Modelagem da vazão anual por aprendizado de máquina na Bacia Hidrográfica do Alto Jaguaribe, CE

Tereza Margarida Xavier de Melo Lopes [1];Leonardo de Sousa Sampaio[2]; José Kerlly Soares de Araújo [3]; Tiago de Abreu Lima[4] ; Samiria *Maria Oliveira da Silva [5] & Francisco de Assis de Souza Filho* †

RESUMO – A previsão de vazões em bacias semiáridas é essencial para subsidiar a gestão dos recursos hídricos em regiões marcadas por elevada variabilidade climática e mudanças no uso e cobertura do solo. Este estudo avaliou o desempenho de diferentes algoritmos de aprendizado de máquina na previsão da vazão média anual da Bacia Hidrográfica do Alto Jaguaribe, Ceará, no período de 1985 a 2023. Foram utilizadas séries anuais de precipitação média, estimada pelo método dos Polígonos de Thiessen, classes de uso e cobertura do solo obtidas pelo MapBiomas e vazão média anual observada no Posto Fluviométrico de Iguatu. Foram comparados os modelos Árvore de Decisão, Random Forest, Support Vector Machine e Rede Neural Artificial, avaliados por NSE, RMSE, MAE e R^2. Os resultados indicaram forte associação entre precipitação e vazão, com correlação de 0,83, além de redução de áreas florestadas e expansão da agropecuária ao longo da série histórica. Entre os modelos avaliados, a Árvore de Decisão apresentou melhor desempenho no conjunto de teste, com NSE de 0,579, RMSE de 29,26 m³/s, MAE de 13,36 m³/s e R^2 de 0,976. Conclui-se que os modelos testados são promissores para representar vazões baixas e médias, embora ainda possam apresentar limitações na estimativa de eventos extremos.

ABSTRACT – Streamflow prediction in semi-arid basins is essential to support water resources management in regions characterized by high climate variability and changes in land use and land cover. This study evaluated the performance of different machine learning algorithms in predicting the mean annual streamflow of the Alto Jaguaribe River Basin, Ceará, from 1985 to 2023. Annual series of mean precipitation, estimated using the Thiessen Polygons method, land use and land cover classes obtained from MapBiomas, and mean annual streamflow observed at the Iguatu Fluviometric Station were used. Decision Tree, Random Forest, Support Vector Machine, and Artificial Neural Network models were compared and evaluated using NSE, RMSE, MAE, and R^2. The results indicated a strong association between precipitation and streamflow, with a correlation of 0.83, as well as a reduction in forested areas and an expansion of agricultural land over the historical series. Among the evaluated models, the Decision Tree showed the best performance in the test set, with an NSE of 0.579, RMSE of 29.26 m³/s, MAE of 13.36 m³/s, and R^2 of 0.976. It is concluded that the tested models are promising for representing low and medium streamflows, although they may still present limitations in estimating extreme events.

Palavras-Chave – Aprendizado de máquina; previsão de vazão; uso e cobertura do solo.

1) Doutoranda em Engenharia Civil (Recursos Hídricos), Universidade Federal do Ceará – Campus do Pici, lopes.melo.tereza@gmail.com.
2) Mestrando em Modelagem e Métodos Quantitativos, Universidade Federal do Ceará – Campus do Pici, leonardo.ssampaio95@gmail.com
3) Doutorando em Engenharia Civil (Recursos Hídricos), Universidade Federal do Ceará – Campus do Pici, kerllysrs@gmail.com.
4) Doutorando em Engenharia Civil (Saneamento), Universidade Federal do Ceará – Campus do Pici, tiagoabreu.engamb@gmail.com.
5) Professora do Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará – Campus do Pici, samiriamaria@gmail.com.
†) In memoriam. O Professor Assis contribuiu significativamente para a concepção, desenvolvimento e orientação deste estudo antes de seu falecimento.

INTRODUÇÃO

A gestão dos recursos hídricos no semiárido representa um desafio para governos e órgãos gestores, uma vez que as condições climáticas severas e o crescimento populacional contribuem para aumentar a vulnerabilidade às secas (Braga e Matushima, 2021). Conforme Fernandes e Moreira (2022), a irregularidade das chuvas exerce influência direta sobre a dinâmica dos recursos hídricos, afetando as condições de vida e de produção tanto nas áreas urbanas quanto nas áreas rurais.

Para mitigar esses impactos e aprimorar a gestão dos recursos hídricos, a compreensão das dinâmicas que afetam as bacias hidrográficas torna-se essencial (Adera et al., 2024). As mudanças no uso e ocupação do solo, aliadas à variabilidade climática, exercem um impacto direto na hidrologia local e na disponibilidade de água (Kumar et al., 2022; Viana et al., 2023)

Diante da complexidade e da não-linearidade das interações desses processos hidrológicos, as abordagens baseadas em aprendizado de máquina têm ganhado um destaque crescente devido à sua capacidade de lidar com grandes volumes de dados geoespaciais e climáticos (Xu e Liang, 2021). Diferentemente dos modelos físicos e conceituais, que frequentemente demandam extensas quantidades de dados paramétricos, os modelos orientados a dados operam de forma estatística, exigem menos parâmetros empíricos e são altamente capazes de capturar relações não-lineares complexas a partir do histórico de observações (Adera et al., 2024). Além disso, a integração contínua de dados dinâmicos de uso e cobertura do solo reduz os riscos de sobreparametrização e aumenta a fidelidade e robustez das simulações nesses modelos preditivos (Viana et al., 2023).

Nesse contexto, a Bacia Hidrográfica do Alto Jaguaribe constitui uma área relevante para analisar a resposta hidrológica em ambiente semiárido. A região abrange 24 municípios e cerca de meio milhão de habitantes, destacando sua importância territorial e socioeconômica para o Ceará (Oliveira et al., 2023). Por integrar a bacia do Rio Jaguaribe, a maior do estado, representa um recorte estratégico para investigar relações entre variabilidade climática, uso e cobertura do solo e dinâmica das vazões (Silva Neto, 2024).

Diante disso, este estudo tem como objetivo avaliar o desempenho de diferentes algoritmos de aprendizado de máquina na previsão da vazão média anual da Bacia Hidrográfica do Alto Jaguaribe, considerando variáveis de precipitação e uso e cobertura do solo no período de 1985 a 2023. Para isso, foram integradas séries temporais de classes de cobertura da terra obtidas pelo MapBiomas, precipitação média anual e vazão média anual observada no Posto Fluviométrico de Iguatu. A pesquisa busca contribuir para a compreensão da influência conjunta da variabilidade climática e das transformações da paisagem sobre a resposta hidrológica em uma bacia semiárida, além de comparar a capacidade preditiva de modelos supervisionados com diferentes níveis de complexidade.

METODOLOGIA

Área de estudo

A área de trabalho escolhida foi a Bacia Hidrográfica do Alto Jaguaribe, a qual tem uma área de drenagem de 24.636 km², correspondente a 16,56% do território Cearense. Localiza-se a montante do açude Orós. Esta bacia é composta por 24 municípios e apresenta a capacidade de acumulação de águas superficiais de 2.792.563.000 de m³, num total de 24 açudes públicos gerenciados pela COGERH (Figura 1).

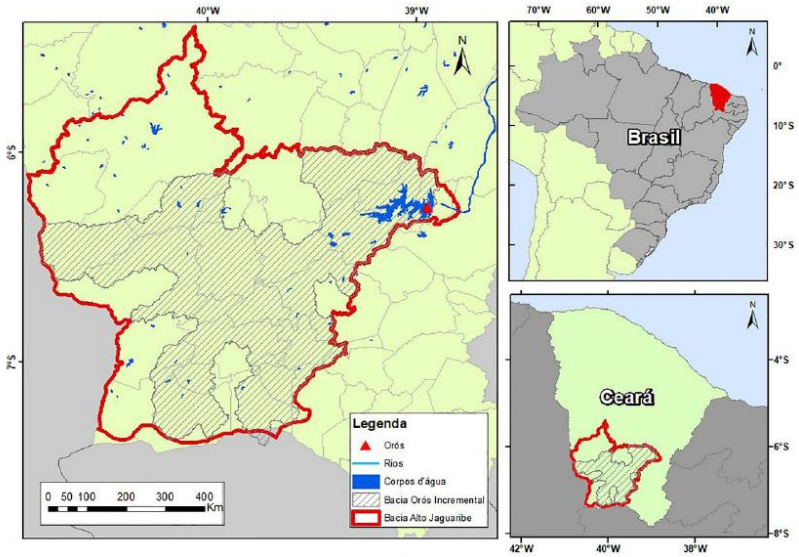


Figura 1 – Mapa de Localização da Bacia
Fonte: Oliveira, Souza Filho e Vasconcelos Junior (2023)

Coleta e estruturação dos dados

A estruturação do dataset desta pesquisa baseou-se na integração de informações de uso e ocupação do solo com variáveis hidrológicas, visando à aplicação de algoritmos de machine learning. A base de dados foi organizada em variáveis preditoras (features) e variável de resposta (target).

As features incluem séries de uso do solo e precipitação média da bacia. Os dados de cobertura da terra foram obtidos do projeto MapBiomas para a região do Alto Jaguaribe, considerando cinco classes: Forest (Floresta), Non_Forest (Vegetação Herbácea e Arbustiva), Farming (Agropecuária), Non_veg. (Área Não Vegetada) e Water_Envir (Superfície de água).

Para a variável climática, foram selecionadas 19 das 106 estações pluviométricas da área de estudo, priorizando-se aquelas com menor número de falhas e maior continuidade temporal. Com os totais anuais acumulados validados, aplicou-se o método dos Polígonos de Thiessen para obter a precipitação média anual espacial da bacia. Como target da modelagem, utilizou-se a série de vazão média anual do Posto Fluviométrico de Iguatu. A matriz final de dados compreendeu o período de 1985 a 2023, garantindo compatibilidade temporal entre todas as séries analisadas.

Modelos de Previsão

Árvore de Decisão

A Árvore de Decisão é uma técnica de aprendizado supervisionado amplamente empregada em problemas de classificação e regressão devido à sua simplicidade e facilidade de interpretação. O modelo organiza o processo de predição em uma estrutura hierárquica composta por nós de decisão e folhas terminais, nas quais os dados são sucessivamente particionados com base nas variáveis explicativas. As divisões são realizadas de forma a maximizar a homogeneidade dos subconjuntos gerados, utilizando critérios como ganho de informação, índice de Gini ou redução da variância (Quinlan, 1986; Loh, 2011).

Uma das principais vantagens das árvores de decisão é a capacidade de representar relações não lineares entre variáveis sem exigir transformações complexas dos dados. Além disso, os resultados podem ser facilmente interpretados por meio das regras geradas pelo modelo. Entretanto, árvores excessivamente complexas tendem a apresentar sobreajuste aos dados de treinamento, reduzindo sua capacidade de generalização. Para minimizar esse problema, são frequentemente utilizados mecanismos de regularização, como poda e limitação da profundidade da árvore (James *et al*., 2021).

Florestas Aleatórias

O *Random Forest (RF)* é um método de aprendizado supervisionado baseado na combinação de múltiplas árvores de decisão, sendo classificado como uma técnica de ensemble. O algoritmo utiliza o procedimento de *bootstrap* para gerar diferentes amostras dos dados de treinamento e, adicionalmente, seleciona subconjuntos aleatórios de variáveis em cada divisão dos nós. Essa estratégia aumenta a diversidade entre as árvores e reduz a correlação entre os modelos individuais (Breiman, 2001).

A estimativa final é obtida pela agregação das previsões de todas as árvores da floresta, utilizando votação majoritária em problemas de classificação ou média das previsões em problemas de regressão (Hastie; Tibshirani; Friedman, 2009). Como consequência, o Random Forest geralmente apresenta maior estabilidade e capacidade de generalização quando comparado a uma única árvore de decisão. Além disso, o método permite avaliar a importância das variáveis preditoras, embora sua interpretação seja mais complexa devido ao grande número de árvores envolvidas no processo decisório (Liaw; Wiener, 2002).

Máquina de Vetor de Suporte

A Máquina de Vetores de Suporte (*Support Vector Machine – SVM*) é um algoritmo fundamentado na teoria da aprendizagem estatística e no princípio da minimização do risco estrutural. Seu objetivo consiste em encontrar um hiperplano ótimo capaz de separar observações pertencentes

a diferentes classes, maximizando a distância entre a fronteira de decisão e os exemplos mais próximos, denominados vetores de suporte (Cortes; Vapnik, 1995; Vapnik, 1995).

Para problemas em que a separação linear não é suficiente, o SVM utiliza funções kernel que permitem projetar os dados em espaços de maior dimensionalidade, possibilitando a modelagem de relações não lineares entre as variáveis (Schölkopf; Smola, 2002). Essa característica torna o método adequado para aplicações ambientais e hidrológicas, nas quais os fenômenos analisados frequentemente apresentam comportamento complexo. Entretanto, seu desempenho depende da escolha adequada dos hiperparâmetros e da função kernel utilizada durante o treinamento (Baydaroğlu; Koçak, 2014).

Rede Neural Artificial

As Redes Neurais Artificiais (*Artificial Neural Networks – ANN*) são modelos computacionais inspirados no funcionamento dos sistemas neurais biológicos. Sua estrutura é composta por neurônios artificiais organizados em camadas de entrada, camadas ocultas e camada de saída, conectados por pesos ajustáveis. Durante o treinamento, esses pesos são atualizados de forma iterativa por algoritmos de otimização, geralmente baseados no método de retropropagação do erro (Rumelhart; Hinton; Williams, 1986; Haykin, 2009).

As redes neurais apresentam elevada capacidade de aprendizado e são capazes de representar relações complexas e altamente não lineares entre variáveis explicativas e variáveis resposta. Essa característica tem favorecido sua aplicação em estudos de previsão hidrológica, climática e ambiental, especialmente quando há grande quantidade de dados disponíveis (Goodfellow; Bengio; Courville, 2016). Por outro lado, a definição da arquitetura da rede, do número de neurônios e dos parâmetros de treinamento pode influenciar significativamente o desempenho do modelo, além de dificultar a interpretação dos resultados quando comparada a métodos mais transparentes, como as árvores de decisão.

Implementação computacional e otimização dos modelos.

A modelagem preditiva foi desenvolvida no ambiente R, utilizando principalmente os pacotes *caret*, *rpart*, *randomForest*, *e1071*, *nnet* e *hydroGOF*. O pacote *caret* foi empregado para o treinamento dos modelos, aplicação da validação cruzada e otimização dos hiperparâmetros; os pacotes *rpart*, *randomForest*, *e1071* e *nnet* foram utilizados, respectivamente, para a implementação dos modelos Árvore de Decisão, Random Forest, SVM com kernel radial e Rede Neural Artificial. O pacote *hydroGOF* foi utilizado para o cálculo das métricas de desempenho.

A otimização dos hiperparâmetros foi realizada por meio de busca em grade associada à validação cruzada do tipo Leave-One-Out Cross-Validation (LOOCV). Para a Árvore de Decisão, foi

testado o parâmetro de complexidade *cp*, variando de 0 a 0,5, sendo selecionado *cp* = 0. No modelo Random Forest, avaliou-se o parâmetro *mtry*, com valores de 1 a 6, sendo selecionado *mtry* = 5. Para o modelo SVM com kernel radial, foram testadas combinações dos parâmetros *C* e *sigma*, sendo selecionados *C* = 5 e *sigma* = 0,05. Para a Rede Neural Artificial, foram avaliados o número de neurônios na camada oculta (*size*) e o parâmetro de regularização (*decay*), sendo selecionados *size* = 1 e *decay* = 0,001, com número máximo de 500 iterações.

Métricas de avaliação

O desempenho dos modelos foi avaliado nos conjuntos de treinamento e teste por meio dos indicadores *Nash-Sutcliffe Efficiency* (NSE), raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE) e coeficiente de determinação (R^2). O NSE foi utilizado para avaliar a eficiência dos modelos na simulação da vazão anual, sendo que valores mais próximos de 1 indicam melhor desempenho. O RMSE e o MAE foram empregados para quantificar os erros entre os valores observados e preditos, na escala real da vazão, enquanto o R^2 foi utilizado para medir o grau de associação entre as vazões observadas e estimadas.

RESULTADOS E DISCUSSÕES

A Figura 2 mostra a variação anual da precipitação e da vazão, bem como a evolução das principais classes de uso e cobertura do solo. Observa-se que os maiores valores de vazão ocorreram, em geral, em anos com precipitação elevada, com destaque para 1985 e 2004. Entretanto, a relação entre chuva e vazão não foi proporcional em todos os anos, indicando que outros fatores, como condições antecedentes de umidade, armazenamento hídrico e características da superfície da bacia, também influenciam a resposta hidrológica.

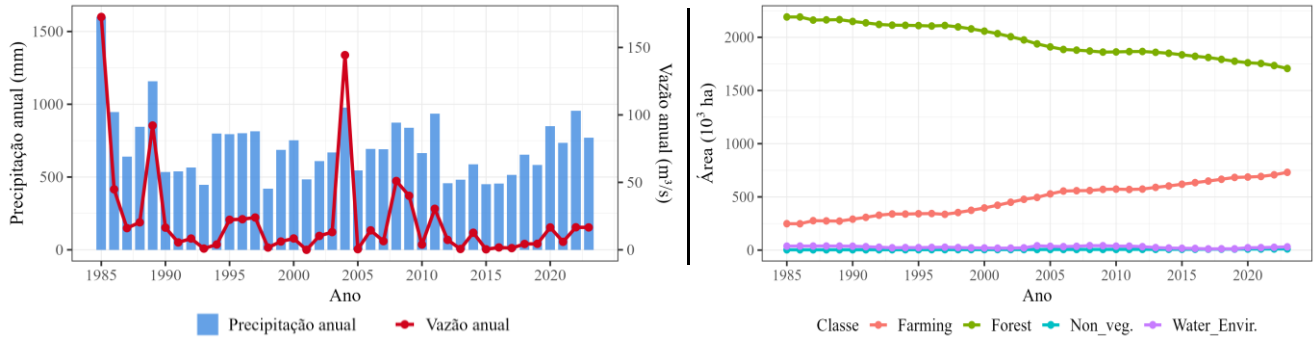


Figura 2 - Evolução temporal da precipitação anual, vazão média anual e uso e ocupação do solo

Em relação ao uso e cobertura do solo, verifica-se redução contínua da classe “Forest” e aumento expressivo da classe “Farming” ao longo da série histórica, indicando um processo de substituição de áreas florestadas por usos agropecuários. A classe “Non_veg” também apresentou crescimento, embora em menor magnitude, enquanto “Water_Envir” mostrou comportamento semelhante, entretanto, com maior oscilação temporal. De modo geral, a Figura 02 evidencia a

importância de considerar conjuntamente a variabilidade climática e as mudanças no uso da terra na interpretação da vazão anual da bacia.

Já a análise da Figura 3, indicaram forte associação entre a precipitação anual e a vazão anual da bacia, evidenciada pelo coeficiente de correlação de 0,83 entre Precipitação e Vazão. Esse comportamento reforça o papel dominante da variabilidade pluviométrica na resposta hidrológica anual, indicando que anos com maiores totais precipitados tendem a apresentar maiores vazões. A variável "Water_Envir" também apresentou correlação positiva com a vazão anual, sugerindo relação entre a presença de superfícies hídricas e o comportamento hidrológico da bacia.

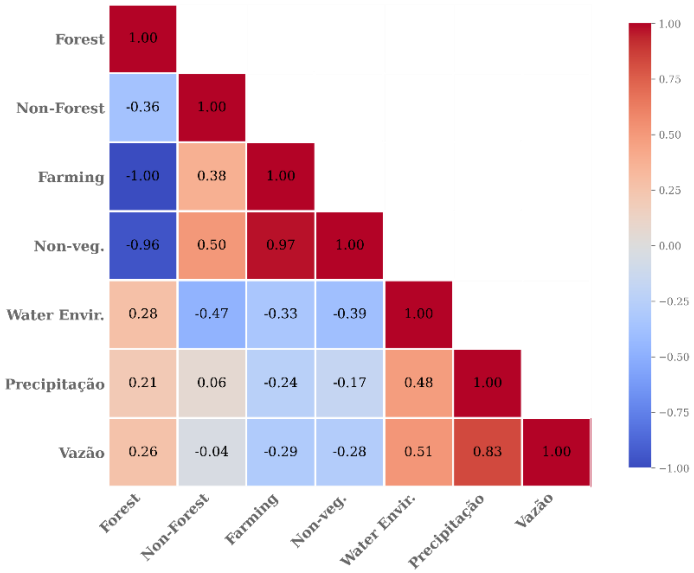


Figura 3 – Matriz de Correlação

As classes de uso e cobertura do solo apresentaram diferentes relações com a vazão. As variáveis "Farming" e "Non_veg" apresentaram correlação negativa com a vazão, enquanto "Non_Forest" exibiu correlação próxima de zero. Observou-se, contudo, elevada colinearidade entre algumas classes, especialmente entre "Forest" e "Farming", indicando que a substituição de áreas florestadas por usos agropecuários ocorreu de forma consistente ao longo da série histórica. Portanto, os resultados associados a essas variáveis devem ser interpretados de forma integrada, considerando a forte interdependência entre as classes de uso do solo.

Quanto ao desempenho dos modelos (Tabela 1), a Árvore de Decisão apresentou o melhor resultado no conjunto de teste, com NSE de 0,579, RMSE de 29,26 m³/s, MAE de 13,36 m³/s e R^2 de 0,976. Apesar de modelos como Random Forest, SVM e Rede Neural terem apresentado melhor ajuste no treinamento, seus desempenhos foram reduzidos no teste, indicando menor capacidade de generalização. Esse comportamento sugere possível sobreajuste, especialmente considerando o número limitado de observações anuais disponíveis.

Tabela 1 - Desempenho dos Modelos Preditivos nos Conjuntos de Treino e Teste

Modelo	Conjunto	NSE	RMSE	MAE	R2
Árvore de decisão	Treino	0.461	24.619	12.795	0.461
	Teste	0.579	29.261	13.360	0.976
Random Forest	Treino	0.887	11.283	5.638	0.936
	Teste	0.290	38.000	17.479	0.925
SVM	Treino	0.961	6.637	5.020	0.964
	Teste	0.457	33.231	15.264	0.911
Rede Neural	Treino	0.962	6.548	5.009	0.962
	Teste	0.462	33.078	15.278	0.900

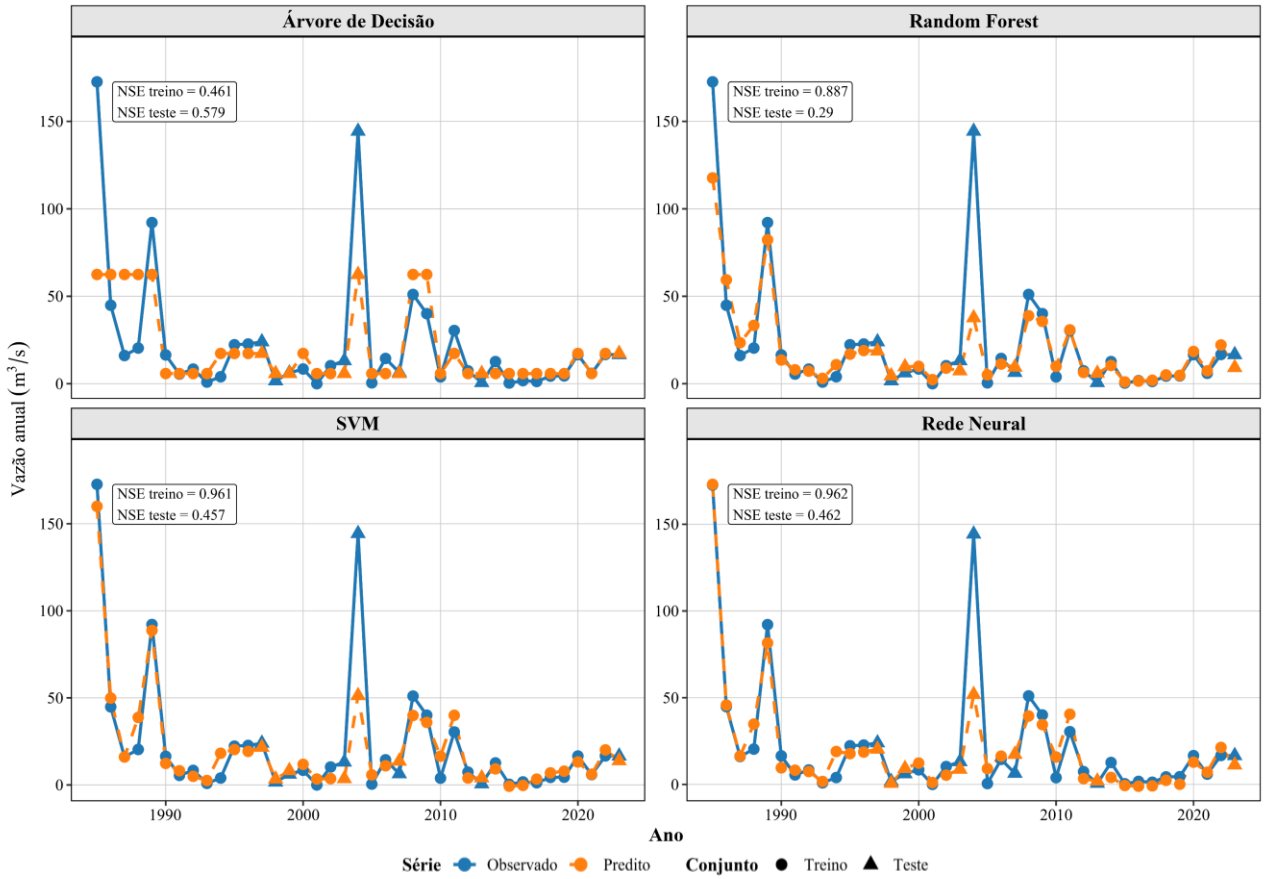


Figura 4 - Avaliação dos modelos na predição de vazão anual

A análise gráfica das vazões observadas e preditas (Figura 4) evidenciou que os modelos foram capazes de representar satisfatoriamente os períodos de baixa e média vazão. Entretanto, verificou-se dificuldade na reprodução dos eventos de maior magnitude, com tendência geral à subestimativa dos picos de vazão. Essa limitação foi mais acentuada nos modelos Random Forest, SVM e Rede Neural, os quais apresentaram maior suavização dos valores extremos.

Por outro lado, a Árvore de Decisão, embora também tenha demonstrado restrições na estimativa dos picos, apresentou desempenho relativamente superior no conjunto de teste. De modo geral, os resultados indicam que os modelos avaliados possuem boa capacidade de representação do comportamento hidrológico predominante, mas ainda apresentam limitações na previsão de eventos extremos.

CONSIDERAÇÕES FINAIS

Os resultados deste estudo evidenciaram que a precipitação anual foi a variável de maior influência na previsão da vazão média anual da Bacia Hidrográfica do Alto Jaguaribe, confirmando a forte dependência da dinâmica hidrológica em relação à variabilidade climática da região semiárida. As mudanças no uso e cobertura do solo, especialmente a redução de áreas florestadas e o avanço da agropecuária, também contribuíram para a compreensão da resposta hidrológica da bacia.

Entre os modelos avaliados, a Árvore de Decisão apresentou o melhor desempenho no conjunto de teste, indicando maior capacidade de generalização em comparação aos demais métodos. Embora Random Forest, SVM e Rede Neural tenham apresentado melhor ajuste no treinamento, seus resultados no teste sugerem possível sobreajuste, especialmente em razão do tamanho limitado da série histórica.

De modo geral, os modelos de aprendizado de máquina mostraram-se ferramentas promissoras para a previsão de vazão anual, sendo mais eficientes na representação de condições de baixa e média vazão. No entanto, ainda apresentaram limitações na estimativa de eventos extremos, reforçando a necessidade de ampliar as bases de dados e incorporar novas variáveis hidrológicas e climáticas em estudos futuros.

AGRADECIMENTOS

Agradecemos à Agência Nacional de Águas e Saneamento Básico (ANA), por meio do Termo de Execução Descentralizada (TED) ANA/UFC nº 06/2023/ANA, intitulado “Revisão, desenvolvimento e aprimoramento da regulação dos serviços hídricos no Brasil”, vinculado ao Processo Administrativo nº 02501.004221/2023-36, documento nº 5839824, pelo apoio ao desenvolvimento desta pesquisa.

REFERÊNCIAS

ADERA, S.; BELLUGI, D.; DHAKAL, A.; LARSEN, L. (2024). *“Streamflow prediction at the intersection of physics and machine learning: a case study of two Mediterranean-climate watersheds”*. *Water Resources Research* 60(7), e2023WR035790.

BAYDAROĞLU, Ö.; KOÇAK, K. (2014). *“SVR-based prediction of evaporation combined with chaotic approach”*. *Journal of Hydrology* 508, pp. 356 – 363.

BRAGA, R.R.; MATUSHIMA, E.R. (2021). *“Vale do Jaguaribe: um oásis em perigo no semiárido brasileiro – revisão sistemática sobre contaminação ambiental e potenciais danos à bacia e seus usuários”*. *Revista de Geociências do Nordeste* 7(2), pp. 24 – 34.

BREIMAN, L. (2001). *“Random forests”*. *Machine Learning* 45(1), pp. 5 – 32.

CORTES, C.; VAPNIK, V. (1995). *"Support-vector networks"*. *Machine Learning* 20(3), pp. 273 – 297.

FERNANDES, A.S.; MOREIRA, A.A.M. (2022). *"A pluviometria do quadrilátero ferrífero em Minas Gerais"*. *Revista Brasileira de Climatologia* 31, pp. 753 – 775.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. (2016). *Deep Learning*. MIT Press, Cambridge, 800 p.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. Springer, New York.

HAYKIN, S. (2009). *Neural Networks and Learning Machines*. 3. ed. Pearson, Upper Saddle River.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. 2. ed. Springer, New York.

KUMAR, M.; DENIS, D.M.; KUNDU, A.; JOSHI, N.; SURYAVANSHI, S. (2022). *"Understanding land use/land cover and climate change impacts on hydrological components of Usri watershed, India"*. *Applied Water Science* 12, 39.

LIAW, A.; WIENER, M. (2002). *"Classification and regression by randomForest"*. *R News* 2(3), pp. 18 – 22.

LOH, W.-Y. (2011). *"Classification and regression trees"*. *WIREs Data Mining and Knowledge Discovery* 1(1), pp. 14 – 23.

OLIVEIRA, T.A.; SOUZA FILHO, F.A.; VASCONCELOS JUNIOR, F.C. (2023). *"Análise da relação entre diferentes índices de seca: caso de estudo na Bacia do Alto Jaguaribe"*. *Revista Brasileira de Meteorologia* 38, e38230014, pp. 1 – 12.

QUINLAN, J.R. (1986). *"Induction of decision trees"*. *Machine Learning* 1(1), pp. 81 – 106.

RUMELHART, D.E.; HINTON, G.E.; WILLIAMS, R.J. (1986). *"Learning representations by back-propagating errors"*. *Nature* 323, pp. 533 – 536.

SCHÖLKOPF, B.; SMOLA, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge.

SILVA NETO, J.O. (2024). *"A estrutura superficial da paisagem das sub-bacias dos riachos Carrapateiras e Cipó - Bacia do Jaguaribe - Tauá - CE"*. Dissertação (Mestrado em Geografia), Universidade Federal do Ceará, Fortaleza, 94 f.

VAPNIK, V.N. (1995). *The Nature of Statistical Learning Theory*. Springer, New York.

VIANA, J.F. de S.; MONTENEGRO, S.M.G.L.; SRINIVASAN, R.; SANTOS, C.A.G.; MISHRA, M.; KALUMBA, A.M.; SILVA, R.M. da. (2023). *"Land Use and Land Cover Trends and Their Impact on Streamflow and Sediment Yield in a Humid Basin of Brazil's Atlantic Forest Biome"*. *Diversity* 15(12), 1220.

XU, T.; LIANG, F. (2021). "*Machine learning for hydrologic sciences: An introductory overview*". WIREs Water 8(5), e1533.