626-w

34. Sousa Jr MF, Uliana EM, Aires RVU, Rápolo LMC, Silva DD, Moreira MC, et al. Streamflow prediction based on machine learning models and rainfall estimated by remote sensing in the Brazilian Savanna and Amazon biomes transition. *Model Earth Syst Environ.* 2024;10:1191-1202. doi: 10.1007/s40808-023-01837-9
35. Farfán-Durán JF, Cea L. Streamflow forecasting with deep learning models: A side-by-side comparison in Northwest Spain. *Earth Sci Inform.* 2024;17:5289-3155. doi: 10.1007/s12145-024-01454-9
36. Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE.* 2007;50(3):885-900. doi: 10.13031/2013.23153
37. Safari MJS, Rahimzadeh Arashloo S, Danandeh Mehr A. Rainfall-runoff modeling through regression in the reproducing kernel Hilbert space algorithm. *J Hydrol.* 2020;587:125014. doi: 10.1016/j.jhydrol.2020.125014
38. Turhan E. A Comparative evaluation of the use of artificial neural networks for modeling the rainfall-runoff relationship in water resources management. *J Ecol Eng.* 2021;22(5):166-78. doi: 10.1016/j.jhydrol.2020.125014
39. Aoulmi Y, Marouf N, Amireche M. The assessment of artificial neural network rainfall-runoff models under different input meteorological parameters Case study: Seybouse basin, Northeast Algeria. *J Water L Dev.* 2021;50:38-47. doi: 10.24425/jwld.2021.138158
40. Morovati R, Kisi O. Utilizing hybrid machine learning techniques and gridded precipitation data for advanced discharge simulation in under-monitored river basins. *Hydrology.* 2024;11(4):48. doi: 10.3390/hydrology11040048
41. Singh AK, Kumar P, Ali R, Al-Ansari N, Vishwakarma DK, Kushwaha KS, et al. An Integrated statistical-machine learning approach for runoff prediction. *Sustain.* 2022;14(13):1-30. doi: 10.3390/su14138209
42. Ditthakit P, Pinthong S, Salaeh N, Binnui F, Khwanchum L, Pham QB. Using machine learning methods for supporting GR2M model in runoff estimation in an ungauged basin. *Sci Rep.* 2021;11(1):1-16. doi: 10.1038/s41598-021-99164-5
43. Shekar PR, Mathew A, Arun PS, Gopi VP. Rainfall-Runoff modelling using SWAT and eight artificial intelligence models in the Murredu Watershed, India. *Environ Monit Assess.* 2023;195(9):1041. doi: 10.1007/s10661-023-11649-0
44. McEachran ZP, Kietzmann J, Johnston M. Parsimonious streamflow forecasting system based on a dynamical systems approach. *J Hydrol.* 2024;641(February):131776. doi: 10.1016/j.jhydrol.2024.131776
45. Eum HI, Simonovic SP, Kim YO. Climate change impact assessment using k-nearest neighbor weather generator: Case study of the Nakdong River Basin in Korea. *J Hydrol Eng.* 2010;15(10):772-85. doi: 10.1061/(asce)he.1943-5584.0000251
46. Liu M, Huang Y, Li Z, Tong B, Liu Z, Sun M, et al. The applicability of lstm-knn model for real-time flood forecasting in different climate zones in China. *Water (Switzerland).* 2020;12(2):1-21. doi: 10.3390/w12020440
47. Anaraki MV, Achite M, Farzin S, Elshaboury N, Al-Ansari N, Elkhrachy I. Modeling of monthly rainfall-runoff using various machine learning techniques in Wadi Ouahrane Basin, Algeria. *Water (Switzerland).* 2023;15(20):1-24. doi: 10.3390/w15203576
48. Santos LS, Mendes LA. Influência da oscilação da temperatura do oceano pacífico equatorial no regime de precipitação da região hidrográfica do Atlântico Leste. *Rev Bras Geogr Física.* 2020;13(4):1502-18. doi: 10.26848/rbgf.v13.4.p1502-1518