

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS / FACULDADE DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM  
COMPUTACIONAL**

**Alfeu Dias Martinho**

**Modelos de Aprendizado de Máquinas Híbridos Aplicados à previsão de curto  
prazo da Vazão do Rio Zambeze Afluente à Barragem Hidroelétrica de  
Cahora-Bassa, em Moçambique**

Juiz de Fora  
2023

Alfeu Dias Martinho

**Modelos de Aprendizado de Máquinas Híbridos Aplicados à previsão de curto prazo da Vazão do Rio Zambeze Afluente à Barragem Hidroelétrica de Cahora-Bassa, em Moçambique**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de Doutor em Modelagem Computacional.

Orientador: Prof. Dr. Leonardo Goliatt da Fonseca

Coorientador: Prof. Dr. Henrique Steinherz Hippert

Juiz de Fora

2023

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Martinho, Alfeu Dias .

Modelos de Aprendizado de Máquinas Híbridos Aplicados à previsão de curto prazo da Vazão do Rio Zambeze Afluente à Barragem Hidroelétrica de Cahora-Bassa, em Moçambique / Alfeu Dias Martinho. -- 2023.

186 f.

Orientador: Leonardo Goliatt da Fonseca

Coorientador: Henrique Steinherz Hippert

Tese (doutorado) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2023.

1. Modelagem Hidrológica. 2. Vazão Afluente. 3. Aprendizado de Máquina. 4. Algoritmos bioinspirados. 5. Modelo Híbrido. I. Fonseca, Leonardo Goliatt da , orient. II. Hippert, Henrique Steinherz , coorient. III. Título.

**Alfeu Dias Martinho**

**Modelos de Aprendizado de Máquinas Híbridos Aplicados à previsão de curto prazo da Vazão do Rio Zambeze Afluente à Barragem Hidroelétrica de Cahora-Bassa, em Moçambique**

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Modelagem Computacional. Área de concentração: Modelagem Computacional.

Aprovada em 24 de março de 2023.

**BANCA EXAMINADORA**

**Prof(a) Dr(a). Leonardo Goliatt da Fonseca** - Orientador

Universidade Federal de Juiz de Fora

**Prof(a) Dr(a). Henrique Steinherz Hippert** - Coorientador

Universidade Federal de Juiz de Fora

**Prof(a) Dr(a). Eliane da Silva Christo**

Universidade Federal Fluminense

**Prof(a) Dr(a). Camila Martins Saporetti**

Universidade do Estado do Rio de Janeiro

**Prof(a) Dr(a). Celso Bandeira de Melo Ribeiro**

Universidade Federal de Juiz de Fora

**Prof(a) Dr(a). Luciana Conceição Dias Campos**

**Prof(a) Dr(a). Carlos Cristiano Hasenclever Borges**

Universidade Federal de Juiz de Fora

Juiz de Fora, 23/03/2023.



Documento assinado eletronicamente por **Carlos Cristiano Hasenclever Borges, Professor(a)**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luciana Conceicao Dias Campos, Professor(a)**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leonardo Goliatt da Fonseca, Professor(a)**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Camila Martins Saporetti, Usuário Externo**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celso Bandeira de Melo Ribeiro, Professor(a)**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **ELIANE DA SILVA CHRISTO, Usuário Externo**, em 24/03/2023, às 11:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Henrique Steinherz Hippert, Professor(a)**, em 24/03/2023, às 13:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Uffj ([www2.uffj.br/SEI](http://www2.uffj.br/SEI)) através do ícone Conferência de Documentos, informando o código verificador **1200805** e o código CRC **0712C3D5**.

Dedico este trabalho à minha família e amigos.

## AGRADECIMENTOS

Quero agradecer em primeiro lugar à Deus pela Vida.

Ao meu orientador Prof. Dr. Leonardo Goliatt, pelo auxílio, atenção, paciência, confiança e ensinamentos.

Ao meu coorientador Prof. Dr. Henrique Hippert, pelo apoio, confiança e atenção.

Aos meus pais, Martinho Nimone e Fátima Valentim e meus irmãos pelos conselhos e incentivos que sempre me ofereceram sempre que precisei.

A todos meus amigos, que sempre estiveram ao meu lado.

A todos os professores e professoras que um dia me ensinaram e contribuíram para minha formação.

Ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora (UFJF), por possibilitar a realização deste trabalho.

Ao GCUB (Grupo Coimbra de Universidades Brasileiras) pelo apoio financeiro a esta pesquisa.

"O bom do caminho é haver volta. Para ida sem vinda basta o tempo."  
(Mia Couto)

## RESUMO

Os Rios são até hoje um dos maiores e mais importantes recursos para a sobrevivência da humanidade. Eles fornecem grande parte da água que consumimos e que usamos para produzir nossos alimentos e para nossa higiene. Além disso, a utilizamos para irrigar o solo das áreas agrícolas, para navegação, pesca, produção de energia elétrica, para além da manutenção da biodiversidade. Devido à escassez da água e da preocupação com a sua existência no futuro, torna-se indispensável o desenvolvimento de estudos que possam ajudar a compreender sua dinâmica, de modo a gerenciá-la. O rio Zambeze é o quarto maior curso de água do continente africano e tem vital importância, porque é a principal fonte de produção de energia elétrica de Moçambique, contribui para o desenvolvimento da economia do delta através da viabilização de atividades econômicas como agricultura, pastorícia, pesca, construção de vias de acesso e, na redução do risco de ocorrência de calamidades naturais, como a seca, cheias ou inundações. Apesar dessa necessidade, os estudos de recursos hídricos para essa região ainda são escassos. O presente trabalho apresenta o desenvolvimento de modelos híbridos de aprendizado de máquina para previsão de vazões naturais de corpos d'água do rio Zambéze na barragem Hidrelétrica de Cahora-Bassa, utilizando valores antecedentes de vazões afluentes, precipitação, evaporação e umidade relativa como variáveis de entrada. Cinco modelos foram considerados na análise: *Extreme Gradient Boosting (XGB)*, *Extreme Learning Machine (ELM)*, *Support Vector Regression (SVR)*, *Elastic Net linear (EN)* e *Multivariate Adaptive Regression Splines (MARS)*. Algoritmos de otimização evolutivos/bioinspirados: *Grey wolf optimization (GWO)*, *Genetic Algorithms (GA)*, *Differential Evolution (DE)* e *Particle Swarm Optimization (PSO)* foram utilizados como estratégias de seleção de todos os parâmetros internos dos modelos. E, além de avaliar o potencial dos cinco modelos de aprendizado de máquinas híbridos para simulação de fluxo, este trabalho também focou em dois métodos de seleção de variáveis de entradas, *Least Absolute Shrinkage and Selection Operator (LASSO)* e *Partial Mutual Information variable selection (PMI)*, usados para escolher o subconjunto que permite a melhor previsão dos valores futuros, e que fornecem informações abrangentes sobre a influência das variáveis no fluxo simulado. Os modelos híbridos desenvolvidos realizaram as previsões para 1, 3, 5 e 7 dias à frente em dados coletados na barragem Hidroelétrica de Cahora-Bassa, em Moçambique. Medidas de desempenho combinadas e testes estatísticos foram aplicados para avaliação da performance dos modelos. Técnicas de análise de incerteza baseadas em intervalos de confiança e simulação de Montecarlo também foram utilizadas para a análise de confiabilidade das previsões fornecidas pelos modelos. Os resultados mostram que todos os modelos obtiveram bons desempenhos para prever a vazão do rio vários passos à frente, indicando que a integração dos algoritmos evolutivos/bioinspirados é uma alternativa bem-sucedida para produzir previsões precisas da vazão. O modelo híbrido XGB com GWO integrado e suas respectivas variações com

seleção de variáveis de entrada, XGB-LASSO e XGB-PMI superaram os modelos híbridos EN, ELM, MARS e SVR, tendo apresentado a melhor performance, para 1, 3, 5 e 7 passos à frente. Esses modelos fornecem previsões confiáveis, e podem ser utilizados como uma alternativa para auxiliar na previsão de vazão, que é crucial para as atividades de gestão e planejamento sustentável ou otimizado de Recursos Hídricos em usinas hidroelétricas.

Palavras-chave: Modelagem Hidrológica. Vazão Afluente. Aprendizado de Máquina. Algoritmos bioinspirados. Modelo Híbrido.

## ABSTRACT

Rivers are still one of the greatest and most important resources for the survival of humanity. They provide much of the water we consume and that we use to produce our food and for our hygiene. In addition, we use it to irrigate the soil in agricultural areas, for navigation, fishing, electricity production, in addition to maintaining biodiversity. Due to water scarcity and concern about its existence in the future, it becomes essential to develop studies that can help to understand its dynamics, in order to manage it. The Zambezi River is the fourth largest watercourse on the African continent and is of vital importance in the region, as it is the main source of electricity production in Mozambique, contributes to the development of the economy of the delta through the viability of economic activities such as agriculture, pastoralism, fishing, construction of access roads and, in reducing the risk of occurrence of natural disasters, such as drought and floods. Despite this need, studies of water resources for this region are still scarce. The present work presents the development of hybrid machine learning models for predicting natural flows of water bodies of the Zambezi River at the Cahora-Bassa Hydroelectric dam, using antecedent values of influent flows, precipitation, evaporation and relative humidity as input variables. Five models were considered in the analysis: Extreme Gradient Boosting (XGB), Extreme Learning Machine (ELM), Support Vector Regression (SVR), Elastic Net linear (EN) and Multivariate Adaptive Regression Splines (MARS). Evolutionary/bioinspired optimization algorithms: *Grey wolf optimization (GWO)*, *Genetic Algorithms (GA)*, *Differential Evolution (DE)* and *Particle Swarm Optimization (PSO)* were used as selection strategies for all parameters internals of the models. And, in addition to evaluating the potential of the five hybrid machine learning models for flow simulation, this work also focused on two input variable selection methods, *Least Absolute Shrinkage and Selection Operator (LASSO)* and *Partial Mutual Information variable selection (PMI)*, used to choose the subset that allows the best prediction of future values, and that provide comprehensive information about the influence of variables on the simulated flow. The developed hybrid models performed predictions for 1, 3, 5 and 7 days ahead on data collected at the Cahora-Bassa Hydroelectric dam in Mozambique. Combined performance measures and statistical tests were applied to evaluate the performance of the models. Uncertainty analysis techniques based on confidence intervals and Montecarlo simulation were also used to analyze the reliability of forecasts provided by the models. The results show that all models performed well to predict river discharge several steps ahead, indicating that the integration of evolutionary/bioinspired algorithms is a successful alternative to produce accurate discharge predictions. The hybrid model XGB with integrated GWO and its respective variations with selection of input variables, XGB-LASSO and XGB-PMI outperformed the hybrid models EN, ELM, MARS and SVR, having presented the best performance, respectively, for 1, 3, 5, and 7 steps ahead. These models provide reliable

forecasts, and can be used as an alternative to aid in flow forecasting, which is crucial for sustainable or optimized management and planning of water resources in hydroelectric power plants.

Keywords: Hydrological Modeling. Affluent Flow. Machine Learning. Bioinspired algorithms. Hybrid Model.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Armazenamento percapita de água em alguns países. . . . .	22
Figura 2 - Número total de Eventos Extremos ocorridos entre 1980-2019 . . . . .	23
Figura 3 - Ciclo Hidrológico, mostrando os processos Escoamento superficial, Evaporação e Evapotranspiração, precipitação e infiltração. . . . .	25
Figura 4 - Régua de Medição de escalas hidrométricas, Boroma-Tete. . . . .	31
Figura 5 - Objetivos de Desenvolvimento sustentável(ODS). . . . .	34
Figura 6 - Limites da Bacia Hidrográfica do Zambeze. . . . .	56
Figura 7 - Localização da Área de estudo. Os pontos EMAs indicam as estações de monitoramento automáticas onde são recolhidos dados em análise neste trabalho. . . . .	57
Figura 8 - Dados históricos diários de vazões do rio Zambeze na barragem de Cahora Bassa. Onde $Q$ é vazão afluente, $R$ a precipitação, $E$ a Evaporação e $H$ a umidade relativa. Dados diários, entre 2003 e 2018 (15anos). De 01/01/2003 a 30/06/2012 treino (azul) (63.3%) e 01/07/2012 a 31/12/2018 teste (laranja) (36.7%). . . . .	59
Figura 9 - Análise da sazonalidade das séries estudadas. $Q$ é vazão afluente, $R$ a precipitação, $E$ a Evaporação e $H$ a umidade relativa. . . . .	60
Figura 10 - Exemplo de um <i>datalogger</i> CR1000 e uma Estação de monitoramento autónoma de Emboque/Chicoa (Cahora-Bassa). . . . .	64
Figura 11 - Representação do esquema de rede do ELM . . . . .	70
Figura 12 - Estrutura do modelo SVR. . . . .	74
Figura 13 - Arquitetura de uma máquina de regressão construída pelo SVR. . . . .	76
Figura 14 - Nós e splines lineares para um exemplo de MARS simples . . . . .	81
Figura 15: Fluxograma mostrando Metodologia Proposta . . . . .	94
Figura 16: Esquema de validação <i>5-Fold walk forward validation</i> . . . . .	106
Figura 17 - Funções de autocorrelação e autocorrelação parcial. As defasagens que se encontram dentro da parte sombreada são consideradas estatisticamente não-significativas. . . . .	111
Figura 18 - Funções de correlação cruzada entre vazão e precipitação, evaporação ou umidade relativa. As defasagens que se encontram entre as linhas tracejadas são consideradas estatisticamente não-significativas. . . . .	112
Figura 19 - Gráficos de Violino para GWO. . . . .	118
Figura 20 - Melhor solução de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente para GWO. . . . .	119
Figura 21 - Gráficos de Violino para DE. . . . .	122
Figura 22 - Melhor solução de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente para DE. . . . .	123
Figura 23 - Gráficos de Violino para GA. . . . .	126

Figura 24 - Melhor solução para cada modelo de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente com GA. . . . .	127
Figura 25 - Gráficos de Violino para PSO. . . . .	130
Figura 26 - Melhor solução para cada modelo de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente com PSO. . . . .	131
Figura 27 - Distribuição dos Parâmetros de XGB com GA. . . . .	137
Figura 28 - Distribuição dos Parâmetros de XGB com DE. . . . .	137
Figura 29 - Distribuição dos Parâmetros de XGB com GWO. . . . .	138
Figura 30 - Distribuição dos Parâmetros de XGB com PSO. . . . .	138
Figura 31 - Número médio de variáveis para XGB-LASSO e XGB-PMI. As barras de erro mostram o desvio padrão. . . . .	144
Figura 32 - Diagramas de Taylor ilustrando modelos com várias etapas à frente: $t + 1$ , $t + 3$ , $t + 5$ e $t + 7$ . . . . .	147
Figura 33 - Distribuição de parâmetros do XGB em 30 execuções. . . . .	151
Figura 34 - Distribuição de parâmetros do XGB-LASSO em 30 execuções. . . . .	152
Figura 35 - Distribuição de parâmetros do XGB-PMI em 30 execuções. . . . .	152
Figura 36 - Análise de incertezas e erros. . . . .	157

## LISTA DE TABELAS

Tabela 1 – Resumo de métodos de aprendizado de máquinas para previsão de vazões. .....	42
Tabela 2 – Resumo de métodos de aprendizado de máquinas híbridos obtidos por combinação de modelos padrões para previsão de vazões. ....	45
Tabela 3 – Resumo de métodos de aprendizado de máquinas híbridos por otimização de parâmetros para previsão de vazões. ....	49
Tabela 4 – Resumo de métodos de aprendizado de máquinas híbridos com aplicação de métodos de seleção de variáveis de entrada ....	52
Tabela 5 – Funções de ativação usadas no ELM. ....	71
Tabela 6 – Treinamento do ELM. ....	72
Tabela 7 – Treinamento do XGB. ....	79
Tabela 8 – Pseudocódigo do algoritmo GA ....	84
Tabela 9 – Pseudocódigo do algoritmo GWO ....	87
Tabela 10 – Pseudocódigo do algoritmo DE ....	89
Tabela 11 – Pseudocódigo do algoritmo PSO ....	91
Tabela 12 – Codificação de soluções candidatas. A coluna IP indica o parâmetro interno na codificação dos algoritmos bioinspirados. ....	113
Tabela 13 – Descrição dos parâmetros específicos dos algoritmos de otimização . . .	113
Tabela 14 – Informações estatísticas do conjunto de treinamento. A precipitação é indicada por $R$ , afluente do vazão natural da barragem de Cahora Bassa por $Q$ , $H$ é a umidade relativa e a evaporação por $E$ ....	114
Tabela 15 – Informação estatística do conjunto de teste. ....	114
Tabela 16 – <b>Média e desvio padrão das medidas de desempenho.</b> Desvios padrão entre parênteses. Seguindo o modelo apresentado na Eq. 4.1, a primeira linha indica o número de dias à frente que o modelo híbrido realizou a previsão. A segunda coluna exibe os modelos de aprendizado de máquina integrados à abordagem híbrida com GWO. As colunas restantes resumem as métricas alcançadas pelo modelo. Um total de 30 execuções foram realizadas para cada entrada na tabela. ....	117
Tabela 17 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados com DE ....	121
Tabela 18 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados pelo GA ....	125
Tabela 19 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados com PSO ....	129
Tabela 20 – $p$ -valores do teste ANOVA para cada medida de Desempenho ....	132

Tabela 21 – Teste de Comparações Múltiplas de Médias das metaheurísticas - Tukey HSD, Significância=0.05 . . . . .	133
Tabela 22 – Teste de Comparações Múltiplas de Médias dos modelos - Tukey HSD, Significância=0.05 . . . . .	134
Tabela 23 – Comparações de modelos híbridos gerados por combinações de modelos e Metaheurísticas em análise . . . . .	134
Tabela 24 – Codificação de soluções potenciais. A coluna IP representa o parâmetro interno na codificação GWO. . . . .	141
Tabela 25 – Métricas de média e desvio padrão. Desvios padrão entre parênteses. DA indica o número de dias antes do modelo híbrido realizar a previsão na primeira coluna. A segunda coluna exibe os modelos de aprendizado de máquina integrados à abordagem híbrida. O acrônimo métrico é mostrado na primeira linha. As colunas restantes resumem as métricas alcançadas pelo modelo. Um total de 30 execuções foram realizadas para cada entrada na Tabela . . . . .	143
Tabela 26 – Variáveis selecionadas para os melhores modelos XGB-LASSO e XGB-PMI para diferentes horizontes de previsão. A vazão natural é indicada por $Q$ , a precipitação por $R$ , $H$ é a umidade relativa e $E$ é a evaporação. . . . .	145
Tabela 27 – Resumo de métodos híbridos computacionais para previsão de vazões fluviais. . . . .	148
Tabela 28 – Estimativas de incerteza para modelos de aprendizado de máquina. . . . .	154
Tabela 29 – Análise de incerteza para o conjunto de teste usando Simulação de Monte Carlo (MCS). . . . .	155

## LISTA DE ABREVIATURAS E SIGLAS

ACF	Funções de Autocorrelação
ANN	Artificial Neural Networks
ANN-BP	Artificial Neural Network with backpropagation
ANFIS	adaptive neural-based fuzzy inference system
ARA	Administração Regional de Águas
ARMA	Autoregressive moving-average
ARIMA	Autoregressivos integrados de Médias Móveis
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BNN	Bayesian neural networks,
CCF	Função de Correlação Cruzada
CART	CART regression trees
DWAF	Department of Water Affairs of Zambia
DE	Diferential Evolution
DE-SVR	SVR com parâmetros otimizados por DE
DENFIS	Sistema de Inferência Neural-Fuzzy de Evolução Dinâmica
E	Evaporação
ECMWF	European Center for Medium Range Weather Forecast
ELM	Extreme learning machine
EN	Elastic net linear model
GA	Genetic Algorithm
GA-SVR	SVR com parâmetros otimizados por GA
GCV	Generalized cross-validation
GPR	Gaussian Process Models
GWO	Gray Wolf Optimization Algorithm
GP	Genetic programming
GRNN	Generalized regression neural networks,
HCB	Hidroelétrica de Cahora-Bassa
H	Umidade relativa
IA	Inteligência Artificial
ITCZ	Intertropical Convergence Zone
KNN	K-nearest neighbor regression,
KGE	Eficiência de Kling-Gupta
KNBPS	Kariba North Bank Power Station
KSBPS	Kariba South Bank Power Station
LSSVM	Least squares support vector machine
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	long short-term memory network,
MAE	Erro médio absoluto
MARS	Multi-Adaptative spline Regression
MAPE	Erro Percentual Absoluto Médio

ML	Machine Learning
MLR	Regressão linear múltipla
MLP	Multi-Layer Perceptron
M5tree	Árvore modelo M5
NSE	Eficiência de Nash-Sutcliffe
NWS-CPC	National Weather Service – Climate Prediction Center
ODS	Objetivos de desenvolvimento sustentável
PIB	Produto interno Bruto
PPR	Pursuit regression
PACF	Funções de Autocorrelação Parcial
PSO	Particle Swarm Optimization
PMI	Partial Mutual Information variable selection
Q	Vazões afluentes
RMSE	Raiz do Erro Quadrático Médio
R	Precipitação
RNA	Rede Neural Artificial
RF	Random Forest
RBF	Radial basis functions,
SWAT	Soil and Water Assessment Tool
SVR	Support vector regression
SMAP	Soil Moisture Accounting Procedure
SAC	Sistema de Aviso Cheias
SLSQP	Least square sequential least square programming
SARCOF	Southern African Regional Climate Outlook Forum,
TOPMODEL	Topography-based Hydrological Model
WI	Índice de Willmot
XBG	Extreme gradient boosting
XGB-LASSO	Modelo híbrido Extreme Gradient Boosting assistido por LASSO
XGB- PMI	Modelo híbrido Extreme Gradient Boosting assistido por PMI
ZINWA	Zimbabwe National Water Authority,
ZRA	Zambezi River Authority
ZESCO	Zambia Electricity Supply Company

## LISTA DE SÍMBOLOS

$\forall$	Para todo
$\in$	Pertence
$t_N$	Instante de tempo
$Z_{t_N}$	Série temporal
$\ \cdot\ _2$	Norma $L_2$
$\ \cdot\ _1$	Norma $L_1$
$R^n$	Espaço Real de dimensão $n$
$H^+$	Pseudoinversa
$\Sigma$	Somatório
$\prod$	Produto
$\vec{D}$	Vetor
$km, km^2, km^3$	kilometro, kilometro quadrado e kilometro cúbico
$m^3$	Metro cúbico
$Mm^3$	Mega Metro cúbico $10^6 m^3$
$mm$	Milímetro
$m^3/s$	Metro cúbico por segundo
$L$	Número de defasagens ou antecedentes
$\int$	Integral
$\iint$	Integral dupla
$O_i$	Dados observados
$P_i$	Valores previstos,
$\bar{O}$	Média dos fluxos observados
$\#$	Cardinal
$\%$	Porcentagem
$\bar{e}$	Média dos logaritmos dos erros de previsão
$S_e$	Desvio padrão dos logaritmos dos erros de previsão
$N$	Tamanho da amostra de dados
$t + j$	Horizonte de previsão $j$ passos a frente
$F$	Função que representa a relação de entrada/saída
$\phi(\mathbf{x})$	Função kernel
$m$	Metros
$H(X)$	Entropia da variável aleatória $X$
$H(X, Y)$	Entropia conjunta de $(X, Y)$
$H(Y X)$	Entropia condicional de $Y$
$p(X), P(X, Y)$	Distribuição de probabilidades de $X$ , distribuição conjunta de probabilidades de $(X, Y)$
$\Sigma\Sigma$	Soma dupla

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>21</b>
1.1	CONTEXTUALIZAÇÃO . . . . .	21
1.1.1	POLÍTICAS E ESTRATÉGIAS NACIONAIS DE GESTÃO DE RECURSOS HÍDRICOS . . . . .	23
1.1.2	MODELAGEM HIDROLÓGICA . . . . .	24
1.1.3	HORIZONTE DE PREVISÃO . . . . .	28
1.1.4	MODELAGEM HIDROLÓGICA NO RIO ZAMBEZE EM MOÇAMBIQUE	29
1.2	MOTIVAÇÃO . . . . .	31
1.3	OBJETIVOS . . . . .	34
1.3.1	OBJETIVO GERAL . . . . .	34
1.3.2	OBJETIVOS ESPECÍFICOS . . . . .	34
1.4	ORGANIZAÇÃO DO TEXTO . . . . .	35
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b> . . . . .	<b>36</b>
2.1	MÉTODOS DE APRENDIZADO DE MÁQUINAS PARA PREVISÃO DA VAZÃO . . . . .	36
2.2	MÉTODOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS PARA PRE- VISÃO DA VAZÃO . . . . .	42
2.2.1	MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS POR COMBI- NAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINAS INDIVIDU- AIS . . . . .	43
2.2.2	MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS POR OTIMI- ZAÇÃO DE HIPERPARÂMETROS . . . . .	45
2.3	MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS COM APLI- CAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS . . . . .	50
2.4	CONTRIBUIÇÃO DA TESE . . . . .	53
<b>3</b>	<b>MATERIAIS E MÉTODOS</b> . . . . .	<b>54</b>
3.1	ÁREA DE ESTUDO . . . . .	54
3.2	CONJUNTO DE DADOS . . . . .	57
3.3	MODELOS DE SÉRIES TEMPORAIS EM HIDROLOGIA . . . . .	64
3.4	MODELOS DE APRENDIZADO DE MÁQUINAS PARA PREVISÃO HI- DROLÓGICA . . . . .	67
3.4.1	<i>ELASTIC NET</i> (EN) . . . . .	68
3.4.2	<i>EXTREME LEARNING MACHINE</i> (ELM) . . . . .	69
3.4.3	<i>SUPPORT VECTOR REGRESSION</i> (SVR) . . . . .	73
3.4.4	<i>EXTREME GRADIENT BOOSTING</i> (XGB) . . . . .	76
3.4.5	<i>MULTIVARIATE ADAPTIVE REGRESSION SPLINE</i> (MARS) . . . . .	79

3.5	AJUSTE DE PARÂMETROS DOS MODELOS DE PREVISÃO PROPOSTOS . . . . .	81
3.5.1	<i>GENETIC ALGORITHM</i> (GA) . . . . .	83
3.5.2	<i>GREY WOLF OPTIMIZATION</i> (GWO) . . . . .	84
3.5.3	<i>DIFERENTIAL EVOLUTION</i> (DE) . . . . .	87
3.5.4	<i>PARTICLE SWARM OPTIMIZATION</i> (PSO) . . . . .	90
4	<b>METODOLOGIA DO TRABALHO . . . . .</b>	<b>92</b>
4.1	ESTRUTURA DE MODELAGEM . . . . .	95
4.2	CONFIGURAÇÃO DE ATRASO/ANTECEDENTES PARA PREVISÃO E SELEÇÃO DE VARIÁVEIS . . . . .	95
4.2.1	<i>LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR</i> (LASSO) . . . . .	96
4.2.2	<i>PARTIAL MUTUAL INFORMATION VARIABLE SELECTION</i> (PMI) . . . . .	97
4.3	PREVISÃO MULTIPASSOS À FRENTE DE VAZÃO . . . . .	100
4.4	ESTIMAÇÃO DO DESEMPENHO DOS MODELOS . . . . .	101
4.4.1	TESTE DE HIPÓTESES ESTATÍSTICAS . . . . .	103
4.5	<i>WALK FORWARD VALIDATION</i> . . . . .	104
4.6	ANÁLISE DE INCERTEZA . . . . .	107
5	<b>EXPERIMENTOS COMPUTACIONAIS . . . . .</b>	<b>109</b>
5.1	DESCRIÇÃO DOS EXPERIMENTOS COMPUTACIONAIS . . . . .	109
5.2	RESULTADOS E DISCUSSÃO . . . . .	113
5.2.1	ANÁLISE DESCRITIVA . . . . .	114
5.3	RESULTADOS E DISCUSSÃO DAS SIMULAÇÕES . . . . .	114
5.3.1	ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS POR GWO	115
5.3.2	ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELO DE	120
5.3.3	ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELOS GA . . . . .	124
5.3.4	ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELO PSO	128
5.3.5	ANÁLISE COMPARATIVA DOS RESULTADOS . . . . .	132
5.3.5.1	ANÁLISE DAS DISTRIBUIÇÕES DOS PARÂMETROS INTERNOS DOS MODELOS DE APRENDIZADO DE MÁQUINAS PROPOSTOS . . . . .	136
5.4	RESULTADOS E DISCUSSÃO DOS MODELOS COM SELEÇÃO DE VARIÁVEIS . . . . .	138
5.4.1	DISTRIBUIÇÃO DOS PARÂMETROS INTERNOS PARA O MÉTODO XGB . . . . .	150
5.5	ANÁLISE DE INCERTEZA . . . . .	153
6	<b>CONCLUSÃO . . . . .</b>	<b>158</b>
6.1	PERSPECTIVAS DE TRABALHOS FUTUROS . . . . .	161
6.2	TRABALHOS PUBLICADOS . . . . .	162
	<b>REFERÊNCIAS . . . . .</b>	<b>163</b>

# 1 INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO

Os Rios são até hoje um dos maiores e mais importantes recursos para a sobrevivência da humanidade. São eles que fornecem grande parte da água que consumimos, que usamos para produzir nossos alimentos, de que necessitamos para nossa higiene, que utilizamos para irrigar o solo das áreas agrícolas, para a navegação, a pesca e a produção de energia eléctrica, para além da manutenção da biodiversidade.

Segundo a Organização Mundial de Saúde (OMS), estima-se que cada ser humano consome direta ou indiretamente quarenta litros de água por dia, no entanto, a quantidade de água que o ser humano precisa para viver com conforto e saúde é de 110 litros por dia. Aproximadamente 70% da água consumida no Mundo vai para a irrigação de campos agrícolas, 20% são usados na indústria e os restantes 10% nas nossas casas (ANJO, 2015). Porém, apenas 2.5% da água do planeta é doce e própria para o uso, requerendo o uso desse recurso de forma sustentável (OLIVEIRA; HENKES, 2015; PINTO, 2017).

Moçambique possui 104 bacias hidrográficas de rios principais, das quais 9 são de rios transfronteiriços, correspondendo a um escoamento superficial total de cerca de  $216Km^3$ , dos quais cerca de  $100Km^3$  são gerados no país, e os restantes  $116Km^3$  nos países vizinhos (BRITO et al., 2008; MUAIEVELA, 2011).

Cada bacia hidrográfica apresenta suas características particulares no que diz respeito às potencialidades de navegação e turismo, extensão territorial, produção e distribuição de energia eléctrica, irrigação, entre outras. Dentre elas, a geração de energia eléctrica constitui a principal atividade.

A localização de Moçambique a jusante <sup>1</sup> de quase todos os rios daquela região revela a dependência do mesmo de água proveniente dos países relativamente a montante <sup>2</sup>. O monitoramento dos volumes de entrada de água nas bacias vindos dos países à montante deve merecer alta prioridade no gerenciamento dos recursos hídricos. O rebaixamento desses volumes de entrada pode trazer consequências negativas (MUAIEVELA, 2011).

As oscilações das vazões dos rios ao longo do ano são condicionadas por fatores climáticos e do relevo, podendo registrar as máximas na época das chuvas e as mínimas na estação seca, sendo a última a época a mais longa.

Essas variações das vazões dos rios trazem sérias dificuldades para o seu uso contínuo sem armazenamento, tornando as barragens indispensáveis para o aproveitamento

<sup>1</sup> Jusante é o fluxo normal da água, de um ponto mais alto para um ponto mais baixo. A jusante é o lado para onde se dirige a corrente de água. Por isso, diz-se que a foz de um rio é o ponto mais a jusante deste rio.

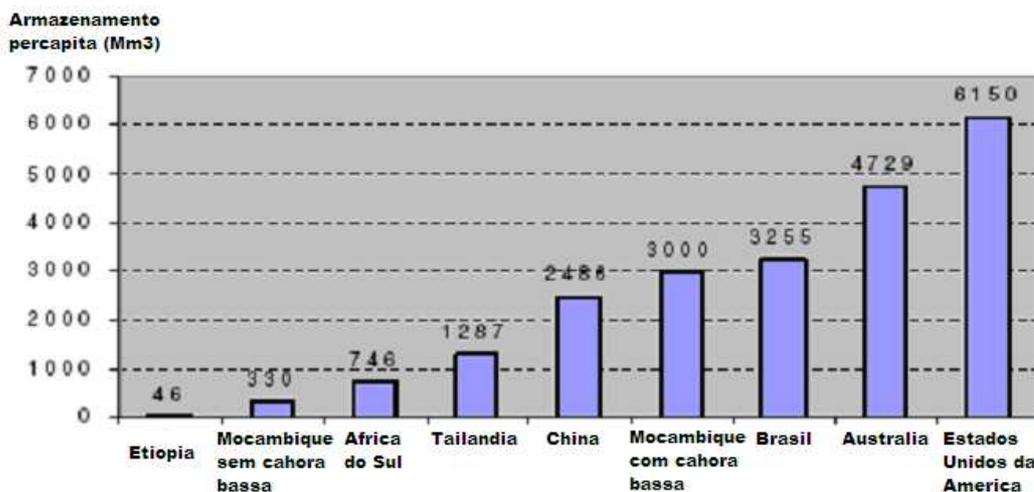
<sup>2</sup> Montante é a direção de um ponto mais baixo para o mais alto. Montante é parte onde nasce o rio. Por isso, diz-se que a nascente é o seu ponto mais a montante

dos recursos hídricos de Moçambique.

A capacidade de armazenamento total do país é cerca de  $56.000Mm^3$ , com capacidade usual de  $45.000Mm^3$ , o que corresponde a 21% da média anual do escoamento dos rios do país (BARROS et al., 1997).

A distribuição dessa capacidade de armazenamento é irregular, cerca de 90% da capacidade do país encontra-se na barragem de Cahora Bassa, no rio Zambéze (BARROS et al., 1997). A Figura 1 evidencia este aspecto ao apresentar o armazenamento per capita, onde se pode observar que excluindo a barragem de Cahora Bassa, a água armazenada disponível por pessoa e por ano é de cerca de  $330m^3$ , contra os  $3000(Mm^3)$  com inclusão de Cahora-Bassa.

Figura 1 - Armazenamento percapita de água em alguns países.

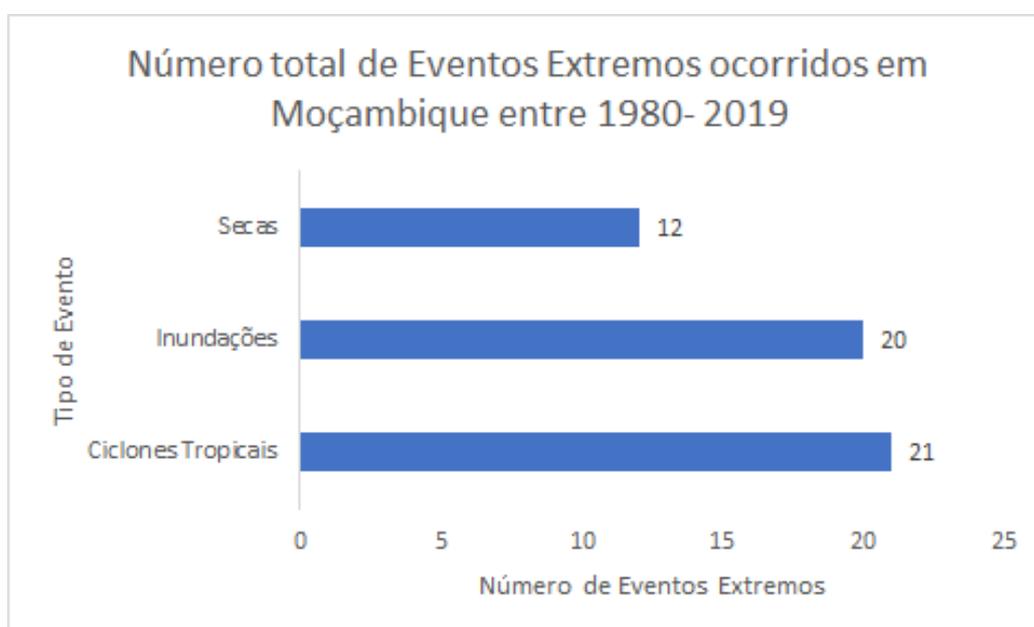


Fonte: Extraído de Barros (2009), Muaievela (2011)

No entanto, Moçambique enfrenta nos últimos anos vários problemas ao registrar secas severas e, por consequência, suas barragens têm atingido níveis de armazenamento muito abaixo do normal na região sul, devido não só a escassez de chuvas, mas também, as questões relacionadas a gestão da água nos países à montante. Por outro lado, inundações ou cheias, tempestades fortes ou ciclones, tem ocorrido nas regiões centro e norte do país, devido aos movimentos sazonais da zona de convergência intertropical que influenciam o estado do tempo. Estes fenômenos afetam o abastecimento de água, a agricultura e pastoreio, causam insegurança, perdas de vidas e a destruição das infraestruturas, agravando a pobreza absoluta das populações e, por conseguinte, retrocedendo o desenvolvimento da economia do país (BRITO et al., 2008).

A Figura 2 mostra a vulnerabilidade de Moçambique à eventos extremos como secas, inundações e ciclones tropicais.<sup>3</sup> Pode-se verificar que em média, o país é afectado por um ciclone tropical ou um evento de inundações a cada dois anos e um evento de seca a cada três anos. Os ciclones tropicais e eventos de inundações representam cerca de 77% dos eventos ocorridos. O impacto direto destes eventos é frequentemente expresso pela quantidade de vidas humanas perdidas, pela perda de bens pessoais e meios de subsistência, pela destruição de infra-estruturas críticas do país tais como estradas, pontes, sistema de abastecimento de água, escolas, hospitais, assim como pela eclosão de doenças transmitidas pela água ( malária, cólera, diarreias, entre outras).

Figura 2 - Número total de Eventos Extremos ocorridos entre 1980-2019



Fonte: Extraído de <https://www.ctc-n.org/sites/>

O monitoramento do comportamento hidrológico dos rios através da previsão dos níveis de afluência é por isso de extrema importância para a gestão de recursos hídricos e planejamento. A previsão de curto a longo prazo, diária, semanal, mensal, sazonal ou mesmo anual é particularmente útil em operações de reservatório e gestão de irrigação, bem como aspectos institucionais e legais da gestão e planejamento de recursos hídricos.

### 1.1.1 POLÍTICAS E ESTRATÉGIAS NACIONAIS DE GESTÃO DE RECURSOS HÍDRICOS

Devido à extrema importância da água, e face ao carácter de escassez atribuído à ela (PAZ, 2004), atualmente são notáveis os esforços do estado Moçambicano na criação e consolidação de políticas ou estratégias com vista ao reconhecimento da importância de preservação e uso racional desse recurso.

<sup>3</sup> <https://www.ctc-n.org/sites/>

O futuro desejado em relação à água é aquele onde esta esteja disponível em quantidade e qualidade adequadas para as gerações atuais e futuras, servindo para o desenvolvimento sustentável, redução da pobreza e promoção do bem-estar e paz, e estejam minimizados os efeitos negativos das cheias e secas (BRITO et al., 2008). A água tem uma presença direta ou indireta na maioria dos objetivos de desenvolvimento sustentável(ODS) (ANJO, 2015).

A Lei das águas de 1991 e respectivos regulamentos, a Estratégia nacional de gestão de recursos de 2007, Política de água de 2007 e os acordos de gestão da bacias hidrográficas compartilhadas com os países vizinhos são exemplos de instrumentos constituídos ou ratificados em Moçambique cujas metas compreendem a satisfação das necessidades básicas do consumo humano, o melhoramento do saneamento, a utilização eficiente da água para desenvolvimento económico, para conservação ambiental, a redução da vulnerabilidade às secas e cheias, promoção da paz e integração regional, bem como para garantir os recursos hídricos para o desenvolvimento sustentável de Moçambique (BRITO et al., 2008).

Porém, o cumprimento dessas metas depende da implementação de projetos prioritários e planos de gestão de recursos hídricos que incluem a avaliação, monitoramento dos recursos hídricos, análise e gestão de desastres, gestão de rios internacionais e consolidação das instituições de gestão de recursos hídricos. Para estes fins a modelagem hidrológica deve ser considerada um elemento fundamental a se ter em conta.

Os modelos hidrológicos são representações matemáticas capazes de reproduzir os fenômenos hídricos dentro de uma escala de tempo definida, estimando componentes do fluxo hidrológico anteriormente desconhecidos ou não quantificados. São úteis na gestão ambiental, com intuito de lidar com problemas ambientais, avaliar impactos futuros e subsidiar planeamentos (ALMEIDA; SERRA, 2017; DUARTE et al., 2019).

### 1.1.2 MODELAGEM HIDROLÓGICA

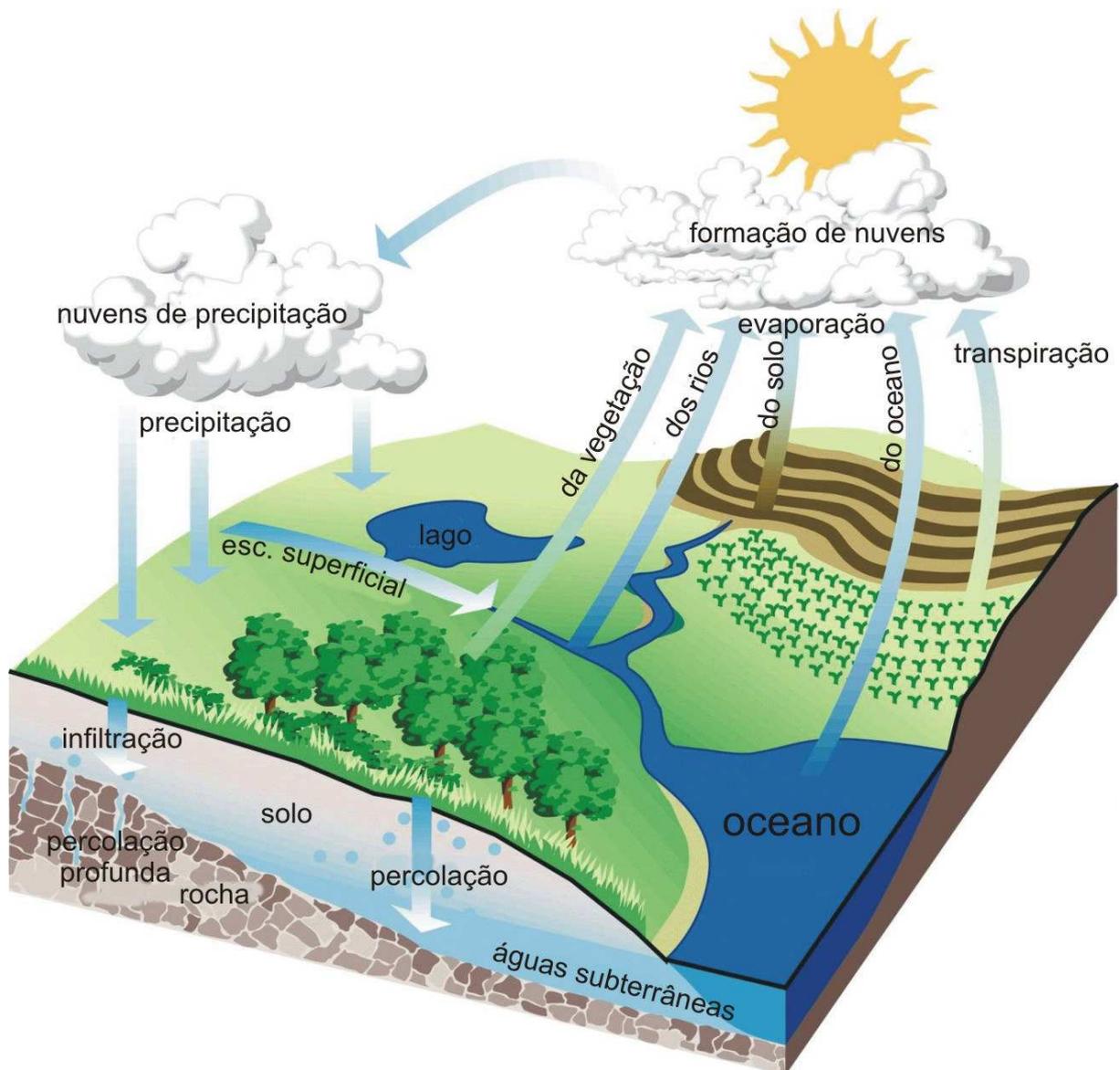
A Hidrologia é uma ciência que lida com a ocorrência, circulação e distribuição das águas na Terra, suas propriedades físicas e químicas, sua interação com o meio ambiente e em todas formas de vida ou seja, o estudo de precipitação e escoamento associados a sistemas de recursos hídricos (abastecimento de água, geração de energia, controle de enchentes, entre outros) (NETO et al., 2007; PAZ, 2004). Portanto, constitui uma ciência que estuda os fenômenos naturais complexos do ciclo hidrológico.

O ciclo hidrológico compreende todos os processos de formação, transferência e acúmulo de água. Nestes estão envolvidos a *precipitação* que abrange toda a água que cai na superfície da Terra, seja na forma de chuva, neve, granizo ou orvalho. O *escoamento superficial* é o resultado do escoamento (fluxo) da água na superfície provocado pelas chuvas, gerando rios, córregos e lagos até atingir o mar. A água pode também *infiltrar* no solo, abastecendo os lençóis e os corpos d'água superficiais durante os períodos secos. E

ainda, dependendo da temperatura e da umidade do ar, a água presente na superfície pode evaporar, gerando a *evaporação*. As plantas, que por diferença de pressão e temperatura perdem água pelos estômatos localizados na superfície da folha, em um processo chamado *transpiração*, que ocorre também nos animais, com intuito de equilibrar a temperatura corporal. Ao conjunto destes fenômenos de evaporação e transpiração dá-se o nome de *evapotranspiração*, importante no ciclo, pois devolve parte da água para a atmosfera.

A Figura 3 mostra o ciclo hidrológico com o resumo dos processos envolvidos.

Figura 3 - Ciclo Hidrológico, mostrando os processos Escoamento superficial, Evaporação e Evapotranspiração, precipitação e infiltração.



A natureza complexa do ciclo hidrológico e da sua relação com o clima, tipos de solo, topografia e geologia fazem com que a análise quantitativa e qualitativa dos seus processos como precipitação, evaporação, infiltração e o escoamento em rios se torne um desafio, requerendo uma abordagem multidisciplinar. Diversos especialistas, entre eles os engenheiros, economistas, estatísticos, químicos, biólogos, matemáticos, geólogos, agrônomos, geógrafos, entre outros, contribuem em suas áreas para entender a situação e alcançar a melhor alternativa, sob determinados critérios (PAZ, 2004).

De acordo com Villela e Mattos (1975), o escoamento superficial é para a engenharia, um dos mais importantes fenômenos do ciclo hidrológico. A sua formação, porém é um dos temas mais complexos da hidrologia, porque envolve as diversas características da bacia. A operação de um reservatório de geração de energia elétrica é definida com base nas vazões afluentes da bacia hidrográfica. Estas vazões são aleatórias, suas estimativas estão sujeita a riscos que se refletem na operação da usina. Para reduzir estes riscos é necessário prever as vazões afluentes de forma confiável.(BRUN; TUCCI, 2001)

Quando um modelo hidrológico é utilizado para a tomada de decisão operativa tem que apresentar alta confiabilidade e grande estabilidade, mesmo quando atuando em um sistema onde entradas, processamento (o próprio modelo) e saídas apresentam algum desvio da realidade. Visto que as previsões são parte de sistemas de tomada de decisão, erros nestes valores causam menor eficiência nas decisões tomadas, o que resulta em perdas, geralmente econômicas ou na pior das hipóteses, em perda de vidas. A minimização dos erros de previsão e estimativa da incerteza associada as previsões são importantes para enriquecer o processo de tomada de decisão, ponderando a relevância da informação contida nesta previsão (GUETTER, 2005).

O desenvolvimento de métodos com vista a alta confiabilidade das suas estimativas é uma atividade bastante complexa, dada a variabilidade das escalas temporal e espacial e a característica de não-linearidade das variáveis que delineiam os processos envolvidos (MARACAJÁ et al., 2005; ADNAN et al., 2019a; GORODETSKAYA et al., 2018), o que torna seu estudo um desafio.

Atualmente existem vários sistemas ou tecnologias de monitoramento inteligente como sensores, telemetria e internet das coisas <sup>4</sup>, que gerenciam a água através do fornecimento de medições em tempo real ou mesmo simulações de cenários da vazão da água (FALSARELLA et al., 2022; PALERMO et al., 2022). São porém necessários enormes investimentos, que podem ser de difícil ou impossível acesso para as comunidades locais ou

---

<sup>4</sup> Sensor é um dispositivo que tem a função de detectar e responder com eficiência algum estímulo. Telemetria é uma tecnologia que permite a transmissão remota de dados, ou seja, por meio de recursos tecnológicos é possível medir, monitorar ou controlar determinado produto ou serviço. E, internet das coisas (IoT) diz respeito a dispositivos do cotidiano existentes em todos os lugares conectados, coletando e enviando informações entre si com o objetivo de tornar mais prática e produtiva a vida.

mesmo do governo local. Neste contexto, há necessidade de aplicar métodos ou modelos inovadores de previsão, não só acessíveis, mas também, altamente precisos e confiáveis.

Há diversas ferramentas que se propõem a criar modelos capazes de prever de forma confiável o comportamento dos níveis de afluência em rios. Em Hidrologia, duas formas de descrever as variáveis envolvidas no ciclo hidrológico e obter previsões são aplicadas: (i) usar equações com base conceitual ou empírica para descrever os processos físicos, em modelos chamados de *modelos determinísticos*; (ii) considerar que os processos são controlados por leis probabilísticas e usar, portanto, *modelos estocásticos* (ALMEIDA; SERRA, 2017).

Os modelos determinísticos têm sido muito utilizados na prática de engenharia para resolução de diversos problemas, com resultados que atendem a grande parte dos problemas existentes. Apesar dos modelos determinísticos possibilitarem maior capacidade de extrapolação e de representação de processos (BRAVO, 2010), requerem uma ampla gama de dados, sobre intervenções humanas, uso do solo, características fisiográficas da drenagem da bacia, quantidade de chuva, intensidade e sua distribuição (OCHOA-TOCACHI et al., 2016; TEUTSCHBEIN et al., 2018; ADNAN et al., 2019a), suas aplicações são mais complexas e são computacionalmente mais onerosos (SILVA; EWEN, 2000).

Devido a obstáculos como a reduzida disponibilidade de séries históricas hidrológicas e a dificuldade de obtenção de uma vasta quantidade de parâmetros de solo e vegetação requeridas por modelos com base física tais como o SWAT (VIANA et al., 2018), SMAP (MIRANDA et al., 2019), TOPMODEL (XUE et al., 2018), IPH II (ULIANA et al., 2018), dentre outros (DUARTE et al., 2019), ganharam destaque os modelos de séries temporais, estudados com a finalidade de simplificar e melhorar o ajuste e previsão de vazões (TUCCI; HIDROLÓGICOS, 1998; MUSHTAQ et al., 2022).

Os modelos de séries temporais constituem a classe aplicada mais frequentemente e que mais evolui atualmente na modelagem e previsão de fluxo de rios. Esta classe de modelos merece um destaque especial não só pelas suas habilidades de modelagem e previsão, mas também, à sua adaptação flexível, tanto os modelos paramétricos quanto os não-paramétricos, muitas vezes obtendo melhores resultados.

Vários modelos de séries temporais foram desenvolvidos na literatura, como mostram os exemplos destacados no Capítulo 2. Contudo, pesquisadores ainda estão enfrentando dificuldades em definir métodos ou algoritmos que devem ser aplicados para um problema em particular. Não existe um único modelo geral perfeitamente adaptável para um problema em particular, pois ainda está-se em um estágio de exploração métodos capazes de modelar os processos hidrológicos (HUSSAIN; KHAN, 2020a; PAIXÃO et al., 2022).

### 1.1.3 HORIZONTE DE PREVISÃO

Um dos objetivos da modelagem matemática do comportamento dos fenômenos da natureza, no caso do comportamento das vazões nos rios, é o de fazer previsões destas.

Os modelos hidrológicos representam o comportamento dos fenômenos naturais aplicando como dados fundamentais para seu funcionamento as vazões naturais médias diárias, semanais ou mensais (de acordo com a aplicação de interesse), e do horizonte em análise, usando um histórico de série de vazões naturais ou apenas alguns valores passados para períodos à frente.

A obtenção de previsões de vazões permite que os diferentes modelos de operação, otimização e simulação energética simulem as condições operacionais das usinas hidrelétricas no futuro, o que traz subsídios para a tomada de decisões sobre a melhor forma de operar o Sistema, e pode auxiliar na tomada de decisão no que diz respeito a minimização dos efeitos de uma cheia excepcional, desde que feita com relativa precisão e antecedência.

As previsões de aflúências servem também como subsídio a outros setores usuários da água, nomeadamente: a irrigação, a pesca, o turismo, o lazer e a navegação, que também utilizam estas previsões de aflúências para programação das suas atividades.

De acordo com a finalidade das previsões de vazões, diferentes horizontes e intervalos de discretização são utilizados. No setor energético existem três horizontes de previsões de vazões, nomeadamente: previsões de longo prazo, previsões de médio prazo e previsões de curto prazo (GUILHON, 2002; GUILHON et al., 2007a; GÓMEZ; HUERTAS, 2016).

Previsão de longo prazo, com estimativas de fluxos naturais médios mensais de até um ano à frente, são usadas para apoiar a tomada de decisões em uma programação anual de enchimento de novos reservatórios, para sinalizar o comportamento de armazenamento de reservatórios, para indicar o risco de déficit de energia e a entrada e saída de um racionamento de energia elétrica. A Previsão de médio prazo (médias semanais em um horizonte de seis semanas) subsidia a programação mensal, auxilia na otimização dos recursos naturais nos reservatórios, no uso múltiplo de reservatórios como navegação, no controle de eventos extremos, abastecimento, turismo e lazer. Previsão de curto prazo (de algumas horas a 12 dias, com discretização em valores médios diários) são usadas para fins de planejamento diário (GUILHON et al., 2007a).

No presente trabalho, a previsão é feita considerando o horizonte temporal de curto prazo, considerando horizontes de previsão de 1, 3, 5 e 7 dias à frente. A estratégia de previsão aplicada é multipassos à frente. A previsão em várias etapas é de grande importância pois permite o planejamento antecipado e correto de um sistema.

#### 1.1.4 MODELAGEM HIDROLÓGICA NO RIO ZAMBEZE EM MOÇAMBIQUE

A barragem de Cahora-Bassa, área de estudo, é de vital importância para Moçambique, pois é a principal fonte de energia elétrica do país e países vizinhos, contribui para o desenvolvimento da economia do delta do rio Zambeze à jusante, através da viabilização de actividades económicas como agricultura, pastorícia, pesca, construção de vias de acesso e, da redução do risco de ocorrência de calamidades naturais como a seca, cheias ou inundações. (MUNGÓI, 2011)

Contudo, devido a maximização da produção de energia através da libertação de água armazenada para gerar energia durante a estação seca, usando o fluxo elevado indutor de cheias no verão para encher o reservatório, e preparando-se ao mesmo tempo para os fluxos baixos da época seca, o fluxo regulado do Zambeze tem vindo a secar as áreas húmidas antes alimentadas pelas águas das cheias, deixando existir múltiplos canais secundários e ramificações que mudavam constantemente, passando a ser de canal principal único. A água liberada pelas barragens provoca a erosão das margens e aprofunda o leito do rio, devido à necessidade desta de equilibrar o seu conteúdo de sedimentos. As planícies de inundação agora secas apresentam graves consequências para a biodiversidade. As cheias tornaram-se imprevisíveis, só as maiores não são retidas por Cahora-Bassa tornando as comunidades ao longo do Zambeze muito mais vulneráveis aos impactos negativos das mesmas (THA; SEAGER, 2013).

O monitoramento e gestão da bacia hidrográfica do rio Zambeze no território Moçambicano está subdividido em dois principais trechos, as estações de monitoramento à jusante à barragem de Cahora-Bassa e as estações à montante da mesma, sob a responsabilidade das empresas ARA-Zambeze (Administração Regional de Águas do Zambeze) e Hidroelétrica de Cahora-Bassa (HCB), respectivamente.

A ARA-Zambeze trabalha em coordenação com HCB, e esta por sua vez, em coordenação com outras autoridades de gestão nos outros trechos à montante do rio nos países vizinhos, ZINWA (Zimbabwe National Water Authority), DWAF (Department of Water Affairs of Zambia), ZRA (Zambezi River Authority), ZESCO (Zambia Electricity Supply Company), KNBPS (Kariba North Bank Power Station) e KSBPS (Kariba South Bank Power Station).

O modelo de previsão hidrológica utilizado para prever inundações na Bacia do Zambeze em Moçambique, que também é utilizado regionalmente é o SARCOF (Southern African Regional Climate Outlook Forum), que é combinado com outros dois modelos: NWS-CPC (National Weather Service – Climate Prediction Center – USA) e o ECMWF (European Center for Medium Range Weather Forecast). Porém, esses modelos são apontados como inadequados, por serem modelos globais, não incluem especificidades locais, fornecendo assim estimativas muito incertas, e requerem o conhecimento da região para a utilização das mesmas (JESSEN; SILVA, 2008).

A ARA-zambeze, por sua vez, usa no seu trecho o SAC (Sistema de Aviso de Cheias), um modelo que permite avaliar as características da onda de cheia e o grau de inundação que a mesma vai gerar em diferentes trechos do Rio, convertendo o volume de precipitação ocorrido na bacia do Baixo Zambeze em termos de escoamento. A HCB dispõe também de um modelo que permite avaliar as alturas hidrométricas causadas pelas descargas, e o tempo requerido para chegar a vários pontos à jusante. Porém, estes modelos dependem dos dados de entrada e condições iniciais, de dados topográficos, de vegetação e outros elementos fisiográficos da Bacia e do leito do Rio, dados já desatualizados devido às mudanças no rio. A Figura 4 mostra uma régua linimétrica <sup>5</sup> fixada nas margens do rio Zambéze, na região de Boroma-Tete evidenciando a desatualização dos dados devido as mudanças do mesmo.

---

<sup>5</sup> Régua Linimétrica é um instrumento construído em madeira, ferro esmaltado, alumínio ou plástico. As escalas linimétricas, fixadas nas margens dos rios, devem cobrir toda a gama de variações de níveis d'água.

Figura 4 - Régua de Medição de escalas hidrométricas, Boroma-Tete.



Fonte: Extraído de Jessen e Silva (2008)

## 1.2 MOTIVAÇÃO

A água é uma necessidade muito importante de todos os seres vivos. Devido à sua escassez e à preocupação com a sua existência no futuro, há necessidade de aprimorar os sistemas de gestão e de apoio à decisão para que ela não venha faltar (FALSARELLA et al., 2022).

Existem duas formas de conhecer a disponibilidade hídrica, o monitoramento hidrológico e a modelagem hidrológica. O monitoramento hidrológico refere-se à coleta de dados através de medidores; a modelagem hidrológica, à estimativa dos dados (COLLISCHONN; DORNELLES, 2013), sendo a primeira aplicada em Cahora-Bassa.

Os sensores inteligentes, atuadores <sup>6</sup>, dispositivos de medidores inteligentes, telemetria, internet das coisas (IoT) são exemplos de instrumentos ou tecnologias digitais emergentes que são utilizadas na atualidade, tornando os sistemas de gestão da água inteligentes mais sustentáveis e eficientes (PALERMO et al., 2022).

Essas tecnologias são úteis pois se comunicam umas com as outras para coletar, armazenar e disponibilizar rapidamente informações que antes eram difíceis de serem coletadas, gerando dados ou grande volumes de dados (*big data*), sem a necessidade da mediação humana. Estes dados são posteriormente processados, e seus resultados fornecidos para subsidiar os gestores no processo decisório (FALSARELLA et al., 2022). A captura do índice pluviométrico, temperatura, evaporação, fluxo, umidade, nível de água, entre outras ao longo dos vários cursos de água existentes em uma bacia hidrográfica são exemplos de informações relevantes para auxiliar o processo decisório na gestão de recursos hídricos.

A necessidade de processamento desses dados ou grandes quantidades de dados gerados tem direcionado a atenção do setor hídrico para soluções digitais avançadas, onde ferramentas orientadas a dados como *machine learning*, *deep learning* ou outras de inteligência computacional são as alternativas utilizadas, gerando valor, e por conseguinte, reduzindo a subjetividade ou o uso da intuição no processo de tomada de decisão (PALERMO et al., 2022).

Existem alguns estudos na literatura sobre a previsão no rio Zambeze que reforçam a importância do monitoramento da bacia hidrográfica. Kling et al. (2014) estudaram o futuro dos recursos hídricos na bacia do Zambeze, com base em previsões do Banco Mundial que incorporam os efeitos da irrigação em grande escala, novos projetos hidrelétricos e cenários hipotéticos de mudança climática usando projeções. Hughes et al. (2020) calibraram duas abordagens hidrológicas (WEAP e Pitman) para comparar os efeitos do desenvolvimento e da variabilidade climática nos recursos hídricos da bacia do Rio Zambeze. Hulsman et al. (2021) fizeram análise de seca baseada em satélite na Bacia do Rio Zambeze na África Subsaariana durante a estação seca de 2019, quando níveis excessivamente baixos de água foram registrados no Zambeze.

Porém, não há registros de estudos sobre a bacia hidrográfica do Zambeze que apliquem modelos de aprendizado de máquina ou de inteligência computacional para essa finalidade. Havendo portanto, a necessidade do desenvolvimento de uma estrutura de modelagem e previsão de cenários futuros dos recursos hídricos no rio Zambeze com maior antecedência, para subsidiar o processo de tomada de decisões no seu gerenciamento.

Este estudo também pode:

- a) fornecer contribuições científicas para a área de modelagem computacional aplicada ao monitoramento de recursos hídricos, na engenharia de recursos

---

<sup>6</sup> Atuador é um dispositivo que produz movimento, convertendo energia pneumática, hidráulica ou elétrica, em energia mecânica.

- hídricos, auxiliando no preenchimento desta lacuna na literatura;
- b) fornecer estimativas que podem ser utilizadas por instituições operadoras de estações fluviométricas para análise de consistência e preenchimento de falhas de dados de vazão;
  - c) fornecer aos tomadores de decisões um conhecimento a priori da vazão do rio no reservatório da hidroelétrica e, permitindo assim planejar de forma antecipada a geração e distribuição de energia elétrica;
  - d) otimizar outras demandas de uso da água, como sistemas de irrigação agrícola, navegação fluvial, construção de vias de acesso, entre outras;
  - e) auxiliar na utilização e distribuição sustentável de água, considerando a diversidade hidrológica da região;
  - f) contribuir na economia, através do fornecimento de energia para os países vizinhos.
  - g) contribuir na prevenção de desastres ambientais, com redução das possibilidades de colapso de obras hidráulicas, controlo de inundações em cidades, zoneamento de planícies de inundação;
  - h) contribuir para a conservação da biodiversidade da bacia hidrográfica;
  - i) Auxiliar nos estudos sobre qualidade, demanda e abastecimento de água nas cidades;
  - j) auxiliar na tomada de decisões sobre disponibilidade hídrica para novas outorgas de água;
  - k) auxiliar na promoção da soberania territorial, integridade territorial, Utilização Equitativa e Gerenciamento Comum das nações através de instrumentos como acordos, tratados, leis ou políticas internas, regionais, continentais e mundiais de partilha de águas.

Como se pode observar, nas contribuições referenciadas anteriormente, estão subjacentes os objetivos de desenvolvimento sustentável (ODS), portanto, os resultados desta pesquisa podem constituir um contributo para o cumprimento destes objetivos, os quais Moçambique é signatário, Figura 5.

Figura 5 - Objetivos de Desenvolvimento sustentável(ODS).



Fonte: Extraído de Ferreira (2018)

### 1.3 OBJETIVOS

Nesta seção serão descritos o Objetivo Geral e os Objetivos Específicos que foram desenvolvidos no trabalho.

#### 1.3.1 OBJETIVO GERAL

Esta tese tem como objetivo geral desenvolver uma estrutura de modelagem e previsão da vazão afluente à barragem hidroelétrica de Cahora-Bassa em Moçambique através de modelos de aprendizado de máquina, utilizando dados hidrometeorológicos (valores antecedentes da vazão, precipitação, evaporação e umidade relativa) como variáveis de entrada.

#### 1.3.2 OBJETIVOS ESPECÍFICOS

Para a operacionalização do objetivo geral, propõem-se os seguintes objetivos específicos:

- a) descrever os modelos de aprendizagem de máquinas na previsão hidrológica da vazão;
- b) avaliar a performance dos modelos de aprendizado de máquinas na previsão diária da vazão afluente à barragem de Cahora-Bassa;
- c) analisar a qualidade de previsão em diferentes horizontes temporais;

- d) avaliar a capacidade de ajuste de cada modelo às vazões extremas (vazões baixas e altas);
- e) estudar a incerteza de previsão dos métodos na modelagem e previsão das vazões;
- f) analisar os fatores que influenciam na previsão diária da vazão na barragem de Cahora-Bassa.

#### 1.4 ORGANIZAÇÃO DO TEXTO

O trabalho está estruturado da seguinte maneira:

O Capítulo 1 apresenta a *Introdução*, contendo uma descrição da situação do problema, uma discussão de políticas e estratégias de gestão de recursos hídricos, da modelagem de hidrológica na área de estudo, da motivação e dos objetivos do trabalho;

No Capítulo 2 faz-se a *Revisão bibliográfica* do trabalho, isto é, os desenvolvimentos na literatura sobre previsão da vazão desde a aplicação dos modelos clássicos até aos mais recentes. Apresentam-se trabalhos utilizando modelos de aprendizado de máquinas padrão, aprendizado online, aprendizado profundo, modelos híbridos por combinações de modelos individuais, modelos híbridos por otimização de parâmetros ou metamodelos, métodos de pré-processamentos aplicados aos modelos de aprendizado de máquinas, e modelo de aprendizados aplicados a previsão de eventos extremos;

O Capítulo 3 é o dos *Materiais e Métodos*, este descreve a área de estudo, o conjunto de dados, os principais conceitos aplicados neste trabalho, as séries temporais em hidrologia, os modelos de aprendizado aplicados e os algoritmos bioinspirados utilizados na seleção dos parâmetros,

O Capítulo 4 é o da *Metodologia*, que explica de forma minuciosa toda ação desenvolvida para obtenção dos resultados desta pesquisa. Descrevem-se a estrutura de modelagem, as técnicas de seleção de variáveis de entrada usadas, a estratégia de modelagem, as formas de avaliação e generalização dos modelos, as técnicas de análise de incerteza usadas;

O Capítulo 5 é o dos *Experimentos computacionais*, que exhibe os principais resultados como a análise descritiva, a análise das simulações, a análise comparativa dos resultados e a validação do estudo;

E, o trabalho é finalizado com as *conclusões, trabalhos publicados, trabalhos futuros, e a lista de Referências Bibliográficas* utilizadas.

## 2 REVISÃO BIBLIOGRÁFICA

Os modelos de séries temporais tornaram-se ferramentas importantes na análise e interpretação de dados na ciência e tecnologia, gerando conhecimento.

Devido a rapidez na sua elaboração e capacidade de se ajustar melhor aos dados prevendo adequadamente dados semelhantes (LIU et al., 2015), essas técnicas tornaram-se alternativas de modelagem sendo encontradas em diversos problemas de previsão de variáveis de interesse na hidrologia como: precipitação, vazão, transporte de poluentes em cursos d'água, evapotranspiração, transporte de sedimentos, qualidade da água e estimação de coeficiente de dispersão longitudinal de córregos naturais, entre outras, (PINI et al., 2020; ADNAN et al., 2019b; ADNAN et al., 2020; LI et al., 2019; GOLIATT et al., 2021).

Os modelos de séries temporais podem ser subdivididos em dois grandes grupos: métodos estatísticos e modelos de aprendizado de máquina.

Métodos estatísticos, também designados por clássicos, como regressão linear múltipla, *Autoregressive integrated Moving Average* (ARIMA) e suas variantes, foram amplamente aplicados para previsão hidrológica desde a década de 1970. Porém, estes métodos somente produzem resultados satisfatórios na previsão de séries temporais quando as mesmas são lineares ou quase lineares; não capturam os padrões não-lineares e não-estacionários ocultos em séries temporais (ZHANG et al., 2016).

As séries temporais hidrológicas são caracterizadas por complexidades dinâmicas e não-estacionariedade (UNNIKRISHNAN; JOTHIPRAKASH, 2018). Por isso, nos últimos anos, várias técnicas de aprendizado de máquina como *Support Vector Regression (SVR)*, *Artificial Neural Networks (ANN)*, *Multi-Adaptive Regression Splines (MARS)*, *Adaptive Neuro-Fuzzy Inference System (ANFIS)*, *long short-term memory network (LSTM)*, entre outras receberam uma atenção considerável pelas suas fortes capacidades de aprendizagem e adequação para modelagem complexa e para processos não-lineares. Como resultado, uma variedade de modelos de aprendizado de máquina foram desenvolvidos e utilizados para produzir previsões satisfatórias de processos hidrológicos não-lineares (NOURANI et al., 2014; CHEN et al., 2018).

### 2.1 MÉTODOS DE APRENDIZADO DE MÁQUINAS PARA PREVISÃO DA VAZÃO

Técnicas de aprendizado de máquina para prever vazões mensais de rios foram relatadas por Hussain e Khan (2020a). Este estudo analisou o potencial de métodos de aprendizado de máquina, nomeadamente, *Multilayer perceptron (MLP)*, *Support Vector Regression (SVR)* e *Random Forest (RF)*. Os métodos foram utilizados para prever o fluxo um passo à frente, utilizando como dados de entrada valores passados da vazão do rio (tempo defasado). Os resultados mostraram que o RF teve o melhor desempenho, seguido por MLP e SVR, e reforçaram o argumento de que algoritmos /modelos de aprendizado

de máquina, particularmente o modelo RF podem ser usados para prever o fluxo de rios com alta precisão e melhorar a gestão de riscos e água.

Tongal e Booij (2018) compararam o SVR, ANN e RF para prever o fluxo simulado como uma função da precipitação, da temperatura, de evapotranspiração e dos valores antecedentes de vazão em quatro rios nos Estados Unidos. O modelo ANN teve melhor desempenho em três bacias, enquanto o modelo SVR foi melhor no período de calibração. No período de validação, SVR e RF tiveram melhor desempenho do que o modelo ANN devido à fraca capacidade de generalização do modelo ANN.

Herrera et al. (2010) descreveram e compararam uma série de modelos preditivos, nomeadamente: ANN, PPR, MARS, RF, e SVR para previsão da demanda de água uma hora à frente, usando como dados de entrada os valores antecedentes da série horária de demanda hídrica, a temperatura, a velocidade do vento, a pressão atmosférica e a chuva. Os resultados desta comparação indicaram o SVR como o modelo mais preciso, seguido por MARS, PPR e RF. A precisão dos resultados obtidos sugere que o SVR é uma alternativa para decisões de manejo em bacias de médio porte.

Wang et al. (2009) compararam o desempenho de vários métodos como ARMA, ANN, ANFIS, GP, e SVR para a previsão de séries temporais de descarga mensal. Os métodos foram testados utilizando registros anteriores dos fluxos mensais dos rios em análise como variáveis de entrada dos modelos. Os resultados indicaram que o ANFIS produziu o melhor desempenho considerando diferentes critérios de avaliação durante as fases de treinamento e validação.

Sun et al. (2014) relataram que o GPR superou os métodos baseados em ARIMA em mais de 400 bacias hidrográficas e sugeriu o GPR como um modelo de aprendizado de máquina eficiente para previsão precisa do fluxo diário. Os valores antecedentes de vazão, precipitação, temperatura mínima e máxima e evaporação potencial foram utilizados como dados de entrada para previsão um passo à frente.

Parisouj et al. (2020) examinaram a capacidade de SVR, ANN-BP, e ELM para os fluxos mensais e diários de quatro rios nos Estados Unidos usando dados diários antecedentes da vazão, precipitação diária, temperatura mínima e máxima diárias como as principais variáveis preditoras e, concluíram que SVR teve o melhor desempenho entre os três modelos, nas escalas mensal e diária.

Modelos de aprendizado de máquina para regionalização de fluxo foram relatados por Ferreira et al. (2021). Esses autores aplicaram RF, MARS e um modelo linear. Os dados de entrada envolvem variáveis ambientais relacionadas com uso e cobertura do solo, morfometria, clima, topografia e condições de superfície como: altitude média, máxima e mínima; inclinação média e máxima; déficit hídrico climático médio, máximo e mínimo; evapotranspiração real e potencial média, máxima e mínima; temperatura máxima e mínima; precipitação total, mínima e máxima anual; umidade média, máxima

e mínima do solo; fluxo equivalente ao volume precipitado, entre outras. Os resultados mostraram que os modelos RF e MARS tiveram desempenhos semelhantes, e notável capacidade de prever o fluxo mínimo e o fluxo médio de longo prazo. Segundo os autores, os métodos de aprendizado de máquina constituem alternativas potentes e promissoras para a regionalização das vazões, auxiliando na gestão e no planejamento integrado dos recursos hídricos ao nível das bacias hidrográficas.

Pini et al. (2020) avaliaram técnicas de Aprendizado de Máquina como ANN, SVR, RF para previsão de fluxo afluente um passo à frente, aplicando como entradas as vazões afluentes e efluentes e seus atrasos, o nível da água e seus valores antecedentes, a média do nível de água em sete dias, o desvio padrão da vazão afluente em sete dias, as temperaturas máxima e mínima, e a precipitação e seus valores antecedentes. Os resultados experimentais mostraram que a ANN tem melhor desempenho para estimativa de fluxo, seguida por SVR e RF.

Abudu et al. (2010) compararam o desempenho de modelos estatísticos na previsão do fluxo mensal utilizando seus valores antecedentes como entrada. Os resultados sugerem que os modelos de série temporal simples (ARIMA e SARIMA) podem ser usados na previsão de fluxo de um mês à frente no local de estudo como uma estrutura de modelo simples e explícita e, um modelo de desempenho semelhante aos modelos ANN de Jordan-Elman.

Niu e Feng (2021) examinaram o potencial de cinco abordagens de inteligência artificial para a previsão de séries temporais de fluxo diário, ANFIS, ANN, GPR, ELM e SVR, utilizando dados antecedentes de vazão e precipitação como entradas. Os resultados aplicando esses modelos em dados obtidos a partir de dois enormes reservatórios hidrelétricos revelaram que as cinco tecnologias de inteligência artificial poderiam produzir boas estimativas de previsão. Em contraste, as técnicas GPR, SVR e ELM podem superar ANN e ANFIS nas fases de teste e treinamento.

Adnan et al. (2020) compararam a precisão dos métodos heurísticos, *optimally pruned extreme learning machine (OP-ELM)*, *least squares support vector machine (LSSVM)*, *multivariate adaptive regression splines (MARS)*, e *M5 model tree (M5Tree)* para modelar mensalmente o fluxo de água, usando dados hidro-meteorológicos (vazão, precipitação e temperatura) como entradas, e concluíram que os modelos LSSVM e MARS forneceram resultados de predição mais precisos.

Sapitang et al. (2020) aplicaram quatro abordagens de aprendizado de máquina supervisionadas, nomeadamente: *Decision Forest Regression (DFR)*, *Boosted Decision Tree Regression (BDTR)*, *Neural Network Regression (NNR)* e *Bayesian Linear Regression (BLR)* para a previsão do nível de água, usando como variáveis de entrada a precipitação e nível de água. A pesquisa revelou que, dos quatro algoritmos propostos, o modelo BLR obteve melhor desempenho que os demais.

Rasouli et al. (2012) compararam três métodos de aprendizado de máquina BNN, SVR e GP com a MLR, utilizando como entradas do modelo: temperatura máxima, precipitação, vazão e seus valores antecedentes, fase sazonal, oscilação ártica, padrão norte americano do pacífico, oscilação do Atlântico Norte, anomalia do Niño, precipitação acumulada, aquecimento, pressão do ar, pressão ao nível do mar, água precipitável, temperatura da superfície do mar, umidade relativa, temperatura e temperatura do ar, fase e amplitude do vento . Os resultados mostraram que os modelos não lineares geralmente superaram o MLR, e que o BNN superou ligeiramente os outros modelos não lineares na previsão diária de vazão.

A viabilidade do modelo *Radial M5 model tree (RM5Tree)* frente a outros modelos como: *M5 model tree (M5Tree)*, *artificial neural networks (ANN)*, *radial basis function neural networks (RBFNN)*, e *multivariate adaptive regression spline (MARS)* foi investigada na previsão e estimativa da vazão diária em um clima frio por Kisi et al. (2022), utilizando os valores antecedentes da vazão como variáveis de entrada. Verificou-se que o RM5Tree oferece melhor precisão na previsão da vazão diária, em comparação com outros modelos, melhorando a precisão do M5Tree.

Vatanchi et al. (2022) aplicaram *Artificial Neural Network (ANN)*, *Adaptive Neuro-Fuzzy Inference System (ANFIS)*, *Bidirectional LSTM (BiLSTM)*, e *Convolutional Neural Network Gated Recurrent Unit Long-Short Term Memory (CNN-GRU-LSTM)* para prever o fluxo diário de longo prazo usando seus antecedentes como dados de entrada. E, foi determinado que o modelo ANFIS superou os outros modelos estudados em termos de confiabilidade e precisão. Embora os modelos CNN-GRU-LSTM e BiLSTM sejam complexos, eles não apresentam melhor desempenho do que os dois modelos padrões ANN e ANFIS.

Lima et al. (2018) implementaram, aplicaram e analisaram o algoritmo SVR (*Support Vector Regression*) com a finalidade de obter um modelo computacionalmente mais preciso de previsibilidade hídrica em relação a outros dois modelos de prognóstico futuro (rede neural e árvore de regressão) através da análise de imagens do conjunto de dados históricos NDVI (*Normalized Difference Vegetation Index*). Comparando a precisão dos três métodos, seus resultados mostraram que o algoritmo SVR conseguiu quantificar de forma mais precisa, estatisticamente, em comparação com outros dois modelos, e ser mais robusto em desempenho computacional para prever o fluxo do rio.

Aghelpour et al. (2022) investigaram a teleconexão de sinais climáticos de grande escala com a vazão do rio. Os Índices climáticos utilizados como entradas dos modelos analisados (*adaptive neuro-fuzzy inference system (ANFIS)* e *multilayer perceptron (MLP)*) incluem o Índice de Oscilação do Sul, temperatura da superfície do mar do Pacífico tropical oriental extremo (Niño), Índice de Temperatura Média Global, Padrão do Pacífico Norte (NP), Oscilação Decadal do Pacífico e Oscilação do Atlântico Norte. Os resultados

indicaram que MLP superou ANFIS na previsão da vazão e, a inclusão de índices climáticos melhoram a previsão.

Osman et al. (2021) testaram e avaliaram os modelos *Extreme Gradient Boosting (Xgboost)*, *Artificial Neural Network(ANN)*, e *Support Vector Regression (SVR)* , para previsão de níveis de água subterrâneas usando dados registrados de precipitação, temperatura e evaporação como variáveis preditoras. O modelo Xgboost superou os ANN e SVR na previsão um passo à frente.

Adnan et al. (2021) investigaram as habilidades dos modelos *Group method of data handling-neural networks (GMDH-NN)*, *dynamic evolving neural fuzzy inference system (DENFIS)* e, *multivariate adaptive regression spline (MARS)* para previsão mensal da vazão. A Precipitação, temperatura e vazões antecedentes foram utilizadas como variáveis de entrada. Os resultados dos testes mostraram que o modelo DENFIS fornece resultados de previsão mais precisos em comparação com os outros.

Souza et al. (2022) compararam as capacidades dos modelos *Group method of data handling (GMDH)*, *seasonal autoregressive integrated moving average with external or exogenous regressors(SARIMAX)*, *Multilayer Perceptron (MLP)* e *Extreme Learning Machine (ELM)* na previsão um passo à frente da vazão utilizando como dados de entrada, a vazão, precipitação, evaporação e umidade relativa. E, verificaram que o modelo GMDH superou os outros modelos.

Essam et al. (2022) testaram e desenvolveram modelos de aprendizado de máquina (ML) baseados na *support vector machine (SVM)*, *artificial neural network (ANN)*, e *long short-term memory (LSTM)* para prever a vazão utilizando variáveis de entrada seus valores passados. Com base nos resultados obtidos, o modelo ANN foi deduzido como o melhor na previsão da vazão um passo à frente.

Versões online de modelos de aprendizado de máquina para estimativa de vazão diária de rios também foram encontradas na literatura. Lima et al. (2016), Lima et al. (2017) empregaram *Online sequential extreme learning machine (OSELM)* e *Variable complexity online sequential extreme learning machine (VC-OSELM)* respectivamente, utilizando como dados de entrada a precipitação e seus valores antecedentes, precipitação mínima e máxima e antecedentes, vazões e seus antecedentes, Médias móveis da precipitação, Médias móveis da precipitação máxima e mínima, pressão média ao nível do mar, Água precipitável, umidade específica, temperatura, Teor volumétrico de umidade do solo, temperatura do solo, velocidade do vento, o equivalente de água da profundidade da neve acumulada, escoamento da água e, precipitação total . Os modelos propostos são comparados com a regressão linear múltipla sequencial online (OSMLR). Os resultados mostraram que o VC-OSELM superou significativamente o OSELM e, este por sua vez, superou o OSMLR para prever fluxos de água. Existem ainda na literatura sobre previsão hidrológica, abordagens viradas na aplicação de modelos de aprendizagem profunda.

Cheng et al. (2020) aplicou uma estrutura de previsão multipassos recursiva com escalas múltiplas para previsão de fluxo de água à longo prazo usando métodos de aprendizado de máquina, aplicando valores antecedentes do fluxo e a precipitação com entradas dos modelos. Os resultados mostraram que os modelos LSTM e ANN produzem previsões diárias confiáveis com até 20 dias de antecedência. Além disso, o modelo LSTM obteve melhores resultados do que o modelo ANN na previsão de longo período de antecedência na escala diária.

Latif et al. (2021) avaliaram a precisão do modelo de aprendizado profundo, *long short-term memory network (LSTM)* em relação aos modelos, *support vector machine (SVM)* e *artificial neural network (ANN)* para previsão de fluxo em reservatório utilizando apenas seus valores antecedentes como entrada. Os resultados indicaram que o modelo LSTM teve um desempenho muito melhor do que SVM e ANN.

Como se pode observar na Tabela 1, diversos modelos de inteligência computacional (IC) foram aplicados para realizar previsões de vazões de rios, indicando maior complexidade para sua caracterização. As séries temporais das aplicações relacionadas à água são comumente caóticas, dinâmicas e de alta estocasticidade (WEGAYEHU; MULUNEH, 2021), o que geralmente interfere na precisão e confiabilidade dos modelos, tornando a previsão um desafio ainda maior.

Tabela 1 – Resumo de métodos de aprendizado de máquinas para previsão de vazões.

Referência	País	Método
Wang et al. (2009)	China	Adaptive Neuro-Fuzzy Inference System (ANFIS)
Abudu et al. (2010)	Turkey	Jordan-Elman Artificial Neural Networks (JE-ANN)
Herrera et al. (2010)	Spain	Multi-Adaptive Regression Splines (MARS)
Rasouli et al. (2012)	Canadá	Bayesian neural network (BNN)
Sun et al. (2014)	United States	Multilayer Perceptron (MLP)
Lima et al. (2016)	Canada	Online Sequential Extreme Learning Machines (OSELM)
Lima et al. (2017)	Canada	Varying Complexity Online Sequential Extreme Learning Machine (VC-OSELM)
Tongal e Booiij (2018)	China	Support Vector Machine (SVM)
Lima et al. (2018)	United States	Support Vector Machine (SVM)
Hussain e Khan (2020b)	Pakistan	Random Forest
Pini et al. (2020)	Italy	Neural Networks (NN)
Sapitang et al. (2020)	Malaysia	Bayesian Linear Regression (BLR)
Parisouj et al. (2020)	United States	Extreme Learning Machine (ELM)
Adnan et al. (2020)	Pakistan	Least Squares Support Vector Machine (LS-SVM)
Cheng et al. (2020)	Thailand	Long Short-Term Memory-based (LSTM)
Niu e Feng (2021)	China	Support Vector Machine (SVM)
Ferreira et al. (2021)	Brazil	Random Forest (RF)
Latif et al. (2021)	Malaysia	Long Short-Term Memory-based (LSTM)
Adnan et al. (2021)	Pakistan	Dynamic evolving neural-fuzzy inference system (DENFIS)
Osman et al. (2021)	Malaysia	Extreme Gradient boosting(XGboost)
Fluet-Chouinard et al. (2022)	Switzerland	Long short-term memory (LSTM)
Aghelpour et al. (2022)	Iran	Multilayer Perceptron (MLP)
Souza et al. (2022)	Mozambique	Group method of data handling (GMDH)
Essam et al. (2022)	Malaysia	artificial neural network (ANN)
Kisi et al. (2022)	Sweden	Radial M5 model tree (RM5Tree)
Vatanchi et al. (2022)	USA	Adaptive Neuro-Fuzzy Inference System (ANFIS)

Fonte: Elaborada pelo autor (2023)

## 2.2 MÉTODOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS PARA PREVISÃO DA VAZÃO

Embora os modelos de aprendizado de máquina obtenham bons resultados individualmente, evidências mostram que pode-se combinar dois ou mais modelos entre si para eliminar suas desvantagens individuais e melhorar suas capacidades de previsão. Os modelos resultantes são chamados *modelos híbridos*, (YASEEN et al., 2020; IBRAHIM et al., 2022; TIKHAMARINE et al., 2019; ADNAN et al., 2021; NGUYEN et al., 2022; DEKA et al., 2014). As estratégias híbridas associam várias estratégias, entre elas, modelos conceituais distribuídos incluindo modelagem física, inteligência artificial, aprendizado de

máquina e modelos estocásticos (GUILHON et al., 2007b). A vantagem desses modelos é de abranger uma classe maior de modelos dentro de sua estrutura, permitindo maior flexibilidade na modelagem de fenômenos dinâmicos (DEKA et al., 2014).

Duas formas de hibridação são frequentemente utilizadas na previsão de variáveis hidrológicas: a combinação de previsões de modelos individuais, e a integração dos algoritmos de otimização nos modelos de aprendizado de máquinas, sendo este último grupo também designado por *metamodelos* (SILVA et al., 2014).

Os modelos obtidos a partir de combinações de previsões de modelos individuais são relatados a seguir.

### 2.2.1 MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS POR COMBINAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINAS INDIVIDUAIS

Séries temporais geralmente consistem em partes lineares e não-lineares (ZHANG, 2003). Com o objetivo de capturar não-linearidades e outros padrões ocultos, Phan e Nguyen (2020) combinaram o modelo estatístico linear ARIMA com os modelos de aprendizado de máquina RF, SVM, KNN, *Long short term memory* (LSTM), resultando em ARIMA-RF, ARIMA-SVR ARIMA-KNN, e ARIMA-LSTM, respectivamente, para previsão do nível de água, utilizando seus valores antecedentes como entradas. Estes modelos híbridos foram comparados entre si e com os seus correspondentes modelos individuais. Os resultados experimentais mostraram a eficácia dos modelos híbridos em relação aos modelos individuais, e os modelos ARIMA-RF e ARIMA-KNN superaram os outros na previsão do nível de água um passo à frente.

Modelos híbridos ARIMA-MLP, ARIMA-LSSVM, ARIMA-NF, e ARIMA-HW, resultantes da combinação de ARIMA com *Multilayer perceptron* (MLP), *Least Square Support Vector Machine* (LSSVM), *neuro fuzzy* (NF) e *Hammerstein-Weiner* (HW), respectivamente foram desenvolvidos por Pham et al. (2019) para modelagem multiestação de chuva utilizando como dados de entrada seus valores passados. Os resultados indicaram que, dos modelos híbridos propostos, ARIMA-NF e ARIMA-HW mostraram-se superiores em termos de habilidades preditivas, e verificou-se ainda que estes aumentaram a precisão da previsão em comparação com os modelos individuais.

Mehdizadeh et al. (2019) apresentaram várias modelagens híbridas de séries temporais com modelos de inteligência artificial, incluindo MARS, MLR e *Gene Expression Programming* (GEP). Tais métodos foram combinados com os modelos *self-exciting threshold auto-regressive* (SETAR) e *fractionally auto-regressive integrated moving average* (FARIMA) resultando em seis modelos híbridos criados GEP-FARIMA, MARS-FARIMA, MLR-FARIMA, GEP-SETAR, MARS-SETAR, e MLR-SETAR. Os resultados mostraram que os modelos híbridos foram mais precisos do que os modelos independentes na modelagem de séries temporais mensais de fluxo de água utilizando como dados de entrada seus

valores passados, e MLR-FARIMA e MLR-SETAR demonstraram maior precisão do que outros modelos.

Um modelo híbrido obtido combinando *Gaussian Mixture Model* (GMM) com XGB (GMM-XGB) foi concebido por Ni et al. (2020b) para previsão um passo à frente, usando a precipitação, temperatura e vazão como variáveis de entrada. Para determinar a precisão da previsão do modelo criado, XGB simples e SVR foram escolhidos como modelos de referência. Os resultados indicaram que o modelo GMM-XGB proposto é uma alternativa superior, que pode fornecer previsões precisas e confiáveis para o gerenciamento ideal dos recursos hídricos.

Brentan et al. (2017) combinaram o modelo SVR com *Adaptive Fourier series*(AFS) para previsão da demanda de água de curto prazo, aplicando como dados de entrada umidade do ar, temperatura, velocidade do vento, chuva e demanda de água. O modelo híbrido SVR-AFS foi comparado com SVR simples. Os resultados mostraram que o modelo híbrido corrigiu a demanda prevista pelo modelo base.

Um estudo apresentado por Saraiva et al. (2021) analisou a precisão na previsão multipassos à frente dos modelos SVR e ANN combinados com um método *bootstrap* e transformada *wavelet* utilizando valores antecedentes da vazão como entradas dos modelos. De acordo com os resultados, o modelo ANN atinge um desempenho mais alto do que o SVR, e a combinação mais eficaz para prever o fluxo de água foi o método *bootstrap*, *wavelet* e *neural network*(BWNN).

Hong et al. (2020) desenvolveram modelos híbridos RF-MLP e GB-MLP, que são resultantes da combinação dos modelos MLP com RF e *Gradient Boosting* (GB), respectivamente na previsão de de vazão afluente de barragens utilizando precipitação, temperatura máxima, temperatura mínima, velocidade do vento, radiação solar, umidade relativa e vazão afluente como variáveis de entrada. Esses modelos foram comparados com outros modelos *decision tree* (DT), *MLP padrão*, *RF simples*, *GB padrão*, *recurrent neural network-long short-term memory* (RNN-LSTM), *convolutional neural network* (LSTM) (CNN-LSTM). Os resultados mostraram que a combinação dos modelos melhorou a precisão da previsão da vazão do que os modelos individuais, e o modelo RF-MLP superou os outros modelos na previsão um passo à frente.

Ni et al. (2020a) aplicaram dois modelos baseados em LSTM (LSTM com transformada wavelet (WLSTM) e LSTM com camadas convolucionais(CLSTM)) para previsão mensal de chuva e fluxo de água, utilizando como dados de entrada a precipitação, umidade e temperatura. Os resultados indicaram que o LSTM era aplicável para a previsão de séries temporais, mas o WLSTM e o CLSTM eram alternativas superiores.

Wegayehu e Muluneh (2021) compararam os modelos MLP, LSTM e *gated recurrent unit* (GRU) com novos modelos híbridos propostos, incluindo CNN-LSTM e *convolutional neural network-gated recurrent unit* (CNN-GRU), para previsão um passo à frente da

vazão em diferentes escalas temporais (diária, semanal e mensal), usando a precipitação, a temperatura máxima, e temperatura mínima como dados de entrada. Os resultados indicaram que o modelo CNN-GRU apresentou resultados promissores, o que significa que, a integração da camada GRU com a camada convolucional pode melhorar substancialmente a simulação de séries temporais de vazões.

A precisão dos modelos de Inteligência Computacional (IC) pode ser melhorada usando modelos híbridos de IC, que consistem em modelos acoplados uns aos outros. Porém, essa combinação de modelos nem sempre fornece resultados superiores aos dos modelos individuais. Vatanchi et al. (2022) observaram que modelos híbridos *Convolutional Neural Network Gated Recurrent Unit Long-Short Term Memory (CNN-GRU-LSTM)*, *Bidirectional Long-Short Term Memory (BiLSTM)* não apresentaram melhores desempenhos do que modelos padrões, embora fossem mais complexos.

A Tabela 2 mostra o resumo de alguns modelos que obtiveram resultados superiores na previsão de vazões dos rios.

Tabela 2 – Resumo de métodos de aprendizado de máquinas híbridos obtidos por combinação de modelos padrões para previsão de vazões.

Referência	País	Método Híbrido
Brentan et al. (2017)	Brazil	support vector Regression- Adaptive Fourier series(SVR-AFS)
Pham et al. (2019)	Vietnam	ARIMA-Neuro fuzzy (ARIMA-NF)
Mehdizadeh et al. (2019)	United States	Multiple Linear Regression- fractionally autoregressive integrated moving average (MLR-FARIMA)
Phan e Nguyen (2020)	Vietnam	ARIMA- Random Forest(ARIMA-RF)
Hong et al. (2020)	Korea	Random Forest-Multilayer Perceptron
Ni et al. (2020a)	china	Wavelet-Long short-term memory network (WLSTM)
Ni et al. (2020b)	China	Gaussian Mixture Model- Extreme Gradient Boosting (GMM-XGBoost)
Saraiva et al. (2021)	Brazil	bootstrap-wavelet-neural network(BWNN)
Wegayehu e Muluneh (2021)	Ethiopia, Italy	convolutional neural network-gated recurrent unit (CNN-GRU)

Fonte: Elaborada pelo autor (2023)

### 2.2.2 MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS POR OTIMIZAÇÃO DE HIPERPARÂMETROS

No entanto, o desempenho adequado dos modelos de aprendizado de máquina também pode depender da escolha dos parâmetros internos (KUHN; JOHNSON, 2013). Por exemplo, as redes neurais exigem que a topologia de camadas, a taxa de aprendizado e o número de neurônios sejam estabelecidos. Os modelos de processos gaussianos dependem da escolha da função *kernel* e dos parâmetros correspondentes. Conforme descrito por (CLAESEN; MOOR, 2015), esse processo pode ser bastante difícil. Assim, usar algoritmos

de busca inteligente para identificar a melhor configuração possível é uma possibilidade. Uma possibilidade é usar um modelo de otimização combinado ao método de aprendizado de máquina escolhido para lidar com essas desvantagens. Isto resulta em uma abordagem também híbrida. A integração de algoritmos de otimização evolutiva/bio-inspirados nos modelos de aprendizado de máquina constitui a forma mais robusta popularmente aplicada, e atualmente em desenvolvimento.

Como resultado, diversos algoritmos de busca ou metaheurísticas evolutiva como *Genetic Algorithms (GA)*, *Particle Swarm Optimization (PSO)*, *Gray Wolf Optimization (GWO)*, *Differential Evolution (DE)*, *Ant Colony Optimization (ACO)*, entre outros, têm sido testados e incorporados para melhorar o desempenho de modelos de aprendizado de máquina em problemas reais como extração de bordas de imagens, segmentação de imagens, imagens médicas, agrupamento de dados, mesclagem de imagens médicas, devido à sua excelente capacidade de busca de ótimo global em comparação com algoritmos clássicos de otimização obtendo bons resultados (MIRJALILI et al., 2014). A exploração desses modelos híbridos para previsão de variáveis hidrológicas é uma metodologia relativamente recente e popular que tem mostrado uma notável melhora na previsão.

Alguns estudos recentes nessa linha para a previsão de vazão de rios têm sido desenvolvidos. O modelo MARS combinado com o DE (MARS-DE) foi desenvolvido por Al-Sudani et al. (2019a) para simulação de fluxo de água em um ambiente semi-árido, utilizando seus valores antecedentes como entradas. De acordo com os autores, o modelo MARS-DE demonstrou boa capacidade de modelagem preditiva híbrida para fluxo de água em uma escala de tempo mensal, em comparação com o LSSVR e o modelo MARS padrão.

A Otimização multiobjetiva (*multi-objective optimization design (MOOD)*) foi usada por Ribeiro et al. (2020) para selecionar e ajustar os pesos dos modelos ELM e *Echo State Network (ESN)*, resultando em modelos híbridos ELM-MOB e ESN-MOB para previsão da vazão afluente usando seus valores passados como variáveis de entradas. Os resultados desses modelos foram comparados com o modelo SARIMA, e observou-se que estes atingiram melhores resultados. O ESN-MOB foi o modelo superior.

Yaseen et al. (2017) avaliaram a precisão de previsão do modelo híbrido ANFIS-FFA baseado na combinação entre ANFIS e o *Firefly optimization algorithm (FFA)* na previsão da vazão, aplicando seus valores antecedentes como entradas. O modelo híbrido proposto foi comparado pelo correspondente modelo clássico (ANFIS) e uma versão do mesmo que utiliza o método de agrupamento *fuzzy c-means (FCM)* na geração de Sistemas de Inferência Fuzzy (FIS). Os resultados revelaram que o FFA é capaz de melhorar a precisão de previsão do modelo híbrido ANFIS.

Adnan et al. (2019b), obtiveram um resultado análogo ao do Yaseen et al. (2017) ao combinar ANFIS e PSO (ANFIS-PSO). A abordagem híbrida proposta mostrou capacidade

de produzir estimativas precisas na modelagem dos fluxos diários a montante e a jusante, em comparação com outras abordagens como MARS e M5tree. A precipitação e a vazão são utilizados com dados de entrada dos modelos.

Em um outro estudo, Yaseen et al. (2020) desenvolveram o modelo híbrido ELM com *Salp Swarm Algorithm* (SSA-ELM). O modelo desenvolvido foi comparado com o ELM clássico e outros modelos de IC na previsão mensal da vazão, utilizando seus valores antecedentes como entradas. A precisão da previsão de fluxo do SSA-ELM superou o ELM clássico e outros modelos de IC.

Um algoritmo recente denominado *Grey Wolf Optimization* (GWO) foi aplicado para melhorar o desempenho de modelos de IC por Tikhmarine et al. (2020). Os resultados mostraram que os modelos de IC com GWO integrado, ANN-GWO, SVR-GWO e MLR-GWO superam os métodos de IA padrão ANN e SVR. Além disso, o SVR-GWO superou, em termos de desempenho, ANN-GWO e MLR-GWO na previsão da vazão mensal. Em outro estudo os mesmos autores, Tikhmarine et al. (2019) aplicaram GWO combinado com Wavelet SVR (GWO-WSVR), e o resultado mostrou que o algoritmo GWO superou outras abordagens de otimização como *particle swarm optimization* (PSO) (PSO-WSVR), *shuffled complex evolution* (SCE)(SCE-WSVR), *multi-verse optimization* (MVO) (MVO-WSVR), que também foram empregadas no ajuste de parâmetros do WSVR, revelando a superioridade do GWO em otimizar os parâmetros do SVR padrão para melhorar sua precisão na previsão da vazão. Nos dois estudos apenas valores passados da vazão foram utilizados com variáveis de entrada.

Malik et al. (2020) avaliaram a capacidade de predição do modelo SVR otimizados por vários algoritmos nomeadamente, ACO(SVR-ACO), MVO (SVR-MVO), Spotted Hyena Optimizer (SVR-SHO), Harris Hawks Optimization (SVR-HHO), PSO (SVR-PSO), Bayesian Optimization (SVR-BO) utilizados para seleção os parâmetros do SVR . Os resultados da comparação demonstraram que SVR-HHO teve desempenho superior aos modelos SVR-ALO, SVR-MVO, SVR-SHO, SVR-PSO e SVR-BO na previsão da vazão diária na bacia de estudo, aplicando valores passados de vazão como variáveis de entrada. O novo algoritmo HHO superou os outros em termos de precisão de previsão.

Wu et al. (2019) avaliaram o desempenho de quatro modelos de ELM otimizados por algoritmos bioinspirados ELM com GA (ELM-GA), ELM com ACO (ELM-ACO), ELM com *cuckoo search algorithm* (CSA) e ELM com *flower pollination algorithm* (ELM-FPA) para previsão de evapotranspiração (ETo). Dados diários de temperatura ambiente máxima e mínima, velocidade do vento, umidade relativa e radiação solar global foram utilizados como variáveis de entrada. Os modelos propostos foram avaliados e contrastados com o modelo ELM padrão. Os resultados indicaram maior capacidade dos algoritmos de otimização bioinspirados para melhorar o desempenho do modelo ELM tradicional na previsão diária de ETo, particularmente os algoritmos FPA e CSA.

Adnan et al. (2021) desenvolveram um novo modelo híbrido para a previsão de vazão mensal, integrando PSO e GWO com ELM como ELM-PSOGWO. O método proposto foi comparado com os métodos ELM padrão, ELM-PSO e ELM-PSOGSA (ELM híbrido com PSO e algoritmo de busca gravitacional binário integrados). Os dados mensais de precipitação e vazão foram usados como entradas para os modelos para examinar sua precisão. Os resultados mostraram que o modelo ELM-PSOGWO teve um desempenho melhor do que os outros, tendo sido capaz de fornecer estimativas mais precisas das vazões de pico com o menor erro relativo médio absoluto em comparação com outros métodos.

Kilinc e Yurtsever (2022) desenvolveram o modelo híbrido de Aprendizado profundo denominado *grey wolf algorithm-based gated recurrent unit (GWO-GRU)*, que é resultante da combinação de GWO com GRU para previsão da vazão diária da vazão, usando seus antecedentes como variáveis de entrada. O modelo proposto foi comparado com o modelo linear. Os resultados indicaram que este supera o modelo linear.

Zaini et al. (2018) avaliaram a performance do modelo híbrido SVR com PSO (PSO-SVR) para prever o fluxo diário do rio a curto prazo, utilizando como entradas do modelo o fluxo passado, precipitação, evaporação, umidade relativa média, velocidade do vento média, temperatura máxima e mínima. Os resultados indicaram que o modelo híbrido superou o SVM padrão na previsão do fluxo 1 à 7 dias a frente. Além disso, constataram que a inclusão de variáveis meteorológicas melhoram a previsão do fluxo.

Meshram et al. (2019) desenvolveram um modelo híbrido baseado na integração de *feed-forward neural network (FNN)* com um algoritmo híbrido *particle swarm optimization and gravitational search algorithms (PSOGSA)* (PSOGSA-FNN) para previsão da vazão mensal, utilizando seus valores antecedentes como preditores. Os resultados indicaram que o modelo proposto melhora a precisão da previsão e é um método viável para prever o fluxo do rio.

Riahi-Madvar et al. (2021) treinaram e testaram a performance de vários algoritmos evolutivos, incluindo, FFA, GA, GWO, PSO e DE com ANFIS para previsão da vazão diária, semanal, mensal e anual, usando antecedentes de vazão com entradas. Os resultados indicaram que, todos os algoritmos híbridos desenvolvidos significativamente superaram o desempenho do modelo ANFIS tradicional para todos os horizontes de previsão. Além disso, ANFIS-GWO foi o modelo híbrido superior. Em um outro estudo desenvolvido por Adnan et al. (2022), um modelo híbrido ANFIS com *gradient-based optimization (GBO)* integrado é proposto na previsão da vazão, utilizando dados de temperatura e valores antecedentes de vazão como preditores. Os resultados mostraram que o modelo proposto é superior que o ANFIS padrão.

Haznedar e Kilinc (2022) desenvolveram um modelo híbrido ANFIS com GA integrado (ANFIS-GA) para previsão da vazão, usando seus valores passados como entrada. Os resultados indicaram que o modelo proposto supera o ANFIS padrão, LSTM e ANN.

O modelo ANFIS otimizado por GWO foi aplicado por Dehghani et al. (2019) na previsão multipassos recursiva da vazão entre 5 minutos e 10 dias a frente, usando seus valores antecedentes como entradas, e observaram que o modelo proposto superou o padrão em todos horizontes de previsão.

A tabela 3 apresenta o resumo de alguns modelos híbridos resultantes da otimização dos parâmetros.

Tabela 3 – Resumo de métodos de aprendizado de máquinas híbridos por otimização de parâmetros para previsão de vazões.

Referência	País	Método Híbrido
Yaseen et al. (2017)	Malaysia	Adaptive Neuro-Fuzzy Inference Systems with Firefly optimization algorithm(ANFIS-FFA)
Zaini et al. (2018)	Malaysia	Support Vector Machine with Particle Swarm Optimization (PSO-SVM)
Meshram et al. (2019)	Tukey	feed-forward neural network (FNN) with a hybrid algorithm of the particle swarm optimization and gravitational search algorithms (PSOGSA) (FFN-PSOGSA)
Dehghani et al. (2019)	Iran	ANFIS with GWO
Al-Sudani et al. (2019a)	Iraq	Multivariate adaptive regression spline integrated with differential evolution (MARS-DE)
Adnan et al. (2019b)	Pakistan	Neuro-Fuzzy Inference Systems with Particle Swarm optimization algorithm(ANFIS-PSO)
Tikhamarine et al. (2019)	Algeria	Wavelet support vector Regression with grey wolf optimization (GWO-WSVR)
Wu et al. (2019)	China	Extreme Learning Machine with flower pollination algorithm (ELM-FPA)
Ribeiro et al. (2020)	Brazil	Echo State Network and multi-objective optimization design (ESN-MOB)
Yaseen et al. (2020)	Iraq	Extreme Learning Machine Model with Salp Swarm Algorithm(SSA-ELM)
Tikhamarine et al. (2020)	Egypt	support vector Regression with grey wolf optimization (GWO-SVR)
Malik et al. (2020)	India	Support vector Regression with Harris Hawks Optimization (SVR-HHO)
Alizadeh et al. (2020)	Iran	Support vector regression optimized by grasshopper optimization algorithm (GOA)
Adnan et al. (2021)	Pakistan	Extreme learning Machine with Particle swarm optimization and Gray wolf optimization (ELM-PSOGWO)
Riahi-Madvar et al. (2021)	Iran	Adaptive Neuro-Fuzzy Inference Systems with grey wolf optimization algorithm (GWO-ANFIS)
Adnan et al. (2022)	Pakistan	ANFIS with gradient-based optimization (GBO) (GBO-ANFIS)
Kilinc e Yurtsever (2022)	Turkey	Grey wolf algorithm (GWO)-based gated recurrent unit (GRU) (GWO-GRU)

Fonte: Elaborada pelo autor (2023)

### 2.3 MODELOS DE APRENDIZADO DE MÁQUINAS HÍBRIDOS COM APLICAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS

O Fluxo da água resulta da interação de processos complexos associado a diversas variáveis naturais do ciclo hidrológico, como não estacionariedade, não linearidade e aleatoriedade. Estudar suas influências no fluxo é central e muito significativo do ponto de vista hidrológico. Portanto, o desempenho de modelos de aprendizado de máquina para sua previsão não só depende apenas da hibridação dos modelos através da combinação de previsões ou ajuste de parâmetros com algoritmos otimização, como também, do pré-processamento de entradas (FAHIMI et al., 2017).

O pré-processamento de dados é uma ferramenta vital, que atualmente vem merecendo muita atenção na análise de dados de variáveis hidrológicas. Como resultado, técnicas de pré-processamento baseadas em seleção de variáveis/características e decomposição têm sido desenvolvidas e aplicadas com objetivo de fornecer aos modelos com dados relevantes para um aprendizado mais eficiente. Entradas relevantes em um modelo preditivo são essenciais para garantir uma estimativa altamente precisa dos dados de fluxo, reduzindo o custo computacional e a complexidade do mesmo.

Métodos de seleção de variáveis de entrada (*Input variables selection (IVS)*) acoplados com modelos de IC foram aplicados na simulação de fluxo.

O *forward stepwise regression* (TONGAL; BOOIJ, 2018) é o método padrão, as variáveis de entrada são subsequentemente adicionadas à única melhor variável de entrada até o número permitido de antecedentes, valores determinados a partir das análises de correlação (funções de autocorrelação (ACF), autocorrelação parcial (PACF) e correlação cruzada (CCF)). Uma Avaliação das mudanças no ajuste indica se o novo parâmetro adicionado melhora o modelo mais do que seria esperado ao acaso.

O *Least Absolute Shrinkage and Selection Operator (LASSO)* resultou como um método mais eficiente na seleção de variáveis que o *Particle Swarm Optimization-Artificial Neural Networks (PSO-ANN)* em Alizadeh et al. (2020). Esses dois modelos de seleção LASSO e PSO-ANN foram acoplados ao modelo SVR otimizado por *grasshopper optimization algorithm (GOA)*, e comparados com o SVR padrão na previsão da vazão. Os resultados indicaram que os modelos SVR com parâmetros ótimos baseados em GOA têm uma notável primazia sobre o modelo SVR padrão, de modo que sua aplicação melhora a precisão da previsão. E, os recursos selecionados pelo método LASSO são mais precisos do que os do algoritmo PSO-ANN.

Quilty et al. (2016) desenvolveram um método não-paramétrico e não-linear de seleção de recursos denominado *bootstrap rank-ordered conditional mutual information (broCMI)*. Os resultados mostraram superioridade deste em termos de capacidade de seleção em comparação com os outros métodos de seleção como *partial mutual information*

*selection (PMIS)*, *conditional mutual information selection (CMIS)* e *partial correlation input selection (PCIS)*. Além disso, este método de seleção foi emparelhado com os modelos ANN e ELM para previsão multipassos da demanda diária urbana de água, usando dados antecedentes da demanda, precipitação e temperatura máxima do ar como entradas, e observou-se que os melhores conjuntos de entrada usados em ANN e ELM foram muito semelhantes, fornecendo evidências de que o método broCMI foi capaz de identificar preditores de processo adequados que podem fornecer previsões de natureza semelhante, mesmo ao usar diferentes técnicas de modelagem (ANN e ELM).

O modelo XGB foi utilizado para selecionar entradas correlacionadas e classificando-as de acordo com sua importância, em seguida, fornecidas ao modelo ELM para o processo de previsão da vazão, resultando no modelo XGBELM (HADI et al., 2019). Variáveis hidrológicas incluindo precipitação, temperatura e evapotranspiração foram usados para construir o modelo e prever a vazão. O modelo XGBELM superou os modelos XGB e ELM independentes. Além disso, este modelo melhora a capacidade de previsão com número mínimo de variáveis.

O método *SHapley Additive exPlanations (SHAP)* foi utilizado com os modelos XGB, RF e Classification and Regression Trees (CART) (MUSHTAQ et al., 2022). Os resultados indicaram que todos os modelos tiveram um bom desempenho na previsão diária multipassos da vazão, usando a precipitação, temperatura, equivalente de Água de Neve, evapotranspiração potencial como entradas. O SHAP fornece informações de orientação no processo de seleção de preditores, permitindo a visualização da contribuição de variáveis individuais na previsão. Além disso, a análise SHAP permite que as razões para fazer cada previsão sejam explicadas e as representações aprendidas sejam analisadas, o que aumenta a transparência do aprendizado de máquina.

Em outro estudo de Lin et al. (2021) utilizaram também o método SHAP com o modelo híbrido DIFF-FFNN-LSTM (*first-order difference (DIFF)*, *feedforward neural network (FFNN)*, and *long short-term memory network (LSTM)*) para previsão horária de fluxo, utilizando o fluxo e a precipitação como entradas do modelo. Os resultados indicaram que o modelo proposto superou outros modelos usados na comparação, e a vazão geralmente contribui mais do que precipitação.

Snieder et al. (2020), compararam diferentes métodos de seleção de variáveis de entrada, nomeadamente *partial correlation (PC)*, *partial mutual information (PMI)*, *improved input omission (IO)* e *improved combined neural pathway strength (CNPS)* acoplados com *artificial neural network (ANN)*, e verificaram que o método CNPS fornece entradas que conduzem ao melhor desempenho do modelo ANN.

Algoritmos evolutivos, como o caso de Algoritmos genéticos foram também utilizados por Afan et al. (2020) como métodos de seleção de variáveis de entrada e acoplados ao modelo *Radial basis function neural network (RBFNN)* para previsão da vazão, verificou-se

que GA pode determinar os parâmetros de entrada efetivos na previsão de séries temporais de vazão.

Zhu et al. (2019) aplicaram o modelo *Random Forest* para seleção de variáveis relevantes para oferecer uma análise sistemática de fatores que podem potencialmente afetar a previsibilidade sazonal da vazão com o modelo *Gaussian Process Regression (GPR)*.

Outras ferramentas de pré-processamento mais recentes baseadas na decomposição nomeadamente *complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)* (REZAIIE-BALF et al., 2019), *Intrinsic Time-scale decomposition (ITD)* (WANG et al., 2021b), *time varying filter-based empirical mode decomposition (TVF-EMD)* (JAMEI et al., 2022), entre outras, foram examinadas para fornecer aos modelos de IC com dados de entrada-saída mais informativos, avaliando e melhorando sua precisão de previsão de fluxo.

A Tabela 4 apresenta o resumo dos modelos computacionais híbridos com aplicação de métodos de seleção de variáveis de entradas.

Tabela 4 – Resumo de métodos de aprendizado de máquinas híbridos com aplicação de métodos de seleção de variáveis de entrada

Referência	País	Método Híbrido
Quilty et al. (2016)	Canadá	Extreme learning Machine with bootstrap rank-ordered conditional mutual information (ELM-broCMI)
Hadi et al. (2019)	Turkey	Extreme learning Machine with Extreme gradient boosting inputs selection method (XGBELM)
Zhu et al. (2019)	China	Gaussian Process Regression (GPR) with random forest input selection method
Afan et al. (2020)	Egypt	Radial basis function neural network (RBFNN) with GA selection inputs method
Alizadeh et al. (2020)	Iran	Support vector regression with LASSO input selection
Snieder et al. (2020)	Canada	Artificial neural network with CNPS input selection method
Lin et al. (2021)	China	First-order difference (DIFF), feedforward neural network (FFNN), and long short-term memory network (LSTM) (DIFF-FFNN-LSTM) with SHAP
Mushtaq et al. (2022)	Pakistan	Extreme gradient boosting with SHapley Additive exPlanations(SHAP)

Fonte: Elaborada pelo autor (2023)

Modelos de aprendizado de máquinas também foram aplicados na modelagem de eventos extremos como desastres naturais. Ding et al. (2019) através do desenvolvimento e integração da função de perda de valores extremos (EVL) na rede de memória a fim de memorizar eventos extremos em registros históricos alcançou uma estrutura para previsão de séries temporais com eventos extremos.

Chen et al. (2018) desenvolveu um modelo ELM em que, um algoritmo emergente

baseado em população dupla denominado *backtracking search algorithm (BSA)* foi aplicado para otimizar os parâmetros do ELM para previsão de enchentes de curto prazo, vários índices de desempenho foram usados para avaliar a eficiência do modelo proposto e, em seguida comparado com a *General regression neural network (GRNN)*, os resultados mostraram que o ELM-BSA sempre pode fornecer melhores resultados do que os modelos GRNN nos períodos de treinamento e teste, tornando ELM-BSA uma técnica alternativa promissora para a previsão de cheias.

O modelo XGB foi aplicado para previsão multipassos de cheias (VENKATESAN; MAHINDRAKAR, 2019), e seus resultados mostraram previsibilidade efetiva em termos de pico de descarga.

Sahoo et al. (2022) aplicaram os modelos MLP e SVR otimizados por GWO (MLP-GWO e SVR-GWO) para a previsão de cheias e, observaram que, SVR-GWO obteve resultados superiores do que MLP-GWO. Os resultados indicaram ainda que o uso de GWO como otimizador leva a um desenvolvimento concebível no desempenho de modelos MLP e SVR para previsão de enchentes.

## 2.4 CONTRIBUIÇÃO DA TESE

Como foi observado nas seções 2.1, **2.2.1**, **2.2.2** e 2.3 diferentes tipos de modelos de Inteligência Computacional nomeadamente, modelos padrões, híbridos e profundos foram utilizados para caracterização de variáveis Hidrológicas e eventos extremos em bacias hidrográficas ou rios de diferentes regiões do mundo.

Esta tese desenvolve uma estrutura de modelagem e previsão multipassos da vazão do rio Zambeze afluente à barragem de Cahora-Bassa, utilizando modelos de aprendizado de máquinas híbridos com algoritmos de otimização evolutivos ou bioinspirados e métodos de seleção de variáveis de entrada integrados. Constitui o primeiro estudo dessa natureza sobre o rio Zambeze; os modelos desenvolvidos enquadram-se nos novos avanços em modelagem e previsão, usando modelos de inteligência computacional em Hidrologia.

### 3 MATERIAIS E MÉTODOS

#### 3.1 ÁREA DE ESTUDO

Com uma bacia hidrográfica de 1.4 milhão de  $km^2$ , o rio Zambeze é o quarto maior curso de água do continente africano, depois de Congo, Nilo e Niger. Ocupa uma área de cerca de 135.000  $km^2$  e abriga cerca de 30 milhões de pessoas. Partindo das montanhas ao noroeste da Zâmbia, onde nasce à 1.5 mil metros acima do nível do mar, o rio corre por Angola, Zâmbia, Namíbia, Botswana, Zimbabwe, Malawi e Tanzânia, para além de Moçambique. Seu longo percurso é pontuado por áreas de alta relevância biológica (Figura 6). Ao ganhar velocidade pelos gradientes negativos, cria as cataratas de Victoria – uma das maiores do mundo - e dois grandes lagos: o Kariba, maior lago artificial do mundo, com 5,4 mil  $km^2$ ; e o Cahora Bassa, com 2,7 mil  $km^2$  – ambos reservatórios de usinas hidroelétricas. Finalmente, estende-se do Cahora Bassa até o oceano Índico, onde deságua. Este último trecho forma um extenso e magnífico delta (THA; SEAGER, 2013).

O clima da bacia do Zambeze é influenciado pela convergência de três correntes de ar distintas: a corrente de sudeste, a corrente da monção de nordeste e a corrente de noroeste, conhecida por massa de ar úmido do Congo. (TUMBARE; AUTHORITY, 2004; JESSEN; SILVA, 2008). A precipitação na bacia do Zambeze é extensivamente governada pela ITCZ (*Intertropical Convergence Zone*), que oscila na zona intertropical do globo geograficamente limitada pelos trópicos de Câncer e Capricórnio (JESSEN; SILVA, 2008).

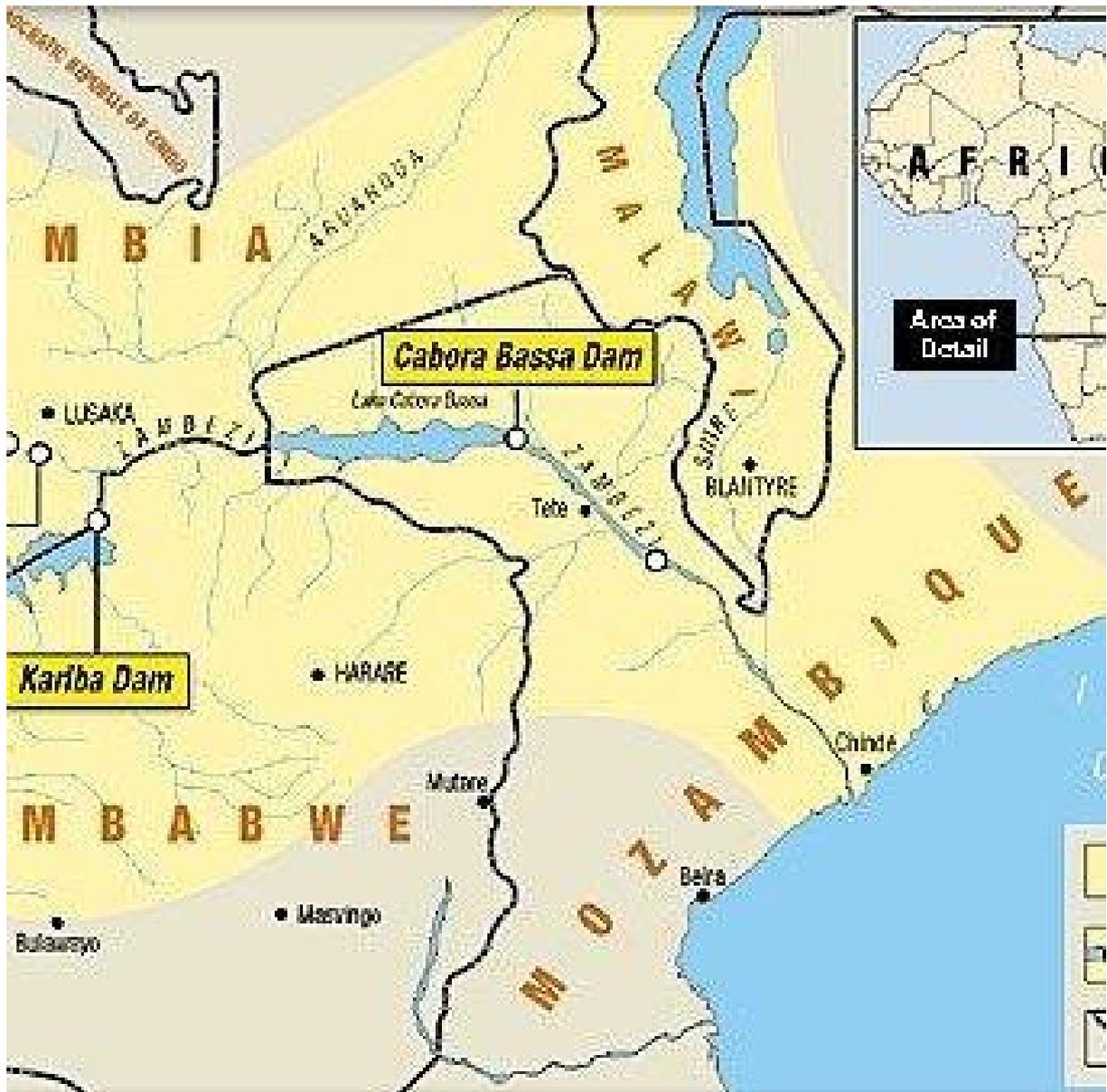
A precipitação anual é mais elevada na parte norte da bacia, sobretudo no território do Congo, Zâmbia, Malawi e norte da Província de Tete, atingindo valores normais máximos de 1400mm, enquanto que na parte sul da bacia, sobretudo no território do Zimbabwe e sul da Província de Tete, atinge valores normais máximos da ordem de 500mm. A época de maior precipitação em toda a bacia compreende habitualmente os meses de Dezembro a Março. A taxa de evaporação é alta (1.600 a 2.300mm) e por esta via ocorre muita perda de água em extensas regiões planas inundadas e pântanos, especialmente ao sudoeste da bacia (JESSEN; SILVA, 2008).

Períodos de estiagem e chuva bem definidos, ocorrendo principalmente à montante, causam respectivamente períodos de secas e cheias no delta, influenciando por conseguinte os ciclos de sua flora e fauna, evitando a erosão e a salinização do solo, nutrindo pescueiros e definindo o modo de vida com que seus habitantes o cultivam. Estima-se que o delta abriga um terço dos mais de 1.5 milhão de moçambicanos que vivem ao longo do rio Zambeze (THA; SEAGER, 2013). Esses habitantes do delta encontram no rio seu objeto central de cultura e modo de vida. Assim como a fauna e a flora do delta, dependem do rio Zambeze para sua sobrevivência. Para além de coletar sua água para beber, lavar e cozinhar, eles cultivam suas culturas graças aos solos férteis e úmidos deixados quando da recessão da inundação; pescam os peixes que se criam e se abrigam no rio; criam animais

que de suas águas bebem diretamente ou por poços; coletam frutas para comer, plantas para curar e material para construir que estão a ele intrinsecamente conectados.

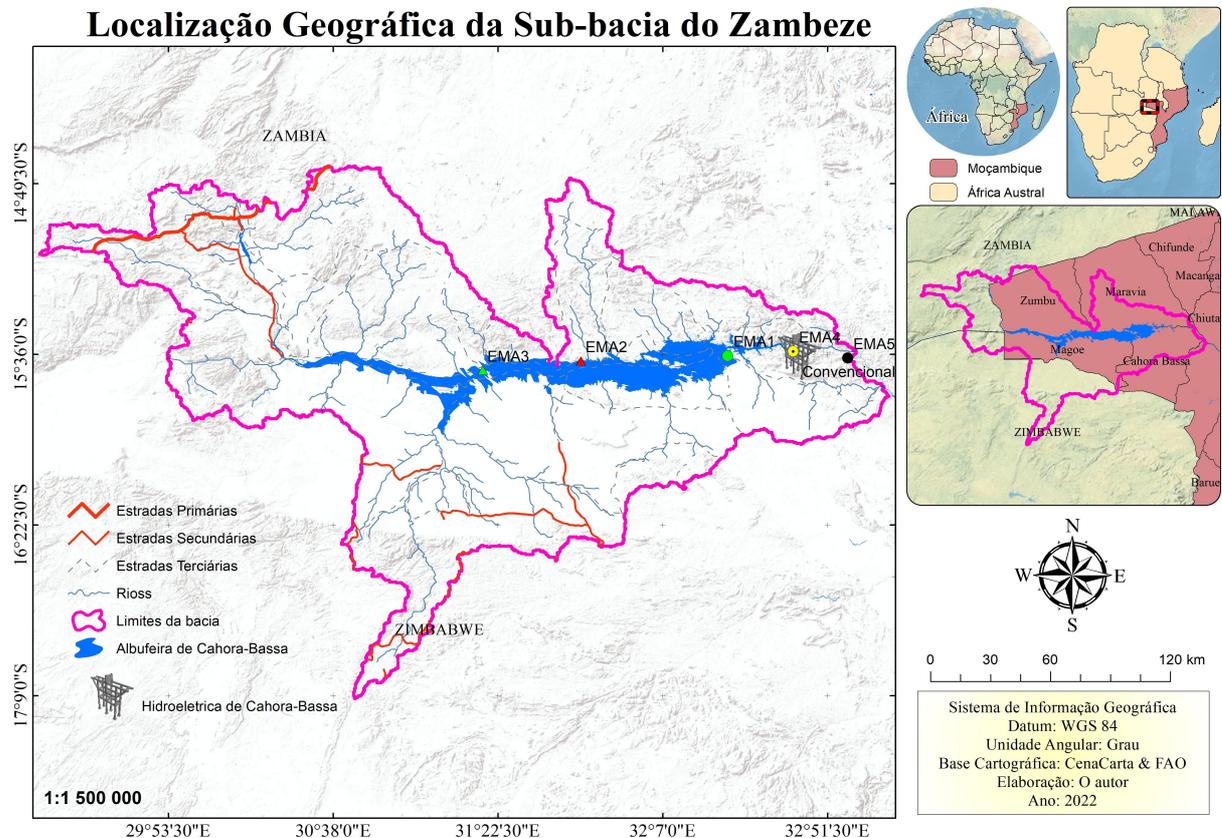
Além do aproveitamento no delta do Zambeze por parte dos habitantes, no trecho imediatamente a montante ( terminal Médio Zambeze) encontra-se a barragem de Cahora Bassa(Figura 7). Localizada no distrito de mesmo nome, província de Tete, em Moçambique, entre as coordenadas S e E, constitui um dos grandes empreendimentos hidroelétricos da África Austral . A albufeira criada em 1975, com uma capacidade máxima de  $65 \times 10^9 \text{m}^3$  e volume útil de  $52 \text{km}^3$ ,  $270 \text{km}$  de comprimento e  $30 \text{km}$  de largura máxima, com uma superfície de inundação de  $2900 \text{km}^2$  ao nível de máxima cheia, é a quarta maior barragem de África e 12º reservatório do mundo. É o maior produtor de eletricidade em Moçambique, com mais de 2000 megawatts, e abastece Moçambique, África do Sul, Zimbabwe e Malawi. Para além da produção de energia elétrica, contribui também para a manutenção de níveis adequados dos regimes hidrológicos, ambientais e ecológicos na albufeira e no delta, e para a manutenção da segurança das populações e bens, para a navegabilidade e para a redução da cheias e seca.

Figura 6 - Limites da Bacia Hidrográfica do Zambeze.



Fonte: Extraído de Jessen e Silva (2008)

Figura 7 - Localização da Área de estudo. Os pontos EMAs indicam as estações de monitoramento automáticas onde são recolhidos dados em análise neste trabalho.



### 3.2 CONJUNTO DE DADOS

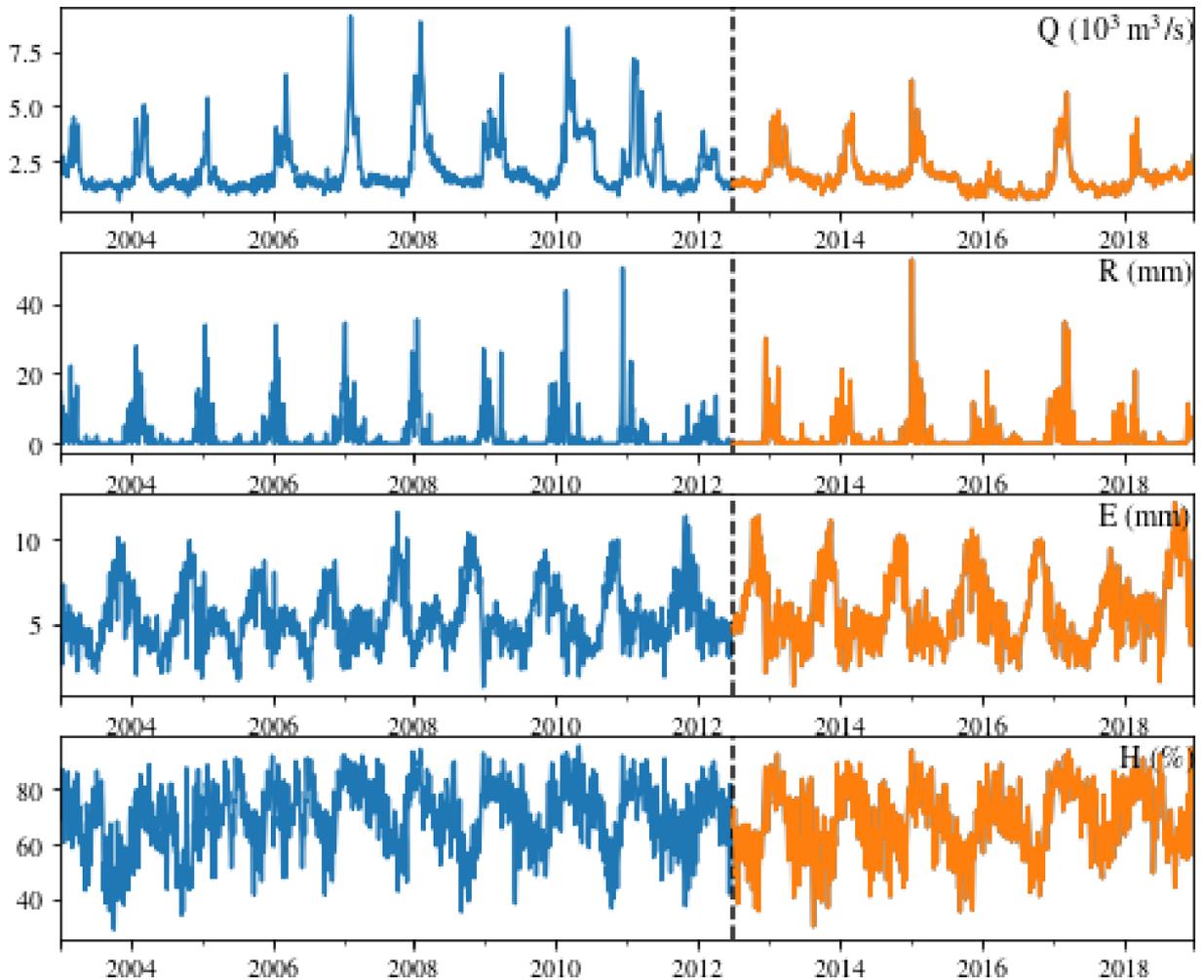
Os dados em análise nesta pesquisa foram fornecidos pelo Departamento de Recursos Hídricos e Meio Ambiente da Hidroelétrica de Cahora-Bassa (HCB), empresa gestora da barragem de Cahora-bassa e maior produtora de energia elétrica em Moçambique. Consistem em séries históricas referentes às variáveis: vazões afluentes ( $Q$ ), precipitação ( $R$ ), Evaporação ( $E$ ) e a umidade relativa ( $H$ ), constituídas por 5844 observações cada uma das variáveis, medidos diariamente. Os dados são correspondentes ao período entre 2003 e 2018, ou seja, um total de 15 anos. Este conjunto de dados, é por sua vez subdividido em dois subconjuntos sendo: de *treinamento* onde se efetua o ajuste do modelo, e de *teste* onde se faz a validação do mesmo. A Figura 8 mostra o conjunto de treinamento em azul que varia de 01/01/2003 a 30/06/2012 e o conjunto de teste varia de 01/07/2012 a 31/12/2018 e aparece em laranja, correspondente a 63.3% e 36.7%, respectivamente.

Uma agregação destes dados diários em médias mensais (Figura 9) mostrou um padrão de sazonalidade em cada uma das variáveis em análise, que é comum em variáveis hidrológicas, que indicam variações nas condições climáticas ou do tempo (estações do ano).

A sazonalidade é definida como um comportamento específico de uma série que tende a se repetir por períodos iguais a medida que o tempo avança (PARMEZAN; BATISTA, 2016). Geralmente, a presença desta propriedade pode dificultar o reconhecimento e a interpretação de movimentos não-sazonais peculiares numa série (como pontos de mudança e outros eventos cíclicos, novos padrões emergentes, ocorrências não esperadas para as quais causas possíveis são procuradas), utilizando modelos. A detecção da sazonalidade pode relevar informações valiosas da série temporal (CHATFIELD; XING, 2019).

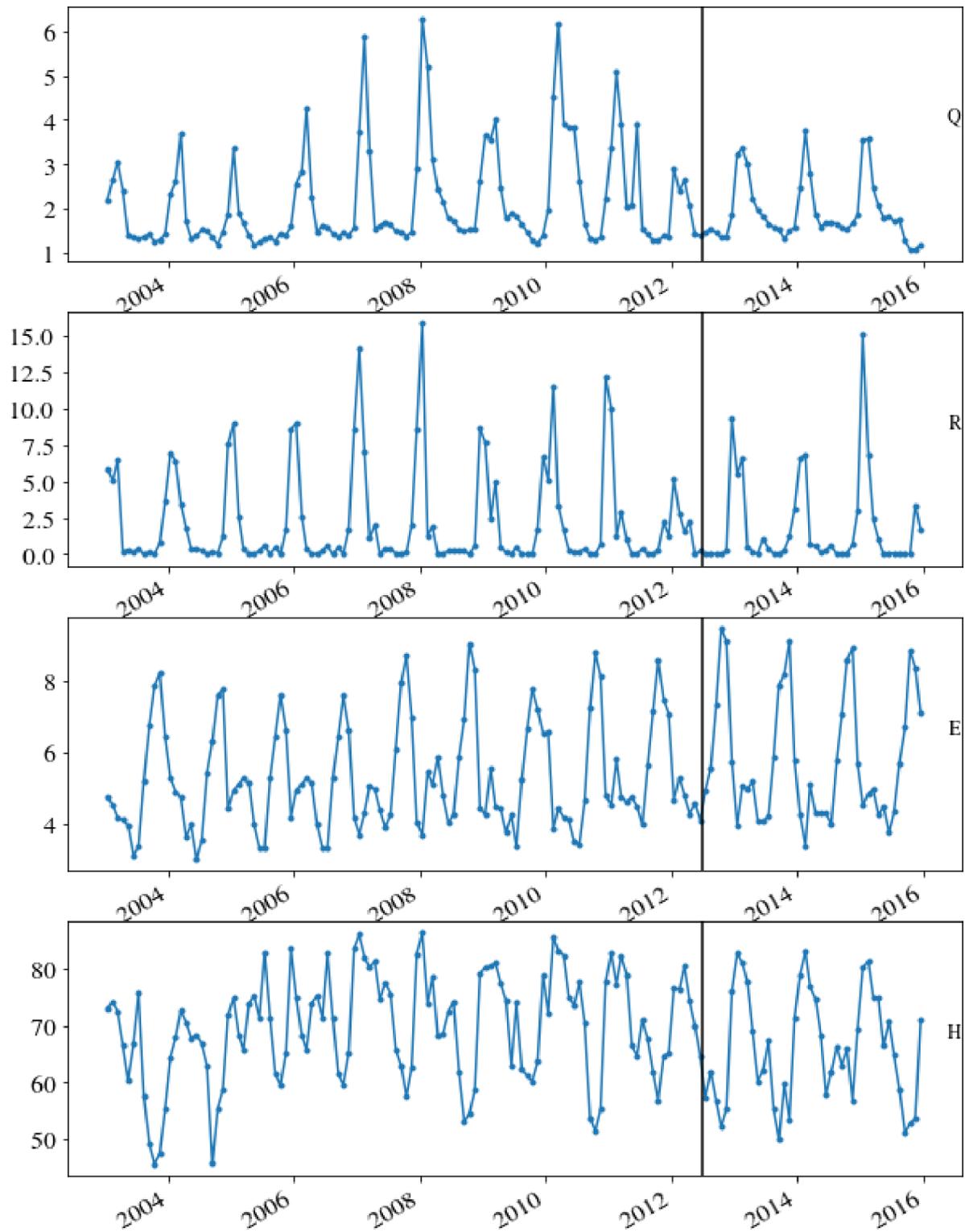
Há que destacar também, a existência de dados faltantes na base dados analisada. Apesar da proporção muito reduzida desses dados faltantes (menos de 5%), o tratamento foi feito mediante a imputação ou preenchimento dados, utilizando médias móveis aritméticas de tamanho  $n = 7$ . A técnica das médias móveis aritméticas é a abordagem mais amplamente aplicada em hidrologia, para atenuar as flutuações espúrias de curto período numa dada série, como para evidenciar as tendências ou os ciclos longos existentes em tal série (PORTELA et al., 2011).

Figura 8 - Dados históricos diários de vazões do rio Zambeze na barragem de Cahora Bassa. Onde  $Q$  é vazão afluente,  $R$  a precipitação,  $E$  a Evaporação e  $H$  a umidade relativa. Dados diários, entre 2003 e 2018 (15anos). De 01/01/2003 a 30/06/2012 treino (azul) (63.3%) e 01/07/2012 a 31/12/2018 teste (laranja) (36.7%).



Fonte: Departamento de Recursos Hídricos e Meio Ambiente da Empresa Hidroelétrica de Cahora-Bassa (HCB)

Figura 9 - Análise da sazonalidade das séries estudadas.  $Q$  é vazão afluente,  $R$  a precipitação,  $E$  a Evaporação e  $H$  a umidade relativa.



Fonte: Elaborada pelo autor (2023)

A estação de Cahora-Bassa é composta por cinco pontos ou subestações de monitoramento designadas por  $EMA_s$  (Estações de monitoramento da Autónomas): EMA1,

EMA2, EMA3, EMA4 e EMA5, localizados em Emboque/Chicoa, Katcholicholi, Carinde, Barragem e Songo, respectivamente, e uma convencional que também é localizada em Songo, como são mostrados espacialmente na Figura 7. Cada subestação de monitoramento registra variáveis específicas.

Em geral, os dados hidrológicos obtidos através da medição das variáveis hidrológicas como chuva, vazão e evaporação, são utilizados para acompanhamento do abastecimento de água, de descarte de efluentes, de controle de cheias e secas em rios urbanos, da construção de pontes, da análise da contaminação de rios, de controle ecológico, da autorização para usos de água (TEODORO et al., 2011) e, no contexto da geração de energia, para análise das séries históricas de vazão, cálculo do volume de reservatório, medição de vazão líquida e sólida, coleta e análise de sedimentos em suspensão, atualização das curvas cota-área-volume <sup>1</sup>, definição e atualização de curva-chave <sup>2</sup>, levantamento de medição de massas de água (batimetria da seção transversal) (SÁ, 2018).

A *vazão* ou *descarga* é uma importante componente de processo-base para realização de estudos hidrológicos, pois é componente do balanço hídrico na fase terrestre (ARNOLD et al., 2012). Esta vazão é o escoamento superficial medido em determinados pontos, ou o calculado a partir dos modelos hidrológicos. É relativamente oneroso medi-la diretamente, seja num intervalo de tempo curto, ou por longos períodos consecutivos. Há muitas alternativas de se medir tal descarga; o nível de água é a maneira indireta mais simples para obter tais resultados (SALES, 2019).

Quando são inexistentes aproveitamentos hidroelétricos à montante ou eles operam com nível constante, esta vazão é chamada de vazão natural afluente, ou seja, é a vazão que existia naturalmente antes de qualquer intervenção humana. A potência de toda usina hidroelétrica é proporcional à vazão de água disponível para movimentar as turbinas (FILHO et al., 2009).

Os equipamentos mais utilizados para medir a velocidade da água são os *molinetes* e os *Acoustic Doppler Current Profilers*(ADCPs). O molinete que consiste, basicamente, em uma hélice cuja rotação é proporcional à velocidade do líquido. O ADCP, que é utilizado para medir a velocidade através de transmissão de pulsações acústicas, um sonar que utiliza o som para medir a velocidade da água através do efeito Doppler(CARVALHO, 2008). O método para determinação da vazão consiste nos seguintes passos: divisão da seção do rio em um certo número de posições para levantamento do perfil de velocidades, levantamento do perfil de velocidades, cálculo da velocidade média de cada perfil, determinação da vazão pelo somatório do produto de cada velocidade média por sua área de influência (PEREIRA et al., 2003).

<sup>1</sup> Curva-área-volume é a curva que fornece o volume e a área de um reservatório em função do nível d'água no momento de medição.

<sup>2</sup> Curva-chave é a relação entre cota e descarga em uma seção.

A vazão afluente em Cahora-bassa é somatório resultante das contribuições da Sub-bacia do rio Luangwa, Sub-bacia do rio Kafue, Sub-bacia do rio Hunyani e afluentes diretos ao reservatório e pelo efluente gerado por Kariba, a barragem à montante de Cahora-Bassa, dependendo do seu estado de armazenamento. Os tributários da margem esquerda contribuem com volumes muito maiores de água do que os da margem direita (Figura 6). Esta vazão é obtida de forma indireta através do nível da água ou cota coletado na subestação de monitoramento situada na barragem (EMA4) entre  $-15.58694$  e  $32.70444$  de latitude e longitude, respectivamente, e altitude de  $334.687m$ .

A *precipitação* é a medida da quantidade de chuva caída em um determinado lugar da terra. Consiste em medir a altura de chuva em milímetros ou em polegadas, geralmente por meio de aparelhos denominados de *pluviômetros* e *pluviógrafos*. Os pluviômetros têm a função de colher uma amostra pontual da altura de chuva precipitada durante determinado período de tempo enquanto que, os pluviógrafos são instrumentos que colhendo uma amostra pontual de chuva, registram em papel apropriado as alturas de chuva em relação ao tempo (*pluviograma*) (PEREIRA et al., 2003).

Os dados de precipitação utilizados são coletados em uma estação meteorológica localizada em Carinde (EMA3) na latitude  $-15.67228$  e longitude  $31.30706$ . Uma característica dos dados desta variável é que, para além da sazonalidade observada, verificou-se a presença de valores nulos, indicando períodos sem precipitação ou de estiagem; um tratamento dos dados eliminando a influência destes e outros na etapa de modelagem, foi crucial.

A *evaporação* é um processo de transferência de água líquida para o vapor do ar diretamente de superfícies como rios, lagos e reservatórios. São necessárias duas condições para o processo de evaporação: que a água líquida receba energia na natureza, na forma de radiação solar, e que o ar acima da superfície líquida não esteja saturado de vapor de água. Os principais fatores da evaporação são a temperatura, a umidade do ar, a velocidade do vento e a radiação solar (COLLISCHONN; DORNELLES, 2013; BUENO et al., 2016).

Geralmente, não existem medidas diretas de evaporação suficientes para cada região e condição climática; portanto, estimativas baseadas em métodos *físicos* ou de relações *empíricas* que usam dados climáticos existentes (como métodos de transferência de energia, balanço de energia, balanço hídrico, empíricos, combinados e medidas diretas) são muitas vezes as alternativas disponíveis (BERTONI et al., 2007). Os métodos físicos são constituídos por equações que descrevem as relações físicas dos fenômenos, ao passo que, os empíricos são elaborados com base em ajustes de regressão das variáveis envolvidas (evaporação, velocidade do vento, pressão do vapor de água, radiação solar, entre outros).

Os métodos de medições diretas são geralmente utilizados através de instrumentos ou aparelhos como os tanques classe A, evaporímetros de Piche, entre outros; porém, são

relativamente caros e geralmente não confiáveis (TUCCI et al., 2004). Assim, devido aos altos custos e logística para coleta dos dados da evaporação, são utilizados dados pontuais, sem considerar a variação espacial sobre o reservatório. Como alternativa para este problema, técnicas com base em dados extraídos de imagens de satélite também tem sido combinados (CURTARELLI et al., 2013). Os dados de evaporação em Cahora-Bassa são estimados através de medidas diretas na subestação de monitoramento convencional, localizada em Songo, entre  $-15.61667$  e  $32.95000$  de latitude e longitude, e uma altitude de  $902.32m$ .

A *umidade relativa* do ar atmosférico é a relação entre a quantidade de vapor de água presente e a quantidade de vapor d'água que o mesmo volume de ar conteria se estivesse saturado, expresso em porcentagem (PEREIRA et al., 2003). Quanto maior for a quantidade de vapor de água no ar atmosférico, tanto maior o grau de umidade e menor a intensidade de evaporação. A umidade relativa pode também ser expressa pela relação entre as pressões parcial (instantânea) e de saturação (máxima) de vapor de água no ar (BUENO et al., 2016). A pressão de saturação de vapor de água depende da variação da temperatura do ar; portanto, quanto maior for a temperatura do ar, maior será a pressão de saturação de vapor de água e, por conseguinte, menor será a umidade relativa. Assim, há uma relação entre a umidade relativa do ar, a evaporação estimada e a temperatura do ar. Os dados de umidade relativa em Cahora-Bassa são obtidos através de medições também na subestação da barragem de Cahora-Bassa (EMA4).

Em substituição dos aparelhos ou instrumentos anteriormente mencionados, podem ser utilizados os *dataloggers* ou *Plataformas de Coleta de Dados*(PCD) (PEREIRA et al., 2003).

Uma Plataforma de Coleta de Dados (PCD) ou *datalogger* é um dispositivo que dispõe de sensores eletrônicos capazes de medir diversas variáveis ambientais, tais como precipitação, temperatura, radiação solar, pressão atmosférica, variação dos níveis de corpos d'água, entre outras. É amplamente utilizado em meteorologia, hidrologia, agricultura, silvicultura, pesquisa científica e outros campos. As PCDs são equipadas com sistemas automáticos que coletam as informações do ambiente, armazenam e disponibilizam os dados, utilizando redes de telefonia ou satélite. A combinação de sensores com uma PCD é chamada de uma *estação de coleta*, que deve ser composta minimamente pelos seguintes equipamentos: um registrador de dados, um sensor de chuva, um sensor de nível d'água, um controlador de carga de bateria, uma bateria estacionária, um painel solar e uma antena (PEREIRA, 2022).

Entre as vantagens de uso dessas estações de coleta destacam-se a precisão nos dados obtidos, a implantação remota do sistema e o baixo consumo, a conexão remota com muitas opções de comunicação suportadas e sistemas personalizáveis para necessidades específicas de aplicação. Contudo, é relativamente alto o orçamento para aquisição e exige

custo manutenção e atualizações para versões mais recentes relativamente constantes.

Os dados em análise neste trabalho são coletados por *dataloggers* ou plataformas de coleta de dados das séries *CR800* e *CR1000*, em estações de coleta autônomas com as descrições: *WxPRO/Entry-Level, Research-Grade* ou *Weather Station*, com transmissão de dados através de periféricos de comunicação *Ethernet*. As séries *CR800* e *CR1000* são usadas em monitoramento climatológico de longo prazo, pesquisa meteorológica e aplicações de medição meteorológica de rotina, também são adequadas para monitoramento remoto e autônomo das condições hidrológicas. A maioria dos sensores hidrológicos comunicam diretamente com o registrador de dados (SCIENTIFIC, 2010). A Figura 10 mostra uma estação de coleta de dados autônoma em Cahora-Bassa.

Figura 10 - Exemplo de um *datalogger* CR1000 e uma Estação de monitoramento autônoma de Emboque/Chicoa (Cahora-Bassa).



Fonte: Departamento de Recursos Hídricos e Meio Ambiente da HCB

### 3.3 MODELOS DE SÉRIES TEMPORAIS EM HIDROLOGIA

As *séries temporais* constituem na atualidade um tipo de estrutura de dados muito comum e frequente em diversas áreas da ciência, e sua análise continua sendo um grande desafio.

As respostas às perguntas como: A bolsa de valores vai subir? Amanhã vai chover? Pode-se saber quando ocorrerá uma enchente? O valor do petróleo vai subir?, podem

significar: redução de custo, maior lucro, menor prejuízo, melhor planejamento, melhoria de processos tornando seu estudo um grande atrativo.

As *séries temporais* são conjuntos de observações de algum fenômeno, ordenados no tempo. Estas podem ser consideradas *contínuas* quando a coleta dos dados é realizada continuamente no tempo, e *discretas* quando a coleta desses dados é realizada em tempos específicos, geralmente equiespaçados. Além disso, a variável observada pode assumir valores discretos ou contínuos (BOX et al., 2015; DUARTE et al., 2019).

Dada uma série temporal  $Z(t_1), \dots, Z(t_N)$  observada nos instantes  $t_1, \dots, t_N$  pode-se estar interessado em (BOX et al., 2015; BAYER, 2010):

- a) investigar o mecanismo gerador da série temporal;
- b) fazer previsões de valores futuros da série podendo ser a curto ou longo prazo;
- c) descrever apenas o comportamento da série através de gráficos;
- d) procurar periodicidades relevantes nos dados.

Portanto, a modelagem e previsão deste tipo de estrutura de dados consiste na identificação do processo gerador da série, estimação de parâmetros do modelo que a representa e a avaliação da qualidade do ajuste (FIGUEREDO, 2008).

Dois enfoques são utilizados na análise de séries temporais com o objectivo de construir modelos: a análise feita no *domínio temporal*, na qual os modelos propostos são paramétricos (com um número finito de parâmetros) e, análise no *domínio de frequências*, na qual os modelos propostos são não-paramétricos (BEZERRA, 2006).

As abordagens de previsão de séries temporais existentes na literatura podem ser organizadas conforme (SOUZA, 2008; DUARTE et al., 2019):

- a) **técnicas Descritivas:** consistem em analisar uma ou mais séries temporais através da representação e análise gráfica dos dados sequencialmente ao longo do tempo, o que pode revelar padrões de comportamento importantes: a tendência de crescimento (ou decrescimento), padrões cíclicos, alterações estruturais, observações aberrantes, entre outros;
- b) **modelos lineares:** incluem modelos probabilísticos, análise espectral, métodos não- paramétricos (alisamento ou suavização), modelos de espaço de estados, séries multivariadas, estudos longitudinais e processos de longa dependência;
- c) **modelos não-lineares:** englobam modelos não-lineares gerais (redes neurais artificiais, sistemas nebulosos, filtro de Kalman estendido, modelos híbridos), modelos pré-definidos, modelos com volatilidade variável, entre outros.

Uma das tarefas mais importantes ao analisar qualquer série temporal é descrever e resumir os dados da mesma. As estatísticas preliminares frequentemente descritas

incluem: medidas de tendência central, de dispersão ou variabilidade, de simetria da distribuição e das estimativas de extremos dos dados. Estas estatísticas têm o intuito de confirmar se as séries temporais possuem todas as suposições ou características requeridas. As características requeridas são: homogeneidade, aleatoriedade, não-sazonalidade, não-tendência e estacionariedade. Contudo, estas suposições parecem ser a exceção em vez do que a regra para séries temporais hidrológicas (MACHIWAL; JHA, 2012).

- a) **homogeneidade** implica que os dados da série pertencem à mesma população estatística, com média invariante no tempo. Os testes para verificar a homogeneidade ou consistência dos dados de uma série são baseadas na avaliação da significância das mudanças no valor médio em diferentes subconjuntos extraídos da série original, como por exemplo, os testes de comparações múltiplas de Tukey, Dunnett, Bartlett, entre outros.
- b) **aleatoriedade** é definida como a independência entre os dados de uma série temporal e é medido por autocorrelação (ou seja, correlação entre dois termos da mesma série temporal).
- c) **sazonalidade** detectada quando as séries temporais forem definidas em intervalos de tempo inferiores ou iguais a um ano; na maioria dos casos, seis e 12 meses. A sazonalidade é muito comum. A série Fourier, a análise da função de autocorrelação periódica, a análise espectral (onde a função de densidade espectral pode ser usada em substituição da Transformação de Fourier da função de autocorrelação) têm sido usadas para o detecção de componentes sazonais na série.
- d) **tendência** é uma componente comum que é uma variação de valores sucessivos: aumento ou diminuição ao longo do tempo. A abordagem mais comumente usada para detecção de tendência é formular um modelo entre os dados e o tempo, na forma de regressão linear.
- e) **estacionariedade**: Uma série temporal é considerada *estritamente estacionária* se suas propriedades estatísticas não variam com as mudanças da origem do tempo. A tendência ou sazonalidade constituem formas de análise da estacionariedade da série. Duas abordagens são utilizadas para verificar a estacionariedade: paramétrica, fundamentada pelas suposições de normalidade e igualdade das médias de subconjuntos extraídos da série, aplicando teste paramétricos, como t-student; e não-paramétrico, usando testes alternativos como Mann-Whitney.

A aplicação dos métodos de análise de séries temporais na análise de séries temporais hidrológicas como precipitação, fluxo, infiltração, qualidade da água superficial, consumo de água, evapotranspiração e vento aumentou dramaticamente no passado recente. Devido

à procura crescente da compreensão científica das alterações climáticas, estes modelos têm sido utilizados para gerar simulações de dados hidrológicos, para prever eventos hidrológicos, identificar tendências e mudanças em dados hidrológicos, preencher as observações faltantes e para estender os dados hidrológicos com registros curtos, tornando-se uma ferramenta vital em ciências hidrológicas. A Estatística é apenas uma das várias ferramentas disponíveis para aplicação em ciências hidrológicas.

Em hidrologia, a modelagem de séries temporais pode incluir modelos estocásticos, isto é, estatísticos, como modelos autoregressivos (AR), modelos de média móvel (MA), modelos autoregressivos de médias móveis (ARMA), modelos autoregressivos integrados de médias móveis (ARIMA), modelos Sazonais autoregressivos integrados de média móvel (SARIMA), entre outras variações (BOX, 1976; BOX et al., 2015; MACHIWAL; JHA, 2012).

Contudo, evidências recentes de tendências não-estacionárias, não-lineares, estocásticas complexas e caóticas em séries temporais de recursos hídricos, como resultado da variabilidade e mudança natural e/ou antropogênica do clima, têm despertado mais interesse em métodos de modelagem de sistemas dinâmicos não-lineares. Os modelos de aprendizado de máquinas, os modelos de redes neurais artificiais, modelos de aprendizagem profunda, entre outros acionados dinamicamente podem ser boas alternativas para a modelagem da dinâmica complexa de um sistema hidrológico com melhor desempenho (MACHIWAL; JHA, 2012).

Entre as vantagens de métodos de aprendizado de máquina em relação aos métodos estatísticos podem ser citados os seguintes: modelos resultantes mais resistentes à multicolinearidade e valores extremos, incluem métodos que reduzem *overfitting*, melhor identificação das variáveis preditoras importantes das relações não-lineares e das interações complexas entre preditores, não serem afetados por transformações de dados (POVAK et al., 2013).

### 3.4 MODELOS DE APRENDIZADO DE MÁQUINAS PARA PREVISÃO HIDROLÓGICA

Vários fatores influenciam a escolha de modelos em hidrologia: a disponibilidade de dados, características físicas da bacia hidrográfica e objetivo da previsão (CASAGRANDE et al., 2017). Outros critérios também podem ser considerados e combinados com os anteriores como o balanço entre : precisão, confiabilidade e generalidade (WHITE, 2017).

Porém, não existe em hidrologia uma recomendação de aplicação universal de um modelo empírico a um dado problema. Na medida em que aplicações relacionadas à água são geralmente também caracterizadas por dados de má qualidade (ou ruído) é recomendada a aplicação de vários tipos de técnicas e a utilização dos resultados de forma comparativa e complementar (SOLOMATINE; XUE, 2004). Devido às suas vantagens de

alta precisão de predição e excelente desempenho ao lidar com problemas não-lineares de alta dimensão, Modelos de aprendizado de máquina (ML) são amplamente utilizados por muitos pesquisadores (SONG et al., 2021; ZHOU et al., 2021).

Em geral, os modelos ML podem ser categorizados em modelos de ML gerais e profundos. Os ML gerais são modelos que usam técnicas (como o aprendizado profundo) que permitem que os computadores usem a experiência para aprimorar as tarefas; os ML profundos são subconjuntos dos ML gerais baseado em redes neurais que permitem que uma máquina se treine para executar uma tarefa (SHINDE; SHAH, 2018). Os MLs profundos geram maior precisão do que MLs gerais. Porém, MLs profundos têm problemas de maior custo computacional, maior tempo computacional, maiores requisitos de hardware, resultando em sua portabilidade deficiente e tornando sua aplicação ainda um longo caminho a percorrer para aprofundá-los (ZHOU et al., 2021).

A seguir se apresentam os métodos de aprendizagem de máquina explorados neste estudo que também se destacam muito na área de recursos hídricos, support vector regression(SVR), extreme learning machine(ELM), Multi-Adaptive spline Regression(MARS), Extreme gradient boosting(XBG) e Elastic net linear model(EN).

### 3.4.1 ELASTIC NET(EN)

Dado um problema de regressão com alta dimensionalidade, em que  $y \sim N(Xw, \sigma^2 I_n)$ , com pesos  $w = (w_1, \dots, w_p)$  de dimensão  $p \times 1$ , matriz de entradas  $X$  de dimensão  $n \times p$  e a saída  $y = (y_1, \dots, y_n)$  de dimensão  $n \times 1$ . Uma solução para esse problema é estimar  $w$  pelo método de mínimos quadrados ordinários. Porém, essas estimativas fornecem resultados pouco viesados e com alta variância. Nesse caso, uma abordagem muito utilizada para obter melhora nos resultados é considerar mínimos quadrados penalizados, ou seja,  $w$  é encontrado de maneiras que minimize

$$G_y(w) = \frac{1}{2N} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2 + \beta \sum_1^p f(w_j^2) \quad (3.1)$$

para uma função de penalidade dada por  $f(w_j^2) = (\sqrt{w_j^2})^\gamma = (|w_j^2|)^\gamma$  e parâmetro de penalidade  $\beta \geq 0$ . Isto é, um problema de otimização da soma dos quadrados dos resíduos sujeita a restrição na norma  $L_\alpha$  dos pesos  $w$  (POPOCA et al., 2013). Para  $\gamma = 2$ , temos a regressão *ridge* (HOERL; KENNARD, 1970), regressão *lasso* quando  $\gamma = 1$  (TIBSHIRANI, 2011) e a que generaliza as duas anteriores, *Elastic net*.

Proposto por Zou e Hastie (2005), EN é uma regressão linear generalizada (HASTIE et al., 2009b) que calcula os coeficientes de um modelo linear  $\mathbf{w} = [w_0, w_1, w_2, \dots, w_p]$  enquanto minimiza a soma dos quadrados dos resíduos entre as saídas no conjunto de dados e a saída prevista pelo modelo EN.

Simbolicamente, o modelo linear generalizado pode ser declarado como:

$$\min_w \frac{1}{2N} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \frac{\alpha\rho}{2} (2\|\mathbf{w}\|_1 - \|\mathbf{w}\|_2^2) + \frac{\alpha}{2} \|\mathbf{w}\|_2^2 \quad (3.2)$$

Onde  $N$  representa a quantidade de amostras,  $\mathbf{X}$  é a matriz de pontos de dados,  $\mathbf{y}$  são os vetores de saída,  $\alpha \geq 0$ ,  $\|\mathbf{w}\|_2$  e  $\|\mathbf{w}\|_1$  são respectivamente a norma  $L_2$  e a norma  $L_1$  do vetor de parâmetro, e  $\rho$  é razão do parâmetro  $L_1$ .

*Elastic Net* é um modelo linear ajustado com regularização  $L_1$  e  $L_2$  no processo de treinamento, isto é, uma combinação linear convexa exata da função de penalidade da regressão *Ridge* e de *Least Absolute Shrinkage and Selection Operator (LASSO)*, respectivamente. Essa combinação permite o aprendizado de um modelo esparsos com poucos pesos diferentes de zero enquanto preserva as qualidades de regularização ou seja, permite simultaneamente uma efetiva regularização em grupos de preditores fortemente correlacionados e qualidade da seleção de características (ZOU; HASTIE, 2005). Esta compensação é regulada aplicando o parâmetro  $\rho$ , denominada a razão  $L_1$ . Entende-se por esparsidade a suposição de que muitos parâmetros podem assumir o valor zero ou valores muito próximos de zero, sem afetar o ajuste do modelo (JUNIOR, 2021).

Portanto, o *Elastic Net* é um estimador que reduz todas as estimativas de parâmetros em torno do zero, enquanto penaliza mais os parâmetros menores e maiores, isto é conseguido de uma forma que reduz o erro quadrado médio (MSE) do estimador de parâmetros e dos erros de previsão do modelo. Estas qualidades permitem a utilização deste modelo para fins de seleção de preditores de maior poder preditivo em problemas de regressão de alta dimensionalidade resolvendo problemas de *overfitting* e multicolinearidade (ZOU; HASTIE, 2005). Geralmente é usado como modelo de base para comparação com outros modelos (AL-JAWARNEH et al., 2021) ou é combinado com outros modelos resultando em modelos híbridos (LIU et al., 2018). A aplicação deste modelo em dados de séries temporais pode ser encontrada em estudos referentes à previsão de variáveis económicas e financeiras (MASINI et al., 2021), Macroeconómicas (GODEIRO et al., 2018), previsão de curto prazo da carga elétrica (LIU et al., 2018), previsão em séries biológicas (GONZALES; SAEGER, 2018), entre outros. Porém, *Elastic net* é uma boa opção quando múltiplas características são fortemente correlacionadas entre si (ZOU; HASTIE, 2005), incluindo também uma ferramenta para extrair de características.

### 3.4.2 EXTREME LEARNING MACHINE (ELM)

Inicialmente proposto por (HUANG et al., 2004), o ELM é uma rede neural artificial do tipo *single-hidden layer feedforward networks (SLFN)*, onde a maioria consiste em apenas uma camada oculta, com topologia menos complexa é uma das técnicas de aprendizado supervisionado mais usadas.

Sua eficiência computacional e robustez permitiu a sua rápida popularidade. Em comparação com a Rede Neural Artificial (RNA) padrão, o ELM tem os benefícios de

um aprendizado rápido e uma forte capacidade de generalização (GUO et al., 2017). Esses modelos, segundo os autores podem apresentar bom desempenho de generalização e aprender milhares de vezes mais rápido do que as redes de retropropagação.

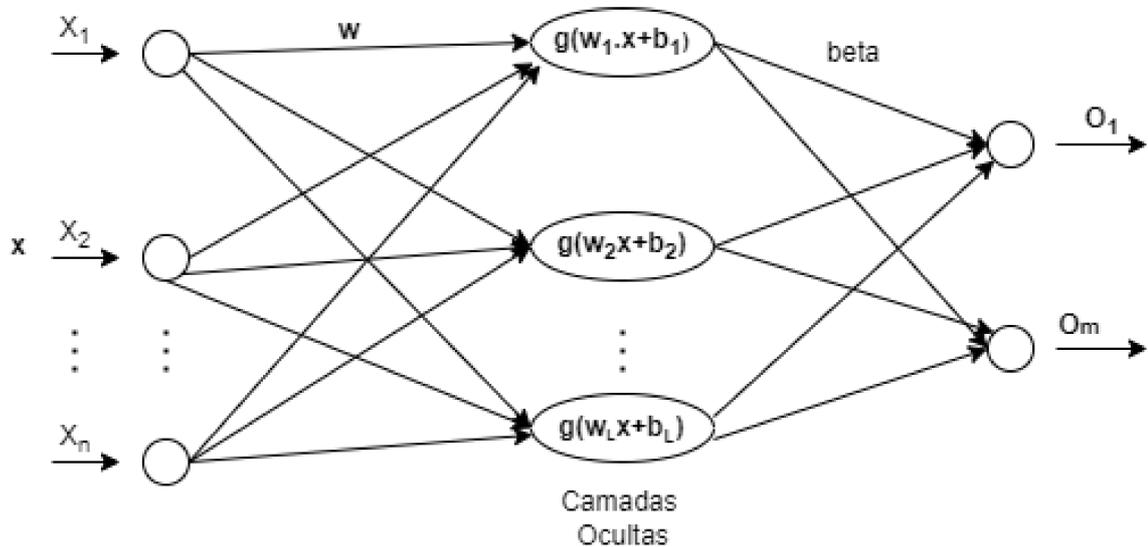
Além disso, esses modelos podem obter melhor desempenho do que máquinas de vetor de suporte em problemas de classificação e regressão.

No ELM existem três níveis de aleatoriedade:

- totalmente conectado, os parâmetros de nós ocultos são gerados aleatoriamente;
- a conexão entre as entradas para nós ocultos pode ser gerada aleatoriamente, nem todos os nós de entrada estão conectados a um nó oculto particular;
- um nó oculto em si pode ser uma sub-rede formada por vários nós resultando no aprendizado de características locais.

O ELM tem como objetivo realizar tarefas supervisionadas em conjuntos de dados. A ideia é que dado  $N$  distintas amostras  $(\mathbf{x}_i, \mathbf{y}_i) \forall i \in \{1, 2, \dots, N\}$  onde  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}] \in \mathbf{R}^n$  e  $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{im}] \in \mathbf{R}^m$ , um modelo de regressão ou classificação possa ser construído ajustado a esses dados. A estrutura do ELM, conforme sugerido por (AKHTER et al., 2019) pode ser observado pela Figura 11.

Figura 11 - Representação do esquema de rede do ELM



Fonte: Extraído de Akhter et al. (2019)

De acordo com Huang et al. (2004), o modelo considera um SLFN padrão e são escolhidos  $L$  neurônios na camada oculta de tal forma que é possível construí-la a partir de uma matriz de pesos aleatórios. O modelo matemático determinado pela equação 3.3 considera as  $N$  amostras com erro nulo, ou seja,  $\sum_{j=1}^N \|\mathbf{y}_j - \mathbf{t}_j\| = 0$ . Isso significa que existe um  $\beta_i$ ,  $\mathbf{w}_i$  e  $b_i$  de tal forma que:

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, \dots, N \quad (3.3)$$

onde  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  representa o  $i$ -ésimo neurônio na camada oculta e  $i \in \{1, 2, \dots, L\}$ ,  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  é o peso da conexão do  $i$ -ésimo neurônio da camada oculta e o neurônio da camada de saída, e o  $b_i$  é o *bias* do  $i$ -ésimo neurônio da camada oculta. Além disso,  $g(\cdot)$  denota uma função de ativação. No presente trabalho aplicam-se as funções de ativação dispostas na Tabela. 5. Além disso, a equação inicial descrita pela 3.3 pode ser representada em notação matricial como  $\mathbf{H}\beta = \mathbf{T}$ , onde:

$$\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (3.4)$$

$$\beta = [\beta_1 \cdots \beta_L]_{L \times m}^T \quad \text{e} \quad \mathbf{T} = [\mathbf{t}_1 \cdots \mathbf{t}_N]_{N \times m}^T \quad (3.5)$$

Tabela 5 – Funções de ativação usadas no ELM.

Nome	Função de ativação $g$
1 ReLU	$g(\mathbf{w}_i, b, \mathbf{x}_j) = \max_i(0, (\mathbf{w}_i \cdot \mathbf{x}_j + b))$
2 Sigmoide	$g(\mathbf{w}_i, b, \mathbf{x}_j) = \frac{1}{1 + \exp(-\mathbf{w}_i \cdot \mathbf{x}_j + b)}$
3 Tangente Hiperbólica (tanh)	$g(\mathbf{w}_i, b, \mathbf{x}_j) = \frac{1 - \exp(\mathbf{w}_i \cdot \mathbf{x}_j + b)}{1 + \exp(\mathbf{w}_i \cdot \mathbf{x}_j + b)}$
4 Gaussiana	$g(\mathbf{w}, b, \mathbf{x}) = \exp(-(\mathbf{w} \cdot \mathbf{x} + b)^2)$
5 Multiquadricas	$g(\mathbf{w}, b, \mathbf{x}) = \sqrt{\ \mathbf{w} - \mathbf{x}\ ^2 + b^2}$
6 Multiquadricas inversas	$g(\mathbf{w}, b, \mathbf{x}) = 1/(\ \mathbf{w} - \mathbf{x}\ ^2 + b^2)^{1/2}$
7 Swish	$g(\mathbf{w}, b, \mathbf{x}) = \frac{\ \mathbf{w} \cdot \mathbf{x} + b\ }{1 + \exp(-\mathbf{w} \cdot \mathbf{x} + b)}$
8 Identity	$G(\mathbf{w}, b, \mathbf{x}) = \ \mathbf{w} \cdot \mathbf{x} + b\ $

Fonte: Extraído de Akhter et al. (2019)

Conforme pode ser observado na Figura 11,  $\mathbf{H}$  é referente a saída da camada oculta da rede neural e cada coluna de  $\mathbf{H}$  diz respeito ao  $i$ -ésimo neurônio da saída da camada oculta em relação a entrada  $\mathbf{x}_i$ . Partindo da notação matricial do problema, o valor de  $\beta$  pode ser encontrado utilizando o método dos mínimos quadrados, ou seja, o valor ótimo do problema de minimização descrito pela 3.6

$$\min_{\beta \in \mathbf{R}^L} (\|\mathbf{H}\beta - \mathbf{T}\|). \quad (3.6)$$

Dessa forma, o valor de  $\beta$  pode ser escrito como em 3.7, onde  $\mathbf{H}^\dagger$  é a pseudoinversa de  $\mathbf{H}$ .

$$\beta = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{T} = \mathbf{H}^\dagger \mathbf{T} \quad (3.7)$$

O treinamento básico do ELM consiste em duas etapas: inicialização aleatória e solução de parâmetros (WANG et al., 2021a; DING et al., 2015; ALBADRA; TIUNA, 2017). Em primeiro lugar, o ELM usa parâmetros aleatórios  $w_i$  e  $b_i$  em sua camada oculta, e eles ficam armazenados durante todo o processo de treinamento. O vetor de entrada é mapeado em um espaço de recursos com configurações aleatórias e funções de ativação não-lineares que é mais eficiente do que os parâmetros treinados. Com funções de ativação contínuas por partes não lineares, o ELM tem a capacidade de aproximação universal (HUANG et al., 2006). Na segunda etapa,  $\beta_i$  pode ser obtido pela inversa generalizada de Moore-Penrose, pois  $H\beta = T$  é um problema linear. Assim, o treinamento do ELM pode ser resumido na Tabela 6.

Tabela 6 – Treinamento do ELM.

Etapas de Treinamento do ELM
<b>Conjunto de treino:</b> $S = (\mathbf{x}_i, \mathbf{y}_i): \mathbf{x}_i \in \mathbf{R}^n$ e $\mathbf{y}_i \in \mathbf{R}^m \forall i \in \{1, 2, \dots, N\}$
<b>Inicialização:</b> atribua valores aleatórios ao peso oculto $w_i$ e viés $b_i$ e calcule a matriz de saída da camada oculta usando o conjunto de treinamento
<b>Solução analítica:</b> Encontre $\beta_i$ a partir de $H\beta = T$ pela inversa generalizada de Moore-Penrose.
$\beta = \mathbf{H}^\dagger \mathbf{T}$

Fonte: Extraído de Huang et al. (2006)

Os nós ocultos aleatórios permitem que o ELM convirja muito mais rápido, porém podem resultar em flutuação no desempenho. O Cálculo do peso de saída  $\beta_i$  é outro problema significativo relacionados ao desempenho do ELM. Estes dois pontos, permitiram o desenvolvimento de variantes dedicadas a melhorar a robustez e a estabilidade do ELM com foco na otimização de parâmetros, métodos robustos para o tratamento de nós ocultos e alterações de estrutura para tarefas específicas tais como aprendizagem online, dados desbalanceados, transferência de aprendizado, e *big data* (NILESH; SUNIL, 2021; DING et al., 2015).

Devido à sua superioridade na velocidade de treinamento, precisão e generalização, o modelo ELM é muitas vezes preferido do que outras abordagens para problemas relacionados à inteligência artificial tais como reconhecimento, previsão, aprendizagem de representação/característica, agrupamento (RATHOD; WANKHADE, 2021). Como resultado foi aplicado em uma ampla gama de problemas de diferentes campos como medicina (em aplicações em campos de imagens médicas, como ressonância magnética, tomografia computadorizada, ultrassom, mamografia), química (na avaliação da gestão verde na geração de energia, previsão de variáveis em processos químicos, previsão de propriedades não lineares de um sistema óptico), transporte (na otimização o sistema de transporte, detecção em tempo real de distração na condução, previsão de temperatura da

superfície da estrada, previsão de atraso dinâmico da rede ferroviária em grande escala, análise de acidentes de trânsito com base em dados de vídeo, na pesquisa de melhores rotas para motorista de táxi com dados de GPS (Sistema de Posicionamento Global), previsão do fluxo de tráfego para motoristas e governos), economia (na previsão do crescimento econômico com dados de comércio, importação e exportação, dados de desenvolvimento de ciência e tecnologia e investimento nas nações, dados de agricultura, manufatura e indústria e dados de demanda e preço do mercado de recursos energéticos, análise da emissão de CO<sub>2</sub> para prever desenvolvimento Econômico), robótica (controle de braços robóticos, interface cérebro-computador, controle de movimento robótico), internet das coisas (detecção de ataque cibernético, gestão de tarefas, geografia (geração de índices de suscetibilidade a deslizamentos de terra, previsão da capacidade de transporte de sedimentos, análise de regras de operação de reservatórios hidrelétrico), indústria alimentar (monitoramento de segurança alimentar, classificação de diferentes tipos de vinhos, previsão durante a análise de amostragem de alimentos em larga escala), entre outros (ALBADRA; TIUNA, 2017).

### 3.4.3 *SUPPORT VECTOR REGRESSION(SVR)*

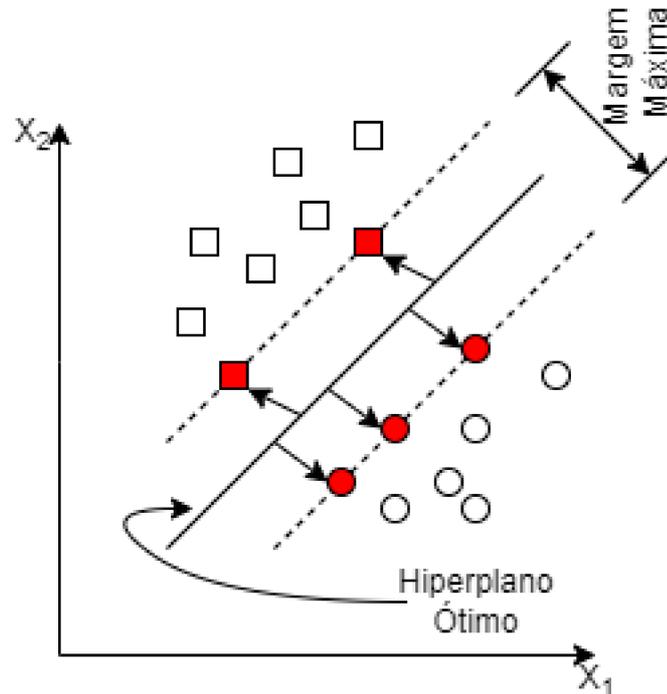
Fundamentado na teoria de aprendizagem estatística, SVR é um método clássico de regressão originalmente proposto por (VAPNIK et al., 1997) em 1995, que caracteriza propriedades de máquinas de aprendizagem que lhes permitem generalizar bem para dados não vistos (CHANG; LIN, 2011) aplicado em vários campos de conhecimento (KARTHIKEYAN; VYAS, 2014). Inicialmente concentrado em reconhecimento de caracteres, em pouco tempo tornou-se competitivo com os melhores sistemas disponíveis para tarefas de reconhecimento de objetos. Também em aplicações de previsão de regressão e séries temporais, excelentes desempenhos foram logo obtidos (SMOLA; SCHÖLKOPF, 2004).

Alguns dos motivos para esse sucesso estão relacionados ao fato dessa técnica exibir bom desempenho de generalização em muitas bases de dados reais, é bem fundamentada teoricamente, o processo de treinamento elimina a possibilidade de mínimos locais, existem poucos parâmetros livres para ajustar e a arquitetura não precisa ser encontrada por experimentação (SANTOS et al., 2002).

A idéia básica do SVR é mapear os dados  $X$  em um espaço de características  $F$  com alta dimensionalidade, através de uma função de mapeamento não-linear, e construir uma hiperplano ideal no novo espaço (RADHIKA; SHASHI, 2009). As máquinas de vetores de suporte podem ser aplicadas tanto à classificação quanto à regressão. No caso de classificação, um hiperplano ótimo é encontrado que separa os dados em duas classes(Figura 12) . Enquanto no caso de regressão, deve ser construído um hiperplano próximo a tantos pontos possível (LIU et al., 2017).

Regressão é o problema de ajustar uma função linear ou não-linear em um determinado conjunto de dados.

Figura 12 - Estrutura do modelo SVR.



Fonte: Extraído de Vapnik et al. (1997)

Consideremos o conjunto de dados  $G = \{x_i, y_i\}_{i=1}^N$  onde  $x_i$  é o vetor de entrada,  $y_i$  é o resultado desejado e  $N$  corresponde ao tamanho do conjunto de dados. A forma geral da função de estimativa da regressão do vetor de suporte pode ser expressa da seguinte forma:

$$f(\mathbf{x}) = (w \cdot \phi(\mathbf{x})) + b$$

onde  $\phi(\mathbf{x})$  é uma função kernel ou uma transformação não-linear em um espaço de entradas,  $\mathbf{w} = [w_1, \dots, w_N]$  é o vetor de pesos,  $b$  é um viés e  $N$  é o número de amostras.

O problema de regressão não-linear pode ser expresso como um problema de otimização minimizando a função de risco regularizada, equação 3.8 (KARGAR; AL., 2020). O  $\mathbf{w}$  e  $b$  ideais são calculados:

$$J(C) = \sum_{i=1}^N w_i^2 + \frac{C}{N} \sum_{i=1}^N L_\epsilon(y_i, f(\mathbf{x}_i)) \quad (3.8)$$

com

$$L_\epsilon(y, f(\mathbf{x})) = \begin{cases} |y - f(\mathbf{x})| - \epsilon & \text{se } |y - f(\mathbf{x})| \geq \epsilon \\ 0 & \text{outros casos} \end{cases}$$

onde  $y_i$  são os dados de saída associados a  $\mathbf{x}_i$ ,  $L_\epsilon$  é a função de perda  $\epsilon$  insensível (GUNN et al., 1998) e  $\epsilon$  é um parâmetro SVR.

Depois da introdução de variáveis de folga, a função de risco pode ser expressa na seguinte forma restrita:

$$J(w, \xi^*) = \sum_{i=1}^N w_i^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (3.9)$$

sujeita a

$$\begin{cases} y_i - w \cdot \phi(\mathbf{x}_i) - b_i \leq \varepsilon + \xi_i \\ w \cdot \phi(\mathbf{x}_i) - b_i - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

A forma dual do SVR não-linear pode ser expressa como

$$Q(\alpha_i, \alpha_i^*) = \sum_i^N y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \quad (3.10)$$

sujeito a

$$\begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C, \text{ com } i = 1, 2, \dots, N \end{cases}$$

onde  $\alpha$  e  $\alpha^*$  são os multiplicadores de Lagrange, também conhecidas como variáveis duais, que atuam como forças que aproximam as previsões para os valores observados  $y$ .

A computação no espaço de entrada é realizada usando uma função kernel  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$  no espaço de características. Qualquer função que satisfaça o teorema de Mercer pode ser usado como uma função do kernel.

O teorema de Mercer (GONÇALVES, 2015; SMOLA; SCHÖLKOPF, 2004) afirma que uma função simétrica  $K(x_i, x_j)$ , ou seja,  $K(x_i, x_j) = K(x_j, x_i)$  com  $x_i, x_j \in R^m$  pode ser escrita com produto interno  $K(x_i, x_j) = \phi_i(x_i)^T \phi(x_j)$  para uma função  $\phi(\cdot) : R^m \rightarrow R^d$  se e somente se  $K(x_i, x_j)$  é semidefinida positiva, isto é,

$$\int K(x_i, x_j) f(x_i) f(x_j) dx_i dx_j \geq 0 \quad \forall f(\cdot)$$

, ou equivalentemente a matriz

$$\begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ K(x_2, x_1) & K(x_2, x_2) & \cdots & K(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_N, x_1) & K(x_N, x_2) & \cdots & K(x_N, x_N) \end{bmatrix}$$

é semidefinida positiva para qualquer coleção de pontos  $x_1, \dots, x_N \in R^m$ . Portanto, uma função de kernel é uma função que é equivalente a um produto interno num espaço de maior de dimensionalidade.

Existem várias formas de kernel aplicadas às máquinas de vetor de suporte, kernel polinomial, kernel de base radial e kernel baseado na função Pearson VII (DMITRIEVA,

2015). Neste trabalho, usamos a função de kernel de base radial (RBF) para realizar a transformação não linear, dada por

$$K(\mathbf{x}, \mathbf{x}_i) = \sum_{i=1}^N \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2)$$

onde  $\gamma = \frac{1}{2\sigma}$  e  $\sigma$  é o comprimento parâmetro de escala, o kernel RBF é o mais utilizado (GU et al., 2011; CHEN et al., 2015).

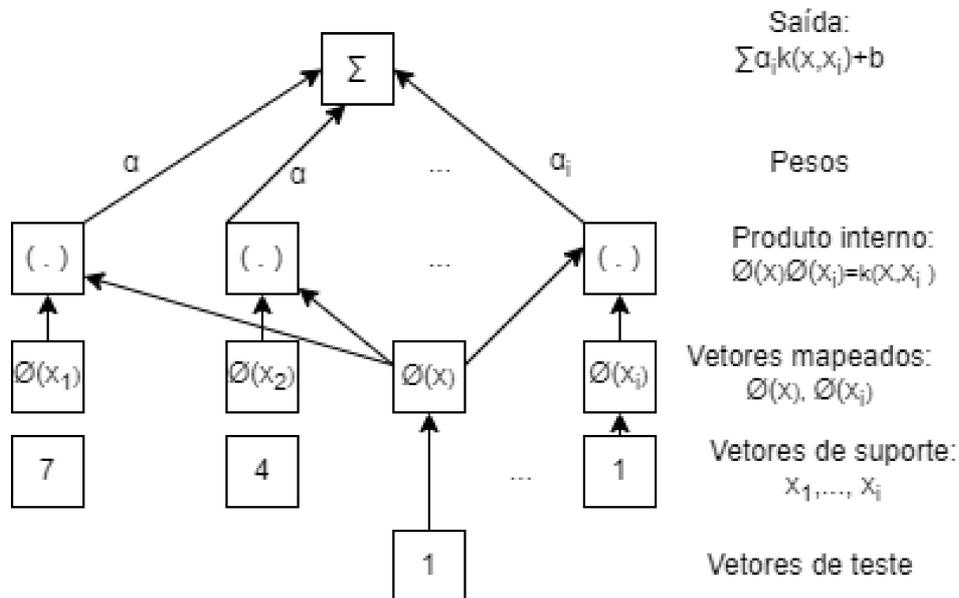
Os parâmetros  $C$ ,  $\varepsilon$  e parâmetros de  $K(\cdot)$  do kernel são definidos pelo usuário.  $C$  controla a suavidade da aproximação função e  $\varepsilon$  determina a margem dentro da qual o erro é tolerado. Finalmente, a função de estimativa SVR pode ser expressa como (WANG et al., 2012):

$$f(x) = \sum_{i=1}^{nsv} (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (3.11)$$

onde  $nsv$  é o número de vetores de suporte.

O SVR pode ser resumido conforme o diagrama abaixo:

Figura 13 - Arquitetura de uma máquina de regressão construída pelo SVR.



Fonte: Extraído de Smola e Schölkopf (2004)

### 3.4.4 EXTREME GRADIENT BOOSTING (XGB)

XGB (CHEN; GUESTRIN, 2016) é um método de *ensembles* que combina vários problemas menores, chamados de preditores fracos, para gerar um problema completo, um preditor forte. No caso do XGB, os preditores fracos são árvores de decisão regularizadas (CHEN; HE, 2015). É conhecido por sua eficiência/flexibilidade, versatilidade, escalabilidade e portabilidade. O XGB também pode automaticamente usar a computação paralela para acelerar a velocidade de processamento, esta característica representa uma grande

vantagem deste sobre outros métodos (ZHAO et al., 2022). Especialmente o XGB tem sido usado no *pipeline* de produção de grandes como Amazon, Uber, Facebook e Google (NGUYEN et al., 2022). Além disso é usado amplamente por cientistas de dados por alcançar resultados excelentes em muitos desafios ou competições de dados e aprendizado de máquina (WANG et al., 2017; ISLAM et al., 2021).

Para um determinado conjunto de dados  $D = (\mathbf{x}_i, \mathbf{y}_i)$  com ( $|D| = n, x_i \in R^m, y_i \in R$ ) A previsão de XGB para uma instância  $i$  é

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F$$

onde  $F = f(x) = w_{q(x)}(q : R^m \rightarrow T, w \in R^T)$  é o espaço de árvores de regressão,  $q$  representa a estrutura de cada árvore que mapeia um dado para o índice de folha correspondente.  $T$  é o número de folhas no árvore. Cada  $f_k$  corresponde a uma estrutura de árvore independente  $q$  e pesos das folhas  $w$ . Ao contrário das árvores de decisão, cada árvore de regressão contém uma pontuação contínua em cada folha, onde  $w_i$  é usado para representar a pontuação na  $i$ -ésima folha. Para um dado usam-se regras de decisão nas árvores (dadas por  $q$ ) para classificá-lo nas folhas e calcular a previsão final somando a pontuação nas folhas correspondentes (dadas por  $w$ ).

Para aprender o conjunto de funções usadas no modelo, minimiza-se a seguinte função de perda regularizada:

$$L(t) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t). \quad (3.12)$$

Na equação 3.12 consideramos  $l$  como o erro quadrático médio onde  $n$  representa o número de amostras e

$$\Omega(f_t) = \alpha_{reg} T + \frac{1}{2} \lambda_{reg} \|\mathbf{w}\|^2$$

é o termo que denota a regularização,  $T$  é o número de folhas,  $\mathbf{w}$  são os pesos das folhas, e  $\alpha_{reg}$  e  $\lambda_{reg}$  são os termos de regularização nos pesos  $L_1$  e  $L_2$ , respectivamente. Isso significa que adicionamos avidamente o  $f_t$  que mais melhora a previsão.

Porém, a expressão 3.12 inclui funções como parametros e não pode ser otimizado usando métodos tradicionais de otimização no espaço euclidiano, aproximação de segunda ordem em séries de Taylor pode ser usada para otimizar rapidamente o objetivo:

$$L(t) = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t)$$

onde  $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$  e  $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$  são estatísticas de gradiente de primeira e segunda ordem na função de perda. Removendo os termos constantes obtem-se a forma simplificada:

$$L(t) = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (3.13)$$

A equação 3.13 acima pode ser reescrita definindo  $I_j = \{i | q(x_i) = j\}$  como o conjunto de instâncias da folha  $j$  e expandindo  $\Omega$ :

$$L(t) = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \alpha T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \alpha T \quad (3.14)$$

Para uma estrutura fixa  $q(x)$ , tem-se o valor ótimo do peso da folha  $j$  dado por:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (3.15)$$

com o seu correspondente valor ótimo igual a:

$$\hat{L}^t(q) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \alpha T \quad (3.16)$$

A Equação 3.16 pode ser usada como uma função de pontuação para medir a qualidade de uma estrutura de árvore  $q$ . Normalmente é impossível enumerar todas as estruturas de árvores  $q$ . Um algoritmo guloso que começa a partir de uma folha única e adiciona ramos iterativamente à árvore é usado. Supondo que  $I_L$  e  $I_R$  sejam os conjuntos de instâncias da esquerda e nós da direita após divisão. Fazendo  $I = I_L \cup I_R$ , então a redução da perda após a divisão é dada por

$$\hat{L}_{split} = \frac{1}{2} \left[ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \alpha \quad (3.17)$$

Esta fórmula é geralmente usada na prática para avaliar os candidatos divididos. Porém, encontrar a melhor divisão conforme indicado na equação 3.17 constitui um dos principais problemas no aprendizado em árvore. Para fazer isso, algoritmos de descoberta de divisão como *Basic Exact Greedy*, *Approximate algorithms*, *Weighted Quantile Sketch* e *Sparsity-aware Split Finding* (CHEN; GUESTIN, 2016) são geralmente utilizados.

Além da regularização, duas técnicas podem ser usadas para evitar *overfitting*, a técnica de encolhimento e a técnica de subamostragem por colunas ou recursos. Semelhante a uma taxa de aprendizado na otimização estocástica, o encolhimento dimensiona os pesos recém-adicionados por um fator  $\eta$  após cada etapa do aumento da árvore reduzindo a influência de cada árvore individual e deixando espaço para futuras árvores para melhorar o modelo. As principais etapas do XGB estão resumidas na tabela 7 (NGUYEN et al., 2022; ISLAM et al., 2021).

Tabela 7 – Treinamento do XGB

Etapas de Treinamento do XGB
Especificar os parâmetros incluindo o número de iterações, $\alpha, \lambda$ e o tamanho de passos $\eta$
Ordenar a entrada definida por valores de recurso
Para cada iteração $t$ :
Encontrar $g_i$ e $h_i$
Examinar a árvore atual e decidir sobre a melhor divisão utilizando a equação 3.17
Os pesos de uma nova árvore $f_t(x)$ é calculado usando a equação 3.16
A nova saída é computada como $y_i^t = y_i^{(t-1)} + \eta f_t(x_i)$
Retornar o modelo XGB treinado

Fonte: Extraído de Nguyen et al. (2022)

Para construir o modelo de aproximação de função baseado em XGB, os seguintes parâmetros devem ser especificados:

1. Eta ( $\eta$ ) taxa de aprendizado usada. Um tamanho grande acelera a fase de aprendizado, mas um valor menor de  $\eta$  torna a fase de aprendizado mais conservadora. O intervalo possível é  $[0, 1]$ , e o padrão é  $\eta = 0.3$ .
2.  $max\_depth$ , profundidade máxima de uma árvore que pode ser permitida por XGB. Aumentar este parâmetro aumentará a capacidade de aprendizagem do modelo, mas é mais provável o *overfitting*. O intervalo possível é  $[0, \infty]$ , e o padrão é  $max\_depth = 6$ .
3. Parâmetro de regularização  $L_2$  lambda ( $\lambda$ ). Valores grandes de  $\lambda$  tornam o modelo mais conservador. O intervalo possível é  $[0, \infty]$ , e o padrão é  $\lambda = 1$ .
4. *Subsample* A razão para a entrada de amostragem. O possível intervalo é  $[1]$  e o padrão é  $Subsample = 1$ .  $Subsample = 0.5$  significa antes do crescimento da árvore, apenas metade ou os dados de treinamento são amostrados.

### 3.4.5 MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)

Desenvolvido por Jerome H. Friedman em 1991 (FRIEDMAN, 1991), MARS é uma abordagem não paramétrica que modela as contribuições de variáveis preditoras para as variáveis resposta usando uma sequência de *splines* de regressão linear por partes para simular a relação não linear entre elas (EVERINGHAM et al., 2011). É uma regressão linear passo a passo que é capaz de receber entradas de dimensões superiores, isto é, os conjuntos de dados de treinamento são divididos em *splines* que são segmentos lineares separados por partes de diferentes gradientes. Nenhuma suposição específica subjacente à relação funcional entre as variáveis de entrada e a saída é necessária (ZHANG; GOH, 2013).

O modelo MARS é representado por funções lineares simples combinadas de forma aditiva e/ou interativa. Sua vantagem está em sua capacidade de detectar e melhorar a

interpretabilidade de interações complexas entre uma variável resposta e um conjunto de variáveis preditoras, maior precisão quando as relações lineares locais estiverem corretas, computacionalmente rápido, podendo funcionar bem em conjuntos de dados grandes e pequenos, fornece seleção automatizada de recursos, pode ser usado para problemas de regressão e classificação, não requer padronização de recursos (ALKHAMMASH et al., 2022; VANEGAS; VÁSQUEZ, 2017).

Pode-se destacar a aplicação com sucesso deste modelo em vários problemas como modelagem da evapotranspiração com dados limitados de temperatura (FERREIRA et al., 2019); agricultura (EYDURAN et al., 2019); análise de sistemas em engenharia geotécnica (ZHANG; GOH, 2013); simulação da vazão (AL-SUDANI et al., 2019b); previsão da demanda do petróleo bruto (ALKHAMMASH et al., 2022); ciência, economia e tecnologia (TAYLAN; WEBER, 2007); indústria de cana-de-açúcar (EVERINGHAM et al., 2011); previsão de séries temporais de mortalidade em menores de 5 anos (VANEGAS; VÁSQUEZ, 2017), entre outras.

O modelo MARS é construído como combinação linear das funções de base (BFs) definidas em pares que formam um nó e subgrupos para definir uma região de inflexão da seguinte forma (CHENG; CAO, 2014):

$$\hat{y}(x) = F_m(x) = c_0 + \sum_{m=1}^M c_m B_m^K(x), \quad B_m^K(x) = \prod_{k=1}^K [\pm(x-t)]_+^q \quad (3.18)$$

Onde  $c_0$  é um valor constante,  $B_m^K(x)$  é a  $m$ -ésima função de base,  $c_m$  é o coeficiente associado desconhecido que pode ser estimado usando o método dos mínimos quadrados à  $m$ -ésima função de base,  $M$  é o número de funções básicas, e  $K$  é o número de graus por partes  $q$  polinômios, que tem a forma:

$$[-(x-t)]_+^q = \begin{cases} (t-x)^q & \text{if } x < t \\ 0 & \text{otherwise} \end{cases}, \quad [(x-t)]_+^q = \begin{cases} (x-t)^q & \text{if } x \geq t \\ 0 & \text{otherwise} \end{cases} \quad (3.19)$$

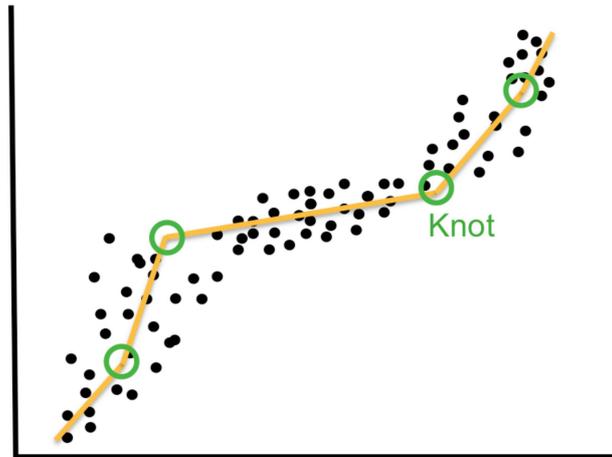
Onde  $t$  são os nós, os pontos de junção dos polinômios. Um nó marca o fim de uma região de dados e o início de outra. As curvas por partes ou simplesmente funções de base, fornecem maior flexibilidade ao modelo.

Os coeficientes MARS na equação (3.18) são escolhidos a fim de minimizar a também chamada Validação Cruzada Generalizada (GCV)

$$\text{GCV} = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i(x))^2}{\left(1 - \frac{\gamma M + (M+1)}{N}\right)^2}, \quad (3.20)$$

onde  $\gamma$  é fator de penalização, e  $M$  é o número de funções base (FRIEDMAN, 2017; AL-SUDANI et al., 2019b).

Figura 14 - Nós e splines lineares para um exemplo de MARS simples



Fonte: Extraído de Raj e Gharineiat (2021)

O modelo MARS caracteriza dados globalmente ou usando regressão linear entre quaisquer dois nós. A sua construção é feita em duas fases, a fase de avanço que adiciona funções e encontra potenciais nós para melhorar o desempenho, resultando em um modelo de *overfit* e, a fase de retrocesso envolve a poda dos termos menos eficazes para descrever um modelo ideal, que consiste na remoção de funções de base uma de cada vez do modelo de *overfit* baseado no critério de soma de quadrados residuais. O modelo é reajustado após cada exclusão e cada modelo reduzido é testado pelo critério GCV (RAJ; GHARINEIAT, 2021).

### 3.5 AJUSTE DE PARÂMETROS DOS MODELOS DE PREVISÃO PROPOSTOS

Os modelos de aprendizado de máquinas embora possuam características matemáticas desejáveis tais como modelo de treinamento convexo, vetores de pesos ótimos esparsos, minimização do risco empírico e termo de capacidade simultaneamente, geralmente solucionados por métodos numéricos baseados em gradientes, Newton, quasi-Newton, Broyden-Fletcher-Goldfarb-Shanno (BFGS) e suas versões, gradiente descendentes, least square (ls)/ sequential least square programming (slsqp) ou ainda máxima verossimilhança aproximada, com *backcasting* ou exato para a estimativa linear ou não linear (SUN et al., 2019) deixam assim algumas decisões que devem ser tomadas pelo projetista de máquinas, por exemplo, parâmetro de regularização, função kernel ou parâmetros do kernel, *epsilon-insensitive* no caso das SVR; número máximo de função de base, parâmetro de penalidade/ suavização, máxima interação entre as variáveis preditoras no caso de MARS (AL-SUDANI et al., 2019b). Esses parâmetros em conjuntos são chamados de hiperparâmetros e a determinação deles Problema de Seleção de Parâmetros ou Problema de Seleção de Modelos. Esses hiperparâmetros têm uma influência essencial nos modelos de aprendizado,

a capacidade de generalização e a complexidade dos modelos estão intimamente ligadas à definição de seus hiperparâmetros. Portanto, a escolha certa desse conjunto de parâmetros determina o compromisso que o modelo possui entre a sua complexidade e a capacidade de generalização ou o balanço entre viés e variância (IZBICKI; SANTOS, 2020).

Os hiperparâmetros podem ser selecionados com base na experiência, na tentativa e erro ou nos parâmetros padrão, no entanto, hiperparâmetros ótimos são dependentes do problema, o que faz com que os usuários não consigam verificar se seus modelos estão estabelecidos com uma configuração ideal. Abordagens mais sistemáticas para selecioná-los incluem *grid search* e *Random search*, onde na primeira os hiperparâmetros são exaustivamente pesquisados a partir de um subconjunto específico do espaço de hiperparâmetros e, na segunda seleciona aleatoriamente combinações de hiperparâmetros para consideração. Ambas as pesquisas *grid search* e *Random search* podem ser feitas em paralelo, no entanto, eles não são confiáveis. Portanto, os pesquisadores têm voltado para metaheurísticas (NGUYEN et al., 2022).

Metaheurísticas e Aprendizado de Máquina são duas áreas de pesquisa em matemática e ciência da computação com um grande número de aplicações, foram amplamente examinados e alcançaram progressos excepcionais individualmente em seus respectivos campos. Técnicas de aprendizado de máquina permitem modelar o conhecimento a partir de dados ou experiência, enquanto esquemas metaheurísticos obtêm a melhor solução ou decisão para um problema complexo. Embora ambas as áreas pertençam a comunidades diferentes, frequentemente são combinadas entre si para eliminar suas desvantagens e melhorar suas capacidades.

Recentemente, essa hibridação de modelos de aprendizado de máquina com metaheurísticas, que consiste em utilizar metaheurísticas para melhorar o desempenho das técnicas de aprendizado de máquina representa um campo de pesquisa emergente com enorme potencial (AKHTER et al., 2019).

O problema de treinar métodos de aprendizado de máquina (ML) é considerado como um problema desafiador com um espaço de busca desconhecido e, as metaheurísticas estão entre as alternativas utilizadas para lidar com a seleção dos melhores parâmetros internos de aprendizado de máquina. O problema é formulado de forma adequada para metaheurísticas, o primeiro passo no treinamento de métodos de aprendizado de máquina usando metaheurísticas é a representação do problema, isto é, encontrar um conjunto de parâmetros internos que forneça a melhor aproximação ou precisão de predição. A função objetivo a ser minimizada é a raiz quadrada do erro quadrático médio (RMSE) calculada por meio de validação cruzada (VAROQUAUX et al., 2017).

No presente estudo metaheurísticas de base populacional, os algoritmos evolutivos ou bioinspirados foram aplicados para a seleção automatizada e inteligente dos hiperparâmetros dos modelos de aprendizado de máquinas em análise. Os algoritmos evolutivos

são técnicas baseadas nos princípios da biologia e do mundo natural (SANTOS, 2019; IBRAHIM et al., 2022). O princípio básico destes algoritmos é, dada uma população de indivíduos (ou seja um conjunto de soluções), pressões do ambiente desencadeiam um processos de seleção natural (ou seja um processo que privilegia as melhores soluções até então encontradas), causando um incremento na adequação das soluções. Estão na classe de algoritmos não determinísticos isto é, possuem uma busca não necessariamente por uma solução ótima, mas sim uma boa solução baseado em ações estocásticas com menor tempo (GABRIEL; DELBEM, 2008). Os algoritmos utilizados foram: genetic algorithms (GA), Differential Evolution (DE), *Grey Wolf Optimization (GWO)* e Particle Swarm Optimization (PSO) porque são algoritmos evolutivos ou bioinspirados clássicos e/ou modernos usados com sucesso no campo da engenharia em geral e, na engenharia de hídricos em particular (MARTINHO et al., 2022; TIKHAMARINE et al., 2020; NGUYEN et al., 2022; NGUYEN et al., 2021) A seguir se apresenta uma breve descrição de cada algoritmo evolutivo utilizado neste trabalho.

### 3.5.1 GENETIC ALGORITHM(GA)

Introduzido por D. Goldberg, J. Holl e K. De Jong em 1989, GA é uma subclasse de algoritmos evolutivos utilizada para fins de busca e otimização com base na genética natural e na seleção natural (RAMSON et al., 2019a). Ao usar um operador estocástico e não determinístico, este algoritmo gera uma solução ótima a cada iteração da população. Este é o motivo de sua maior eficiência do que outros algoritmos de otimização na busca de espaços descontínuos e não lineares. As operações genéticas são cruzamento, reprodução e mutação (AKHTER et al., 2019).

O GA gera aleatoriamente algum conjunto de soluções possíveis para um problema. Cada solução é submetida à função de adequação que se busca otimizar para avaliar cada solução. Novas possíveis soluções serão geradas de forma probabilística a partir das melhores soluções da etapa anterior, sendo que alguns desses são inseridos diretamente na nova população enquanto outros são utilizados como base para gerar novos indivíduos, utilizando operadores genéticos, principalmente o cruzamento e a mutação . O processo continuará até que a solução aceitável seja encontrada (RAMSON et al., 2019a; WHITLEY, 1994)

Os passos do GA podem ser resumidos abaixo:

1. escolha o melhor da população
2. aplicar recombinação e mutação
3. Atribua valor de adequação à nova solução e compare
4. repita até que a solução seja encontrada

## 5. atribuir valor de aptidão a indivíduo

O algoritmo GA pode ser sumarizado pelo pseudocódigo mostrado na Tabela 8 (ROZA et al., 2016):

Tabela 8 – Pseudocódigo do algoritmo GA

---

Pseudocódigo do Algoritmo GA

---

$GA(Fitness, Fitness_{threshold}, p, r, m)$

Fitness: Uma função que retorna um valor que avalie uma hipótese dada

$Fitness_{threshold}$ : Um limiar que especifique um critério de parada

p: O tamanho da população

r: A fração da população a sofrer o crossover a cada iteração

m: A taxa de mutação

- 1 Inicializar população:  $P \leftarrow$  Gerar  $p$  indivíduos aleatoriamente.
- 2 :Avaliar: Para cada  $h$  em  $P$ , calcular  $Fitness(h)$
- 3 **:while**  $\max_h Fitness(h) < Fitness_{threshold}$  **do**:
- 4 : Criar uma nova geração  $P_S$ :
- 5 : Seleção: Selecionar ao azar  $(1 - r)p$  indivíduos de  $P$  para adicionar a  $P_S$ .  
A probabilidade  $Pr(h_i)$  de selecionar o indivíduo  $h_i$  de  $P$  é dada por:  
$$Pr(h_i) = \frac{Fitness(h_i)}{\sum_{j=1}^p Fitness(h_j)}$$
- 6 : Cruzamento: Selecionar ao azar  $\frac{rp}{2}$  pares de indivíduos de  $P$ ,  
de acordo com  $Pr(h_i)$  definido acima.  
Para cada par  $(h_1, h_2)$  produzir dois descendentes aplicando o operador cruzamento.  
Adicionar todos os descendentes a  $P_S$ .
- 7 : Mutação: Escolher  $m\%$  dos indivíduos de  $P_S$  aleatoriamente. Para cada um, aplicar o operador de Mutação.
- 8 :Atualização:  $P \leftarrow P_S$
- 9 :Avaliação: Para cada  $h$  em  $P$ , calcular  $Fitness(h)$

Retornar indivíduo mais apto

---

Fonte: Extraído de Roza et al. (2016)

### 3.5.2 GREY WOLF OPTIMIZATION (GWO)

O algoritmo de Otimização GWO é um algoritmo metaheurístico baseado em enxame recentemente proposto que reproduz a caça e o comportamento social dos lobos cinzentos na natureza. Uma hierarquia com quatro categorias de lobos é utilizada em um processo de caçar, buscar, cercar e atacar a presa (MIRJALILI et al., 2014). O GWO apresentou maior competitividade em termos de resultados obtidos em relação a outras metaheurísticas popularmente conhecidas (NIU et al., 2019). A exploração deste algoritmo é muito intensa que ajuda a população a escapar do mínimo ou máximo local. Além disso, o equilíbrio entre exploração e aproveitamento é direto e eficiente no tratamento de problemas difíceis de acordo com os resultados dos problemas reais, nomeadamente extração de bordas de imagens, segmentação de imagens, imagens médicas, agrupamento de dados, fusão de imagens médicas encontrados em (MIRJALILI, 2015; MIRJALILI et al., 2014; RAMSON et al., 2019b) como também nas áreas de *Feature Selection* (AL-TASHI

et al., 2020), *Global Optimization , Machine Learning*, engenharia e design, processamento de imagens (FARIS et al., 2018), com resultados promissores.

No algoritmo GWO, quatro grupos dividem a população: alfa ( $\alpha$ ), beta ( $\beta$ ), delta ( $\delta$ ) e ômega ( $\omega$ ). Os lobos mais adaptados ao meio ambiente são os primeiros lobos, nomeadamente  $\alpha$ ,  $\beta$  e  $\delta$ , que conduzem outros lobos  $\omega$  na direção de localizações promissoras do espaço de busca. Durante a otimização, os lobos se movem e ajustam suas posições  $\alpha$ ,  $\beta$  ou  $\delta$  da seguinte maneira:

$$\vec{D} = | \vec{C} \vec{X}_p(t) - \vec{X}(t) | \quad (3.21)$$

$$\vec{X}_{(t+1)} = \vec{X}_p(t) - \vec{A} \vec{D} \quad (3.22)$$

onde  $t$  representa a época mais recente,  $\vec{A} = 2a\vec{r}_1$ ,  $\vec{C} = 2\vec{r}_2$ ,  $\vec{X}_p$  é o vetor que representa a posição da presa,  $\vec{X}$  é o vetor posição do Lobo,  $\vec{r}_1$ ,  $\vec{r}_2$  são vetores aleatórios no intervalo  $[0,1]$  e o parâmetro  $a$  é o que decresce linearmente de 2 a zero.

O algoritmo GWO assume que  $\alpha$ ,  $\beta$  e  $\delta$  são supostamente a localização da presa (ótimo). Ao longo da otimização, as três melhores soluções geradas até agora são consideradas  $\alpha$ ,  $\beta$  e  $\delta$  respectivamente. Outros lobos são então indicados como  $\omega$  e com a capacidade de reposicionar em relação a  $\alpha$ ,  $\beta$  e  $\delta$ . O modelo matemático proposto que descreve a posição a ser reajustada de  $\omega$  lobos é o seguinte (MIRJALILI et al., 2014):

$$\vec{D}_\alpha = | \vec{C}_1 \vec{X}_\alpha - \vec{X} | \quad (3.23)$$

$$\vec{D}_\beta = | \vec{C}_2 \vec{X}_\beta - \vec{X} | \quad (3.24)$$

$$\vec{D}_\delta = | \vec{C}_3 \vec{X}_\delta - \vec{X} | \quad (3.25)$$

onde  $\vec{X}_\alpha$ ,  $\vec{X}_\beta$ ,  $\vec{X}_\delta$  mostra a localização de  $\alpha$ ,  $\beta$  and  $\delta$  respectivamente,  $\vec{C}_1$ ,  $\vec{C}_2$ ,  $\vec{C}_3$  são vetores arbitrários e o vetor  $\vec{X}$  denota a posição da solução atual.

As equações 3.23, 3.24 e 3.25 calculam uma aproximação da distância entre a solução atual e  $\alpha$ ,  $\beta$  e  $\delta$ . Seguindo a definição das distâncias, a posição final da solução atual é calculada da seguinte forma:

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \vec{D}_\alpha \quad (3.26)$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \vec{D}_\beta \quad (3.27)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \vec{D}_\delta \quad (3.28)$$

$$\vec{X}_{(t+1)} = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (3.29)$$

onde  $\vec{X}_{alpha}$ ,  $\vec{X}_{beta}$ ,  $\vec{X}_{delta}$  mostra a localização atual dos lobos  $\alpha$ ,  $\beta$  e  $\delta$ ,  $\vec{A}_1$ ,  $\vec{A}_2$ ,  $\vec{A}_3$  são vetores gerados aleatoriamente e  $t$  indica a quantidade de épocas.

Como pode ser observado, as Equações 3.23, 3.24 e 3.25 determinam a magnitude do passo do lobo  $\omega$  em direção a  $\alpha$ ,  $\beta$  e  $\delta$ , respectivamente. As Equações 3.26, 3.27, 3.28 e 3.29 então determinam a posição final dos lobos  $\omega$ . Também existem dois vetores, como pode ser observado:  $\vec{A}$  e  $\vec{C}$ .

As etapas do GWO são resumidas a seguir:

1. Uma população de lobos é inicializada arbitrariamente de acordo com os limites inferior e superior das variáveis.
2. O valor da função objetivo é calculado individualmente considerando as soluções candidatas.
3. Selecione os três melhores lobos e armazene-os como  $\alpha$ ,  $\beta$  e  $\delta$ .
4. Mude a posição do restante da população de lobos ( $\omega$ ) empregando a Eq. 3.26 para 3.29
5. Atualize parâmetros  $a$ ,  $A$  e  $C$
6. Volte para o passo 2 caso o critério de parada não seja atendido
7. Retorne a melhor posição

A população de lobos considerada neste trabalho é cada conjunto formado por hiperparâmetros de cada modelo de aprendizado de máquinas em análise. A inicialização dos mesmos é feita a partir de uma distribuição uniforme. Cada solução candidata é inserida no respectivo modelo de aprendizado de máquinas em estudo para previsão da vazão e registrado o seu valor da função de aptidão. A função objetivo utilizada neste trabalho é o erro quadrático médio. A seleção dos três melhores lobos ou soluções é elitista, isto é, são os lobos que retornam os melhores valores da função objetivo. O Pseudocódigo do algoritmo GWO é (FARIS et al., 2018):

Tabela 9 – Pseudocódigo do algoritmo GWO

---

Pseudocódigo do Algoritmo GWO

---

Inicializa a população de lobos cinzas  $X_i (i = 1, \dots, n)$   
 Inicializa  $a, A$ , e  $C$   
 Calcula a aptidão de cada agente de busca  
 $M =$  Número máximo de iterações  
 $X_\alpha =$  melhor solução  
 $X_\beta =$  a segunda melhor solução  
 $X_\delta =$  a terceira melhor solução  
 $X_\omega =$  qualquer solução  
**while**  $M > t$  **do**  
   **for** cada agente **do**  
     Atualiza a posição do agente atual com equação 3.26 para 3.29  
   **end for**  
   Atualiza  $a, A$ , e  $C$   
   Calcula a aptidão de cada agente de busca  
   Atualiza  $X_\alpha, X_\beta$  e  $X_\delta$   
    $t = t + 1$   
**end while**  
 Retornar  $X_\alpha$

---

Fonte: Extraído de Faris et al. (2018)

### 3.5.3 *DIFERENTIAL EVOLUTION* (DE)

Proposto por Storn e Price na década de 1990, o algoritmo DE é um dos algoritmos robustos bio-inspirados ou inspirados na natureza que otimiza um problema ao tentar iterativamente melhorar uma solução candidata no que diz respeito a uma determinada medida de qualidade (ZAFAR et al., 2017; ZHANG et al., 2017; AL-SUDANI et al., 2019a). Foi inicialmente usado para resolver o problema de ajuste polinomial de Chebychev, mas atualmente é utilizado para encontrar soluções de boa qualidade para problemas de otimização complexos (WANG et al., 2012). Suas principais vantagens em comparação com outros algoritmos evolutivos consistem na fácil implementação pois envolve menos parâmetros, convergência rápida e não há requisitos para qualquer função de otimização entrega a melhor solução (MANDAL et al., 2011).

Este algoritmo adapta os indivíduos por meio da utilização sucessiva dos operadores genéticos mutação, recombinação e seleção que são realizados por várias gerações. As operações mutação e recombinação são responsáveis por modificar as características dos indivíduos, enquanto o operador seleção é responsável por escolher os indivíduos mais adaptados ao meio ambiente (SANTOS, 2019). O DE consiste nas seguintes etapas (WANG et al., 2012):

1. **Inicialização dos parâmetros:** Os principais parâmetros do algoritmo DE são o tamanho da população  $N$ , comprimento do cromossomo  $D$ , o fator de mutação  $F$ , taxa cruzamento  $C$  e o número máximo de gerações  $g$ . o fator de mutação  $F$  é

selecionado em  $[0,2]$ , a taxa de cruzamento  $C$  é selecionada em  $[0,1]$ , quando maior for  $C$  o algoritmo atinge uma convergência mais rápida e prematura.

2. **Inicialização da população:** Definir  $g = 0$ . Gerar uma matriz  $N * D$  com valores aleatórios de distribuição de probabilidade uniforme. O método de geração é:

$$X_{i,j} = rand_{i,j}[0, 1](X_{j,max} - X_{j,min}) + X_{j,min} \quad (3.30)$$

em que,  $i = 1, 2, \dots, N$ ,  $j = 1, 2, \dots, D$ ,  $rand$  é um número aleatório com uma distribuição de probabilidade uniforme, e  $X_{j,min}$ ,  $X_{j,max}$  é o limite superior e limite inferior da coluna  $j$ , respectivamente.

3. **Avaliação da população:** Calcular e registrar os valores de aptidão de todos os indivíduos.
4. **Operação de mutação:** Esta operação usa dois vetores escolhidos aleatoriamente  $X_b, X_c$  para produzir um vetor mutante  $X_a$  da seguinte forma:

$$X'_a = X_a + F(X_b - X_c) \quad (3.31)$$

em que,  $F$  é o fator de mutação no intervalo  $[0, 1]$ ,  $a, b, c$  e  $1, 2, \dots, N$  são escolhidos aleatoriamente e devem manter diferentes uns aos outros.

5. **Operação de cruzamento:** A operação de cruzamento pode aumentar a diversidade da população, e a equação é mostrada a seguir:

$$\begin{cases} X'_b(j) = X'_a(j) & \text{se } C \geq rand(j) \text{ ou } j = randn(j) \\ X'_b(j) = X_a(j) & \text{Outros casos} \end{cases} \quad (3.32)$$

6. **Operação de seleção:** A operação de seleção retém os melhores descendentes na próxima geração. O princípio de seleção são os valores de aptidão, se o valor de aptidão da prole  $f(U_{i,G})$  for melhor que o  $f(X_{i,G})$  do pai, a descendência  $U_{i,G}$  selecionaria, caso contrário, o pai  $X_{i,G}$  reteria.

$$U_{i,G+1} = \begin{cases} U_{i,G} & \text{se } f(X_{i,G}) \geq f(U_{i,G}) \\ X_{i,G} & \text{Outros casos} \end{cases} \quad (3.33)$$

onde  $j$  é a localização do gene de uma população,  $rand(j)$  é um número aleatório e  $randn(j)$  também é um inteiro aleatório no intervalo de  $[1, D]$ ,  $j = randn(j)$  garante que pelo menos um elemento da população possa obter a operação de cruzamento e  $C$  é a taxa de cruzamento.

O algoritmo DE pode resumido pelo pseudocódigo (WANG et al., 2012):

Tabela 10 – Pseudocódigo do algoritmo DE

---

Pseudocódigo do Algoritmo DE

---

**inputs:**  $N, D, F$  e  $C$

**Início:**  
 Inicializa a população  
 Avalia cada indivíduo

**repeat:**  
**for**  $i = 1 : D$  **do**  
 Escolha aleatoriamente  $r_1, r_2, r_3$ .  
 $X'_{r_1, G} = X_{r_1, G} + F(X_{r_2, G} - X_{r_3, G})$   
**for**  $j = 1 : n$  **do**  
**if**  $(C \geq rand(j)) \vee (j = randn(j))$  **then**  
 $u_{j, i, G} = X'_{i, j, G}$   
**else**  
 $u_{j, i, G} = X_{i, j, G}$   
**end if**  
**end for**  
**end for**  
 Avalia cada indivíduo  $u_i$   
 $C = U \cup X$   
 $X = \text{seleção}(C, D)$   
 $G = G + 1$   
**until**  $G \geq N$   
**fim**

---

Fonte: Extraído de Wang et al. (2012)

Os detalhes do algoritmo DE aplicado aos modelos em análise neste trabalho são mostrados a seguir: Na primeira etapa, Os parâmetros são inicializados: número da população ( $N$ ), o fator de mutação ( $F$ ) e a taxa de cruzamento ( $R$ ) do algoritmo e um cromossomo  $x = (p_1, p_2, \dots, p_n)$  que é vetor composto pelos hiperparâmetros para cada modelo em análise. O comprimento do cromossomo é determinado pelo número de hiperparâmetros a ser determinados em cada modelo. Na segunda etapa, a evolução começa, define-se  $g = 0$  e emprega-se a equação 3.30 para gerar a população aleatoriamente. A terceira etapa é a dos cálculos preliminares, insere-se a população gerada em cada modelo para previsão da vazão e, de acordo com os resultados obtidos, calculam-se e registram-se os valores da função de aptidão. A função de aptidão que é empregada é a raiz quadrada do erro quadrático médio que é comumente utilizada. Na quarta etapa, que é a de Geração de filhos, empregam-se as equações 3.31-3.33 para gerar a prole de cada modelo correspondente, então insere-se cada prole em cada modelo de aprendizado correspondente para previsão da vazão e calcula-se o valor da função de aptidão novamente. Define-se  $g = g + 1$ . A quinta etapa é a etapa de repetição até que o critério de parada seja satisfeito, portanto, se  $g$  é igual ao número máximo de gerações, então o melhor cromossomo de solução é obtido, o que significa que os melhores parâmetros de cada um dos modelos correspondentes são obtidos, caso contrário, volta-se a terceira etapa.

### 3.5.4 PARTICLE SWARM OPTIMIZATION (PSO)

A Otimização de Enxame de Partículas (PSO) foi proposta por Kennedy e Eberhart em 1995 (RAMSON et al., 2019a). Inspirado pelos movimentos naturais de enxame biológico baseado em enxame sistemas de otimização consistem em agentes que cooperam uns com outros localmente seguindo regras simples, esses agentes buscam por uma solução ótima entre um conjunto de possíveis soluções em um determinado espaço de busca (BENIAND; WANG, 1993). A Partícula também conhecida como o candidato pode melhorar sua posição considerando a inércia, a influência pessoal e a influência social até obter uma solução melhorada para o candidato.

Devido à sua flexibilidade para aplicações práticas e facilidade de implementar (por ter menos parâmetros), metaheurísticas baseadas em enxame têm sido também utilizadas para otimizar os desempenhos e resultados em diferentes aplicações em engenharia, medicina, militar e economia (MARTENS et al., 2011). O PSO é desenvolvido com base em duas principais características do comportamento de movimento dos bandos de aves: (i) sua posição e (ii) sua velocidade (ADNAN et al., 2021). Considerando o procedimento otimização, a fórmula do PSO na 1ª iteração pode ser escrita da seguinte forma:

$$P_{i+1} = P_i + V_{i+1} \quad (3.34)$$

sendo  $P_i$  e  $V_i$  a posição da partícula e a velocidade da partícula, respectivamente. Na equação 3.34, a velocidade atualizada pode ser alcançada de acordo com o melhor valor pessoal ( $P_b$ ) e à melhor posição do enxame ( $P_g$ ) pela seguinte relação:

$$V_{i+1} = aV_i + c_1r_1(P_i - P_b) + c_2r_2(P_i - P_g) \quad (3.35)$$

As etapas do algoritmo PSO, podem ser resumidas:

1. Inicialização dos parâmetros de PSO.
2. Geração do primeiro enxame.
3. Avaliação da adequação de todas as partículas.
4. registro a melhor aptidão pessoal de todas as partículas.
5. Encontrar a melhor partícula global.
6. Caso o critério de parada não seja atendido, atualizar a velocidade e a posição das partículas e voltar ao passo 3.

O método visa buscar a solução ótima em um espaço de busca por meio da troca de informações entre os indivíduos de uma população. Os indivíduos ou partículas,

representam soluções candidatas para o problema de otimização. O algoritmo trabalha para determinar qual trajetória que cada indivíduo deve seguir no espaço de busca. Durante a busca, as trajetórias são alteradas seguindo o comportamento que reflete o movimento do grupo. Em termos computacionais, cada indivíduo é um vetor (ou uma partícula) em um espaço multidimensional. O algoritmo atualiza a velocidade e a posição de cada partícula (indivíduo) usando um conjunto de equações apresentadas anteriormente. A seguir apresenta-se o pseudocódigo do algoritmo PSO (ROSENDO, 2010; NGUYEN et al., 2022):

Tabela 11 – Pseudocódigo do algoritmo PSO

Pseudocódigo do Algoritmo PSO
Especifica $N_p$ , Max e Atribui parâmetros $a, c_1, c_2$
<b>for</b> $i = 1 : N_p$ <b>do</b>
Inicia $\vec{X}^i$ ( $i = 1, 2, \dots, N_p$ ) com uma solução aleatória para o problema
Inicia $\vec{V}^i$ ( $i = 1, 2, \dots, N_p$ ) com uma velocidade aleatória $< V_{max}$
$\vec{P}_{best}^i \leftarrow \vec{L}_{best}^i \leftarrow \vec{X}_{best}^i$
<b>end for</b>
<b>while</b> Não atingir a condição de parada <b>do</b>
<b>for</b> $i = 1 : N_p$ <b>do</b>
Atualiza $\vec{V}^i$ equação 3.35
$\vec{X}^{i+1} = \vec{X}^i + \vec{V}^i$
$\vec{P}_{best}^i \leftarrow$ melhor entre $\vec{X}^i$ e $\vec{P}_{best}^i$
$\vec{L}_{best}^i \leftarrow$ melhor entre $\vec{P}_{best}^i$ e $\vec{L}_{best}^i$
<b>end for</b>
<b>end while</b>
Retornar $\vec{L}_{best}$

Fonte: Extraído de Rosendo (2010)

## 4 METODOLOGIA DO TRABALHO

A metodologia do trabalho consistiu em aplicar os modelos de aprendizado de máquinas: *Elasticnet* (EN), Máquina de Vetor Suporte (SVR) , *Extreme learning Machine*(ELM), *Extreme gradient boosting* (XGB) e *Multivariate Adaptive Regression Spline* (MARS) à previsão da vazão afluyente em vários horizontes temporais (previsão multipassos) . Para melhorar a capacidade preditiva destes modelos, metaheurísticas bionspiradas de base populacional: Algoritmos Genéticos (GA), Evolução Diferencial (DE), Enxame de partículas (PSO) e *Gray Wolf Optimization*(GWO), foram utilizadas para obtenção dos seus respectivos hiperparâmetros, resultando em modelos otimizados. Uma análise comparativa da performance destes foi aplicada, através de métricas de avaliação de desempenho, escolhendo-se o melhor modelo.

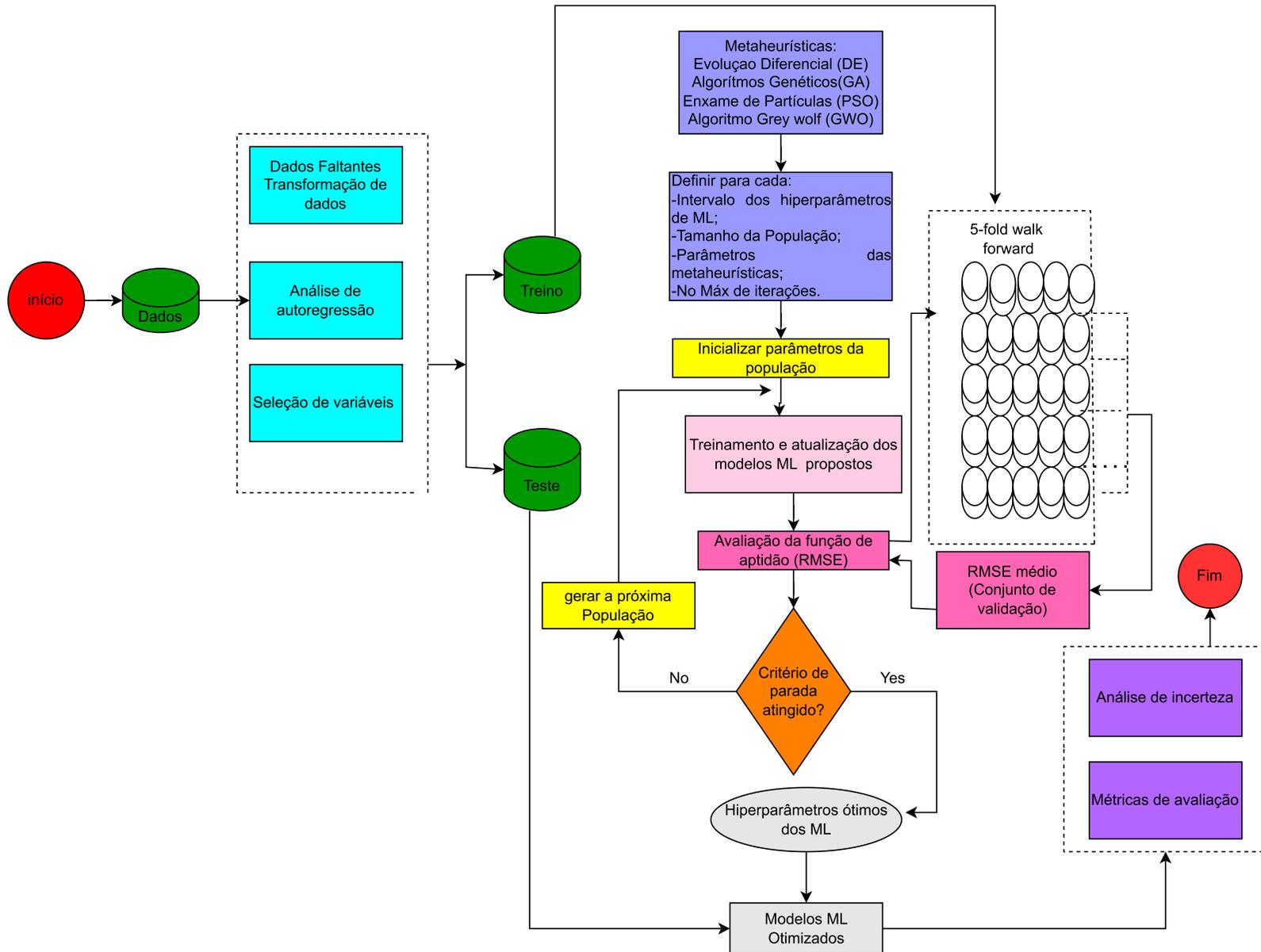
Após a escolha do melhor modelo, métodos lineares e não-lineares de seleção de variáveis de entrada : *Least Absolute Shrinkage and Selection Operator* (LASSO), *Partial Mutual Information variable selection* (PMI), respectivamente, foram utilizados para o desenvolvimento de um modelo simplificado , com a mesma capacidade preditiva que o modelo sem considerar a seleção anteriormente escolhido.

E, finalmente, métodos de análise de incerteza, estatísticos e de (re)amostragem (Método de Monte Carlo), foram usados para quantificar a incerteza dos modelos com a finalidade de avaliar a confiabilidade dos mesmos.

O diagrama de fluxo apresentado na Figura 15 ilustra a metodologia proposta, de maneira geral. A descrição detalhada de cada uma das etapas é feita nas seções seguintes.



Figura 15: Fluxograma mostrando Metodologia Proposta



#### 4.1 ESTRUTURA DE MODELAGEM

O modelo preditivo tem o seguinte formato:

$$Q_{t+j} = F(\underbrace{Q_t, \dots, Q_{t-L}}_{\text{Vazão}}, \underbrace{R_t, \dots, R_{t-L}}_{\text{Precipitação}}, \underbrace{H_t, \dots, H_{t-L}}_{\text{Umidade Relativa}}, \underbrace{E_t, \dots, E_{t-L}}_{\text{Evaporação}}) \quad (4.1)$$

onde  $Q_{t+j}$  é a vazão do rio no dia  $t+j$ ,  $t, \dots, t-L$ ,  $L$  é o número de defasagens consideradas;  $j = 1, 3, 5, 7$  é o número de passos à frente (horizontes de previsão); e  $F$  é uma função que representa a relação de entrada/saída fornecida por um modelo de aprendizado de máquina.

Como alternativa a função  $F$  na Equação 4.1, métodos de aprendizado de máquina foram usados. As unidades de medida das variáveis em análise são: vazão de água ( $m^3/s$ ), evaporação ( $mm$ ), precipitação ( $mm$ ), umidade relativa (%). Em outra análise, esta função  $F$ , também estima as relações ou interações entre a vazão (variável resposta) e as entradas.

#### 4.2 CONFIGURAÇÃO DE ATRASO/ANTECEDENTES PARA PREVISÃO E SELEÇÃO DE VARIÁVEIS

O modelo proposto na Eq. (4.1) requer a determinação do número de defasagens ( $L$ ) para prever as previsões de vazão do rio em diferentes horizontes .

Na teoria, mais variáveis deveriam prover mais poder preditivo, na prática, porém, a excessiva quantidade de variáveis não só torna a aprendizagem do processo mais lenta como também causa excessivo custo de processamento e, muitas vezes confunde a aprendizagem do modelo. A seleção de variáveis, que possibilita a transferência eficaz das informações contidas nos dados para o modelo, tem sido aplicada para esses casos. Esta seleção baseia-se no princípio de que um bom subconjunto de recursos é aquele que contém características altamente correlacionadas com a classe, mas não correlacionadas entre si (FARIS et al., 2018).

O processo de previsão de vazão em análise neste trabalho é baseado na teoria da autoregressão (HADI et al., 2019), ou seja, as previsões são obtidas através de defasagens de uma mesma série ou de variáveis de entrada que contribuem para a variável de saída e, em seguida, na aplicação de métodos de seleção de variáveis (*inputs variables selection*(IVS)). Funções de Autocorrelação (ACF) ou Funções de Autocorrelação Parcial (PACF) para a mesma variável e Função de Correlação Cruzada (CCF), quando outras variáveis são consideradas, são alternativas lineares simples para sugerir o número de antecedentes, ajudando a detectar as variáveis que contribuem para a variável de saída (sem multicolinearidade) (SIQUEIRA et al., 2020; TONGAL; BOOIJ, 2018).

Embora ACF e CCF sejam lineares, esses métodos foram aplicados na escolha das defasagens e na determinação de um conjunto limitado delas. No entanto, um aspecto importante a ser considerado é que essa estratégia pode não ser eficiente, pois o especialista/usuário pode escolher as defasagens de acordo com sua conveniência. Por outro lado, os métodos lineares apresentam limitações, pois detectam apenas a dependência linear do tempo, levando a erros, uma vez que as não-linearidades normalmente presentes nas variáveis hidrológicas não são explicadas. A verificação da existência da dependência temporal que possa incluir não linearidades é relevante no contexto da predição de séries temporais de variáveis hidrológicas, assim modelos mais eficientes de seleção de variáveis de entrada, resumidamente discutidos nas seções 4.2.1 e 4.2.2, foram examinados para fins de uma seleção efetiva de subconjuntos ótimos de variáveis de entrada.

A seleção de variáveis é realizada de forma independente do modelo de aprendizado de máquina preditor, isto é, o modelo de seleção é aplicado ao conjunto de dados original e algumas variáveis são selecionadas para serem eliminadas. A eliminação é feita identificando variáveis com coeficientes nulos ou redundantes, ou ainda pouco informativas. Este procedimento reduz o número de variáveis no conjunto de dados, permitindo a construção de um modelo simples. Em seguida, o modelo de aprendizado preditor é ajustado a esses dados reduzidos, resultando em um modelo com precisão semelhante, mas menos complexo, em consequência diminuição de memória e do esforço computacional requerido.

Formalmente, seja um conjunto de variáveis  $T = \{T_1, T_2, \dots, T_N\}$  uma amostra de variáveis dentro de um determinado contexto, com  $T_i$  sendo uma variável deste conjunto. Seja também um conjunto de classes  $C = \{C_1, C_2, \dots, C_k\}$  de modelos utilizados para a previsão de um determinado processo. Considerando  $G$  um subconjunto de  $T$ , geralmente o objetivo da seleção de variáveis é selecionar um subconjunto mínimo  $G$  tal que  $P(C|G)$  é igual ou tão próxima possível de  $P(C|T)$ , em que  $P(C|G)$  é a distribuição de probabilidade de  $C$  dado o subconjunto  $G$  e  $P(C|T)$  é a distribuição de probabilidade de  $C$  dado  $T$  (KOLLER; SAHAMI, 1996; FREITAS et al., 2007)

#### 4.2.1 LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO)

Séries temporais são sequências temporais de observações  $y_i(t)$ ,  $[i = 1, \dots, n, t = 1, \dots, m]$ , onde  $i$  indexa os dados coletados no momento  $t$ . Se  $n$  for 1 é chamada de série temporal *univariada*; se  $n$  for maior ou igual a 2, é designada como *multivariada*. Mesmo análises de séries temporais univariadas ou multivariadas através de modelos de aprendizado de máquina levam a problemas de multicolinearidade, não-linearidade e não estacionalidade, o que faz com que esses modelos apresentem menor desempenho (AL-JAWARNEH et al., 2021). Para lidar com esse problema, métodos comuns usados para seleção de subconjunto de variáveis devem ser utilizados. O problema de seleção de variáveis de entrada recentemente ganhou muito interesse na modelagem de séries

temporais, especialmente dentro de aplicações de recursos hídricos, demonstrando que algoritmos de seleção de variáveis de entrada fornecem uma melhor representação do processo modelado (QUILTY et al., 2016)

O *Least Absolute Shrinkage and Selection Operator* (LASSO) (TIBSHIRANI, 2011) é um modelo linear que constrói estimativas com coeficientes esparsos. Por construção, LASSO leva a soluções com menos coeficientes diferentes de zero, reduzindo efetivamente o número de variáveis das quais a solução depende. O LASSO reduz o número de variáveis do conjunto de dados original, removendo com parâmetros nulos. Por esta razão, LASSO pode ser visto como um tipo de seleção de recursos, realizado com base em uma análise de regressão (ALIZADEH et al., 2020; DUAN et al., 2016; KWON et al., 2017; CHU et al., 2020; RIZVON; JAYAKUMAR, 2022).

Matematicamente, consiste em um modelo linear com um termo de regularização adicionado. A função objetivo a ser minimizada é:

$$\min_w \frac{1}{2N} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{w}\|_1 \quad (4.2)$$

onde  $N$  é o tamanho da amostra de dados,  $\mathbf{X}$  é a matriz de pontos de dados,  $\mathbf{y}$  são os vetores de saída,  $\alpha \geq 0$ , e  $\|\mathbf{w}\|_1$  é a norma  $L_1$  do vetor de parâmetros.

Durante a minimização da função de custo, o LASSO seleciona automaticamente os variáveis úteis, descartando as redundantes (GUO et al., 2015). Uma variável descartada tem seu coeficiente igual a 0. Além disso, à medida que  $\alpha$  aumenta, mais variáveis são descartadas; portanto  $\alpha$  controla a proporção de variáveis descartadas, sua determinação é geralmente feita por meio de validação cruzada; ou ainda por outras técnicas, como metaheurísticas (AL-JAWARNEH et al., 2021).

Este modelo já mostrou maior capacidade efetiva de seleção de variáveis do que modelos baseados em redes neurais com parâmetros otimizados (PSO-ANN) (ALIZADEH et al., 2020) e métodos de seleção de subconjuntos como: *Best-Subset Selection*, *Forward-and Backward-Stepwise Selection*, and *Forward-Stage-wise Regression*, por estes exibirem muitas vezes alta variância e, conseqüentemente, não reduzirem o erro de previsão do modelo completo (HASTIE et al., 2009a).

#### 4.2.2 PARTIAL MUTUAL INFORMATION VARIABLE SELECTION (PMI)

A teoria da informação <sup>1</sup> se baseia na teoria de probabilidades e estatística, preocupando-se com medidas de informação das distribuições de probabilidade associadas com variáveis aleatórias. As Grandezas importantes da teoria da informação são a

<sup>1</sup> A teoria da informação é um ramo da teoria da probabilidade que estuda a informação e tudo relacionado a ela: canais, compressão de dados, criptografia, entre outros. O modelo da teoria da informação, trata de determinar a forma mais econômica, rápida e segura de codificar uma mensagem, sem a presença de qualquer ruído.

*entropia*, uma medida de informação de uma única variável aleatória, e PMI, uma medida de informação em comum entre das variáveis aleatórias.

A entropia é uma medida de *incerteza*, ligada à previsibilidade do valor que será assumido por uma dada variável aleatória, ou seja, dada uma variável aleatória  $X = \{x_1, x_2, \dots, x_n\}$  associada a distribuição de probabilidade  $p(X) = \{p_1, p_2, \dots, p_n\}$  defini-se entropia (ou informação de shannon <sup>2</sup> média) como:

$$H(X) = - \sum_{i=1}^n p_i \log_2 p_i \quad (4.3)$$

A entropia para uma distribuição conjunta  $P(X, Y)$  pode ser obtida de forma direta da definição de entropia, analogamente dada por:

$$H(X) = - \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} p_{ij} \log_2 p_{ij} \quad (4.4)$$

onde  $p_{ij}$  é a probabilidade de ocorrência do par  $C_{ij} = (x_i, y_j)$ . Essa definição extremamente importante ao definir o PMI.

As limitações da análise de correlação linear criaram interesse em análises estatísticas alternativas, e em medidas de dependência que sejam mais adequadas para identificar e quantificar a dependência que pode ser caótica ou não-linear (MAY et al., 2011).

Derivado da teoria da informação, *Partial Mutual Information variable selection (PMI)* é um critério multivariado não-linear para medir a dependência entre as variáveis (HU et al., 2015; REN et al., 2020; YANG et al., 2016). É um método robusto para detectar relacionamentos entre conjuntos de dados (KRASKOV et al., 2011; ROSS, 2014; LI et al., 2015; LAARNE et al., 2021).

Uma medida para a dependência parcial entre um novo candidato de entrada  $U$  e saída  $V$  dadas as variáveis de entrada  $W$  é dada por:

$$PMI(U, V|W) = \iint f_{U',V'}(u', v') \ln \left[ \frac{f_{U',V'}(u', v')}{f_{U'}(u') f_{V'}(v')} \right] du' dv' \quad (4.5)$$

com

$$u' = u - E(u|W), v' = v - E(v|W) \quad (4.6)$$

onde  $u'_i$  e  $v'_i$  são  $i$  residuais no conjunto de dados com  $n$  amostras,  $f_{U'}(u')$  e  $f_{V'}(v')$  são as funções de densidade de probabilidade marginal, e  $f_{U',V'}(u', v')$  é a densidade de probabilidade conjunta. O PMI é simétrico e não negativo. Se  $PMI = 0$ ,  $X$  e  $Y$  são variáveis independentes.

<sup>2</sup> Claude Elwood Shannon (1916-2001) foi um matemático, engenheiro eletrônico e criptógrafo estadunidense, conhecido como o pai da teoria da informação.

As verdadeiras formas funcionais das funções densidade de probabilidade referenciadas na equação 4.5 são tipicamente desconhecidas. Portanto, se as amostras estiverem disponíveis, a pontuação do PMI pode ser dada por:

$$PMI(U, V|W) = \frac{1}{n} \sum_{i=1}^n \ln \left[ \frac{f_{U',V'}(u'_i, v'_i)}{f_{U'}(u'_i) f_{V'}(v'_i)} \right] \quad (4.7)$$

O PMI também pode ser expresso em termos de entropias como (LIANG et al., 2019):

$$PMI(u, v) = H(u) - H(u|v) \quad (4.8)$$

com

$$H(u) = - \int P_U(u) \log(P_U(u)) du$$

onde  $H(u|v)$  é a entropia condicional, ou seja, a incerteza média do valor de  $u$  após  $v$  ser observado.

Se  $v = u + \eta$ , pode se demonstrar a partir da equação 4.8 que a entropia do ruído é igual a entropia condicional, ou seja,  $H(v|u) = H(\eta)$ , onde  $\eta$  é o ruído:

$$PMI(u, v) = H(v) - H(v|u)$$

$$PMI(u, v) = H(v) - H([u + \eta]|u)$$

Se o valor de  $u$  é conhecido, a incerteza em  $u$  é nula, portanto, não tem nenhuma contribuição na entropia condicional  $H([u + \eta]|u)$ ,

$$PMI(u, v) = H(v) - H(\eta|u)$$

Pela propriedade de independência de  $\eta$  e  $u$ ,  $H(\eta|u) = H(\eta)$ . E por comparação com a equação 4.8, conclui-se que  $H(v|u) = H(\eta)$ . logo, o PMI pode ser obtido considerando-se a entropia da saída em relação ao ruído do canal.

O problema de seleção de variáveis é formulado da seguinte forma: Dado um conjunto inicial  $F$  com  $n$  variáveis, encontra-se um subconjunto  $S \in F$  com  $k$  variáveis que maximiza o  $PMI(C; S)$  entre a variável de classe  $C$  e o subconjunto de selecionado  $S$  (ESTÉVEZ et al., 2009).

É um método vantajoso em relação a outros métodos, pois permite que a correlação linear e a não linear ou de qualquer outro tipo, seja capturada, visto que, é construído a partir de funções densidades de probabilidades (PDFs) conjuntas e marginais das variáveis, e não utiliza estatísticas de qualquer grau ou ordem. É invariante sob transformações do espaço, porque o argumento do logaritmo é adimensional, portanto o valor integral não depende das coordenadas escolhidas (transformação no espaço de características) (KULLBACK, 1997; FRASER; SWINNEY, 1986; ESTÉVEZ et al., 2009).

O PMI é modelo neutro, o que significa que a solução pode ser aplicada a vários tipos de modelos de ML. A solução PMI é rápida (BISHOP; NASRABADI, 2006).

Este método pode ser entendido também como um método de redução de incerteza, pois assumindo que haja alguma incerteza em torno de uma observação  $y$  de uma variável de saída aleatória  $Y$ , dada uma variável de entrada aleatória  $X$  da qual  $Y$  seja dependente, então a observação mútua de  $(x, y)$  reduz essa incerteza, uma vez que o conhecimento de  $x$  permite inferir o valor de  $y$  e vice-versa (MAY et al., 2008).

A aplicação deste método no domínio da hidrologia, com particular destaque em problemas de escoamento (LI et al., 2021), vazão (CHU et al., 2021), evapotranspiração potencial e transporte de sedimentos (DELFORGE et al., 2022) e recursos hídricos (QUILTY et al., 2016) é relativamente recente estando ainda em desenvolvimento.

### 4.3 PREVISÃO MULTIPASSOS À FRENTE DE VAZÃO

A previsão multipassos à frente dos valores futuros de uma série temporal é de extrema importância, visto que permite o planejamento antecipado e correto dos sistemas em estudo (COSTA, 2015). A principal tarefa da previsão multi-passos em séries temporais com  $T$  observações consiste em estimar as observações futuras de  $\{y_1, \dots, y_{T+H}\}$ , em que  $h = \{1, \dots, H\}$  representa o horizonte da previsão, em curto, médio ou longo prazo (CHENG et al., 2020).

Existem duas principais abordagens à previsão multi-passos: a *recursiva* e a *direta*. Na abordagem recursiva ou iterativa, estima-se um único modelo para a série temporal, que produz previsões um passo à frente recursivamente. Existe uma maior sensibilidade a acumulação de erros de estimativa, uma vez que os valores estimados são cada vez mais usados como entrada quando se avança no tempo. Na abordagem direta ou independente, especifica-se e estima-se um modelo diferente para cada horizonte de previsão. Não há então, acumulação de erros, pois não são utilizados valores aproximados, apesar de haver maior complexidade funcional a fim de modelar a dependência estocástica (AN; ANH, 2015).

A estratégia direta desenvolve  $N$  modelos de previsão separados para prever  $N$  passos (MASUM et al., 2018). Para prever por exemplo dois pontos de qualquer cenário usando uma estratégia direta, o primeiro ponto de previsão  $F_1(t+1)$  precisa ser calculado através de um modelo e, em seguida, um modelo diferente seria usado para prever o segundo  $F_2(t+2)$ . O segundo ponto não depende da estimativa do primeiro.

Perante um problema de previsão multipassos, deve-se escolher a melhor estratégia de previsão a utilizar. No presente estudo aplicou-se a abordagem direta de previsão em curto prazo porque, para além desta não apresentar maior sensibilidade a acumulação de erros, não é sensível ao horizonte de previsão, como acontece na estratégia recursiva. Isto permite analisar a robustez absoluta de cada modelo, ou seja, a performance dos modelos propostos independente do passo considerado. Por outro lado, muitas vezes é difícil distinguir os modelos entre si utilizando a previsão um passo à frente, pois estes

tendem a obter melhores resultados. Este facto pode influenciar as previsões nos passos seguintes, fazendo com que seus resultados sejam pouco confiáveis quando uma previsão recursiva é considerada, dado que previsões mais recentes dependem das imediatamente anteriores. Na estratégia recursiva apenas um modelo é estimado; isso faz com que este modelo inalterado reduza a flexibilidade de previsão.

Uma representação da abordagem direta para o problema em estudo, é dada pelas equações:

$$Q_{t+1} = F_1(Q_t, \dots, Q_{t-L}, R_t, \dots, R_{t-L}, H_t, \dots, H_{t-L}, E_t, \dots, E_{t-L})$$

$$Q_{t+3} = F_2(Q_t, \dots, Q_{t-L}, R_t, \dots, R_{t-L}, H_t, \dots, H_{t-L}, E_t, \dots, E_{t-L})$$

$$Q_{t+5} = F_3(Q_t, \dots, Q_{t-L}, R_t, \dots, R_{t-L}, H_t, \dots, H_{t-L}, E_t, \dots, E_{t-L})$$

$$Q_{t+7} = F_4(Q_t, \dots, Q_{t-L}, R_t, \dots, R_{t-L}, H_t, \dots, H_{t-L}, E_t, \dots, E_{t-L})$$

#### 4.4 ESTIMAÇÃO DO DESEMPENHO DOS MODELOS

A avaliação do desempenho dos modelos é feita para verificar se o modelo comporta-se de maneira suficientemente satisfatória com relação aos dados de estudo. As métricas utilizadas hoje para avaliar a qualidade das previsões de vazão são baseadas no erro entre a série simulada e a observada. Independentemente da medida utilizada, sempre se comparam esses valores ponderando esses erros de forma particular. Normalmente, os melhores modelos são considerados para menores valores de erro (COSTA, 2014).

Existem diversas medidas de desempenho utilizadas para quantificar a qualidade para predição de séries temporais (isto é, a qualidade do ajuste do modelo obtido a partir dos dados reais). Neste trabalho, serão utilizados alguns retirados de (COSTA, 2014; GUILHON et al., 2007a). Sabendo que  $O_i$  representa os dados observados e  $P_i$  os valores previstos,  $\bar{O}$  é a média dos valores observados:

- a) **erro percentual absoluto Médio (MAPE)**– Traduz o quanto as previsões se desviam em valor percentual relativo, em relação aos valores observados. Este valor em geral será muito elevado para locais com vazões pequenas, pois qualquer desvio poderá significar muito, em termos percentuais, e deverá ser menor para locais com vazões maiores, pois mesmo grandes variações, podem significar pouco em relação ao total a ser previsto. O MAPE é calculado de acordo com a seguinte equação:

$$MAPE = 100 \times \frac{1}{N} \sum_{i=1}^N \frac{|O_{(i)} - P_{(i)}|}{|O_{(i)}|} \quad (4.9)$$

- b) **eficiência de Nash-Sutcliffe (NSE)** - O NSE corresponde a uma estatística normalizada da variância do resíduo, comparada com a variância dos dados observados.  $NSE = 1$  corresponde a uma aproximação perfeita dos dados

observados,  $1 > NSE > 0$  indica que a aproximação do modelo é melhor do que a média dos dados observados e  $NSE < 0$  indica que a média dos dados observados é um preditor melhor do que o modelo ajustado. A seguir é mostrada a expressão para obter  $NSE$ :

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4.10)$$

- c) **raiz do erro quadrático médio (RMSE)** corresponde a raiz quadrada da média do quadrado da diferença entre valores observados e os valores previstos pelo modelo. O valor do RMSE possui a mesma unidade da variável. No caso, da vazão, seu valor será em  $m^3/s$ , e deve ser o menor possível. Esta métrica é útil pois reúne todos os desvios entre valores previstos e o observados ao longo dos pontos dos dados. A expressão utilizada para calcular é:

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (O_i - P_i)^2} \quad (4.11)$$

- d) **eficiência de Kling-Gupta (KGE)**-é uma medida de adequação que foi desenvolvida por Gupta et al. (2009) para fornecer uma decomposição diagnóstica interessante da eficiência de Nash-Sutcliffe (e, portanto, NSE), o que facilita a análise da importância relativa de seus diferentes componentes (correlação, viés e variabilidade). No contexto da modelagem hidrológica, se propôs uma versão revisada deste índice, para garantir que as razões de viés e variabilidade não sejam correlacionadas. O KGE é obtido com a seguinte expressão:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (4.12)$$

$r$  é o coeficiente de correlação momento-produto de Pearson e  $\alpha$  é a razão entre o desvio padrão dos valores previstos e o desvio padrão dos valores observados. Finalmente,  $\beta$  é a razão entre a média dos valores previstos e a média dos valores observados.

- e) **erro médio absoluto (MAE)**- é a média dos erros das diferenças absolutas entre o valor previsto e o valor real, em que todas as diferenças individuais têm peso igual. O Erro Médio Absoluto mede a magnitude média dos erros em um conjunto de previsões, sem considerar sua direção. O erro médio absoluto também expressa o erro médio de previsão do modelo em unidades da variável de interesse. A expressão para determinar o MAE é descrita a seguir:

$$MAE = \frac{1}{N} \sum_{t=1}^N |(O_i - P_i)| \quad (4.13)$$

- f) **índice de Willmot (WI)**- de acordo com Haveroth et al. (2012), o Índice de Concordância de Willmott mede a dispersão dos dados em relação à reta 1 : 1, isto é, mede a concordância dos valores simulados em relação aos medidos ou observados. Este índice varia de 0 a 1, sendo que quanto mais próximo de 1, menor a amplitude dos erros. O Índice de Concordância de Willmott, é descrito pela equação:

$$WI = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (4.14)$$

- g) **coeficiente de determinação**, também chamado de  $R^2$  é uma medida de ajuste de um modelo aos valores observados de uma variável aleatória. O  $R^2$  varia entre 0 e 1, por vezes sendo expresso em termos percentuais. Nesse caso, expressa a quantidade da variância dos dados que é explicada pelo modelo linear. Assim, quanto maior o  $R^2$ , mais explicativo ou seja, melhor ele se ajusta à amostra. o  $R^2$  é dado por:

$$R^2 = \frac{\sum_{i=1}^N (\bar{O} - P_i)^2}{\sum_{i=1}^N (P_i - \bar{O})^2 + \sum_{i=1}^N (O_i - P_i)^2} \quad (4.15)$$

- h) **coeficiente de correlação linear** que mede uma dependência linear entre valores observados e previstos pelo modelo. O coeficiente de correlação linear é dado por:

$$r = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2 \sum_{i=1}^N (O_i - \bar{O})^2}} \quad (4.16)$$

#### 4.4.1 TESTE DE HIPÓTESES ESTATÍSTICAS

Os testes de hipóteses nos permitem rejeitar, ou não rejeitar, uma hipótese estatística, através da evidencia fornecida pela amostra (OLIVEIRA, 2008). Uma vez formuladas as hipóteses, testam-se para que uma decisão seja tomada, seja em favor da hipótese nula ou da hipótese alternativa.

Os testes utilizados foram o teste ANOVA e o teste de comparações múltiplas de Médias- Teste de Tukey <sup>3</sup>. A Análise de Variância ou ANOVA é uma técnica utilizada em estudos em que busca-se comparar a distribuição de três ou mais grupos de amostras independentes (GOMES, 1954). O resultado desta evidencia que a distribuição de pelo menos um dos grupos se difere das demais, mas não indica entre quais grupos essa diferença é significativa. Assim, é utilizado o teste de comparações múltiplas (Teste de Tukey). O

<sup>3</sup> O teste de Tukey foi desenvolvido por John Wilder Tukey (1915- 2000) e apresentado em 1949 no artigo titulado “Comparing Individual Means in the Analysis of Variance” (Biometrics. 5 (2): 99–114. JSTOR 3001913).

Teste de Tukey consiste em comparar todos os possíveis pares de médias e se baseia na diferença mínima significativa, considerando os percentis do grupo (TUKEY, 1949) .

Neste estudo, essas ferramentas foram utilizadas também para avaliação de desempenho dos modelos, onde foi comparada a distribuição das medidas de desempenho (cada conjunto com 30 execuções independentes) através do teste ANOVA de um fator. E, em seguida, foi utilizado um teste de tukey para comparações múltiplas das médias dos desempenhos dos estimadores (modelos), para verificar qual é o estimador superior, entre os analisados em termos de performance, e outro para comparações múltiplas das médias de desempenho de metaheurísticas.

#### 4.5 WALK FORWARD VALIDATION

Uma técnica utilizada para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados chama-se *Validação Cruzada*. É uma técnica amplamente empregada em problemas onde o objetivo da modelagem é a previsão, buscando estimar quão preciso é este modelo na prática; ou seja, o seu desempenho perante um novo conjunto de dados (IZBICKI; SANTOS, 2020).

Os dados de séries temporais são caracterizados pela correlação entre observações próximas no tempo (autocorrelação); portanto, a técnica de validação cruzada utilizada para a avaliação do desempenho dos modelos estudados é a *Walk Forward Validation* ou *Rolagem* ou mesmo *janelas deslizantes*. A *Walk Forward Validation*, é uma variação de *k-fold*, que retorna os primeiros  $k$  lotes como conjunto de treinamento e o lote  $k + 1$  como conjunto de teste; isto é, os conjuntos de dados de teste são os mais recentes, ou posteriores aos conjuntos de dados de treinamento.

Na prática, a ideia é separar as séries temporais em intervalos de tempo fixos expandindo o conjunto de treino em cada iteração; ou seja, utiliza-se o primeiro lote de dados para treinar e validar o modelo com um conjunto de hiperparâmetros, para testá-lo depois com o segundo. Ou ainda, treina-se e valida-se o modelo com os dois primeiros lotes e avalia-se com a terceira parte dos dados, assim sucessivamente, em  $k - 1$  vezes de validação cruzada.

O processo de *Walk Forward Validation* prevê que sejam definidas amostras de teste sequenciais ao longo da série que podem ser feitas pelo método *Sliding Window* onde as amostras de treino e teste são fixadas e deslocadas a um passo definido de observações ou *Expanding Window* em que é fixado o início do conjunto de treino e o fim aumenta (PARMEZAN et al., 2019).

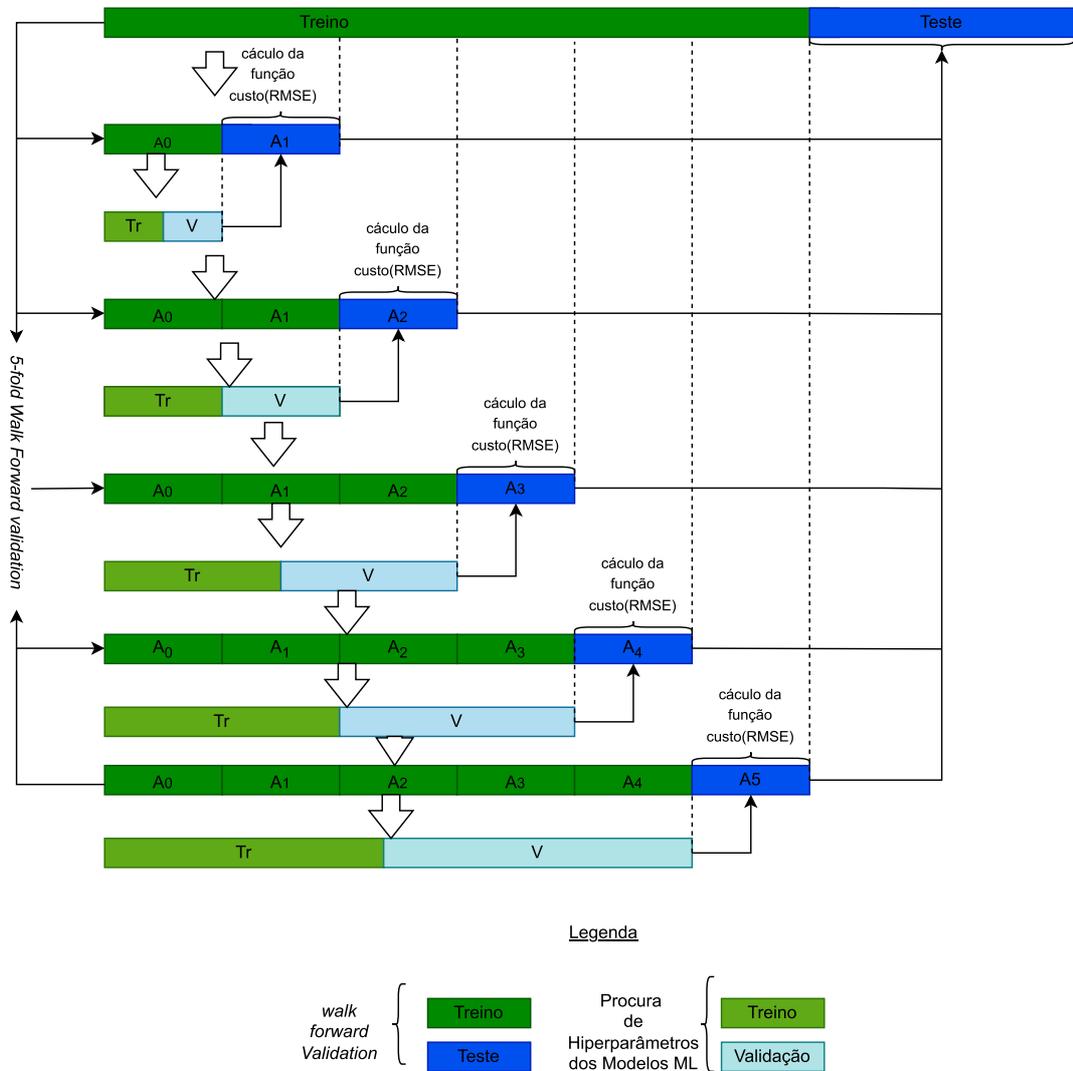
Devido a possibilidade de aumento da amostra de treino, uma validação *walk forward validation* com técnica *Expanding window 5-folds* ( $k = 5$  fold) (valor demonstrado empiricamente que produz estimativa de taxa de erro de teste não sujeitos ao viés e variância

muito altos (BERGMEIR; BENÍTEZ, 2012)) (Figura 16), esta técnica foi utilizada para avaliar cada solução candidata. O conjunto de treino é subdividido em 6 subconjuntos  $A_0, A_1, A_2, A_3, A_4$  e  $A_5$ , que são usados como segue:

- a) fold 1 : treinando  $A_0$ , teste  $A_1$
- b) fold 2 : treinando  $A_0 \cup A_1$ , teste  $A_2$
- c) fold 3 : treinando  $A_0 \cup A_1 \cup A_2$ , teste  $A_3$
- d) fold 4 : treinando  $A_0 \cup A_1 \cup A_2 \cup A_3$ , teste  $A_4$
- e) fold 5 : treinando  $A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4$ , teste  $A_5$

O tamanho desses conjuntos é calculado de modo que  $\#A_0 = \#A_1 = \#A_2 = \#A_3 = \#A_4 = \#A_5 = \frac{N}{k+1}$ , onde  $N$  é o tamanho total de dados disponíveis para validação cruzada e  $k$  é o número total de simulações realizadas.

Figura 16: Esquema de validação *5-Fold walk forward validation*.



Fonte: Extra\u00eddo de Bergmeir e Ben\u00edtez (2012)

## 4.6 ANÁLISE DE INCERTEZA

O fluxo de um rio é uma das principais variáveis do ciclo hidrológico. Sua análise quantitativa e qualitativa é complexa devido a muitos fatores que influenciam as incertezas na previsão de seus valores. A quantificação dessas incertezas é essencial para o estudo da capacidade preditiva dos modelos propostos, pois uma previsão precisa traz diversos benefícios na operação dos reservatórios (BRAVO et al., 2009).

Portanto, além da validação *5-folds walk forward validation* anteriormente apresentada, métodos de quantificação do erro de previsão também foram utilizados. Aplicados em cenários de alto risco onde para além da previsão existe maior interesse em analisar o grau de certeza do modelo para tomada de decisão, esses métodos consistem na especificação da distribuição preditiva do erro  $e$ , como consequência, na obtenção de estatísticas de incerteza e a construção de seus intervalos de confiança. Geralmente são utilizados métodos analíticos (integração numérica) ou estatísticos, onde suposição de normalidade é uma regra e/ou métodos de (re)amostragem que consistem em sortear com reposição dados pertencentes a uma amostra, onde será encontrada a distribuição empírica dos dados (FERREIRA, 2013).

Os métodos estatísticos ou paramétricos foram utilizados para análise do erro de previsão através da construção dos intervalos de confiança. Devido ao grande erro relativo, que pode ultrapassar 100% na previsão em aplicações hidrológicas, usou-se o erro logarítmico sugerido por Sattar e Gharabaghi (2015) para qualquer análise estatística das previsões. A análise de incerteza define o erro de previsão individual como:

$$e_j = \log_{10}(Q_{tp_j}) - \log_{10}(Q_{tm_j}), \quad j = 1, \dots, N \quad (4.17)$$

Os erros de previsão calculados para todo o conjunto de dados são usados para calcular a média e o desvio padrão dos erros de previsão como

$$\bar{e} = \sum_{j=1}^N e_j \quad \text{and} \quad S_e = \sqrt{\sum_{j=1}^N \frac{(e_j - \bar{e})^2}{(N - 1)}} \quad (4.18)$$

, respectivamente. Um valor médio negativo indica que o modelo de previsão subestimou os valores observados, e um valor positivo indica que modelo superestimou os valores observados.

Dado um conjunto de  $N$  amostras observadas no conjunto de dados de teste, o intervalo de confiança de 95% em torno do valor previsto  $Q_{tm_j}$  é construído por (SATTAR; GHARABAGHI, 2015)

$$(Q_{tm_j} \times 10^{-\bar{e}-1.96S_e}, Q_{tm_j} \times 10^{-\bar{e}+1.96S_e}) \quad (4.19)$$

onde  $\bar{e}$  e  $S_e$  são a média e o desvio padrão dos logarítmos dos erros de previsão para cada modelo proposto, respectivamente, e  $Q_{tm_j}$  é a vazão real e  $Q_{tp_j}$  é a vazão estimada a  $i$ th amostra.

Um método de (re)amostragem designado por "método de simulação de Monte Carlo"(MCS) (SHIELDS, 2022) também foi utilizado.

Apropriado para resolver problemas de grande escala, este método não-paramétrico é muitas vezes usado para calcular o valor esperado de uma variável que é função de várias variáveis estocásticas, e não pode ser tratado analiticamente (HARRISON, 2010). É uma alternativa simples e de fácil implementação, e não assume que a distribuição de amostragem seja normal. A técnica permite a construção da distribuição empírica dos quantis a partir de amostras das séries obtidas por simulação (PHAM et al., 2020).

O resultado determinístico para cada simulação de Monte Carlo (MCS) é calculado para cada modelo desenvolvido neste trabalho. Um total de 25.000 resultados foram determinados para  $Q_t$  aleatoriamente, usando de uma distribuição uniforme. O desvio médio absoluto (MAD) em torno da distribuição de saída da mediana é escrito da seguinte forma:

$$\text{MAD} = \frac{1}{25000} \sum_{i=1}^{25000} |Q_{tp_i} - \text{median}(Q_{tp})| \quad (4.20)$$

enquanto a incerteza da saída de cada modelo pode ser especificada por

$$\text{Uncertainty \%} = \frac{100 \times \text{MAD}}{\text{median}(Q_{tp})} \quad (4.21)$$

onde  $Q_{tp_i}$  é a vazão afluyente prevista para a  $i$ th amostra. Uma porcentagem alta de incerteza indica menos confiabilidade ao modelo.

O MAD constitui um estimador robusto (mais resistente a *outliers*), que traduz o quanto as previsões de vazões se desviam, por unidades de medida, em relação aos valores observados; ou seja, em relação ao seu objetivo, que é acertar a previsão. A tendência é que esse valor seja mais elevado para locais com maiores vazões, e seja menor para locais com valores menores de vazões. Quanto menor é o seu valor, melhores são as previsões (GUILHON et al., 2007a).

## 5 EXPERIMENTOS COMPUTACIONAIS

### 5.1 DESCRIÇÃO DOS EXPERIMENTOS COMPUTACIONAIS

Para atender aos objetivos desta pesquisa, um conjunto de dados reais constituído por quatro séries temporais: vazão do rio afluente ao reservatório de geração de energia elétrica ( $Q$ ), precipitação ( $R$ ), evaporação ( $E$ ) e umidade relativa ( $H$ ), foram utilizadas na análise. A vazão afluente é considerada como variável de saída, e as restantes como variáveis de entrada ou preditoras.

A tarefa de previsão da vazão afluente vários passos à frente em análise neste trabalho foi realizada construindo inicialmente uma estrutura própria constituída pelas variáveis de entrada e seus defasamentos para receber os modelos de aprendizado de máquinas propostos.

A determinação do número de defasamentos ( $L$ ) a serem consideradas na equação 4.1 para fazer previsões de vazão afluente do rio  $Q_{t+j}$  no horizonte temporal ( $j = 1, 3, 5, 7$ ) foi feita utilizando as Funções de Autocorrelação (ACF) ou as Funções de Autocorrelação Parcial (PACF) para uma mesma variável, e a Função de Correlação Cruzada (CCF) quando outras variáveis são consideradas. Esses métodos são alternativas simples para sugerir o número de defasamentos, ajudando a detectar as variáveis que contribuem para a variável de saída. A autoregressão ou estudo da relação entre as defasamentos usando ACF/PACF e CCF para as variáveis analisadas são ilustradas nas Figuras 17 e 18, respectivamente. A Figura 17 indica que os primeiros defasamentos podem ser considerados preditores para ACF/PACF e, a Figura 18 mostra que nenhum (ou todos) os defasamentos podem ser potencialmente selecionados como critério de seleção para CCF. Por outro lado, pode-se observar ainda na Figura 18 um padrão cíclico (sazonalidade) identificado através da recessão da correlação em certos intervalos de tempo (dias).

Neste estudo, um total de sete defasamentos, o que gera a vinte e oito (28) variáveis de entrada (4 variáveis originais x 7 defasamentos). Portanto, uma matriz  $5844 \times 28$  foi considerada como conjunto de dados de entrada para os modelos de aprendizado de máquina, utilizando ACF/PACF como critério de seleção. O CCF não foi informativo, dado que, como pode ser observado na Figura 18, a maioria dos valores de CCF estão dentro do intervalo de confiança definido em linhas tracejadas.

A correlação não-significativa entre a variável resposta e as variáveis preditoras observada na Figura 18 pode estar relacionada à utilização de ACF/PACF ou CCF, que são medidas lineares e, por conseguinte, podem não detectar relações não-lineares ocultas; ou ainda, à variação espacial muito frequente em variáveis hidrológicas, visto que foram considerados dados de apenas uma e única região. Além disso, esses modelos não incorporam equações ou relações físicas entre a vazão e outras variáveis. O ruído, além de diferentes características hidrometeorológicas entre treinamento e períodos de teste

conhecidas como não estacionariedade também podem deteriorar essa relação.

Um modelo de previsão hidrológica conecta a saída de um sistema hidrológico não apenas com variáveis hidroclimáticas, mas também com valores observados antecedentes da saída considerada do sistema hidrológico (KLEMEŠ, 1986; TONGAL; BOOIJ, 2018).

Os modelos de aprendizado de máquinas utilizados foram: EN, ELM, XGB, MARS e SVR, cujo objetivo foi de avaliar suas habilidades na previsão da vazão em diferentes passos à frente.

Esses modelos tiveram seus parâmetros determinados de maneira inteligente, através de algoritmos bioinspirados: GA, GWO, PSO e DE. A Tabela 12 mostra a codificação das soluções candidatas para cada modelo de aprendizado de máquina a ser utilizado em cada um dos algoritmos bioinspirados.

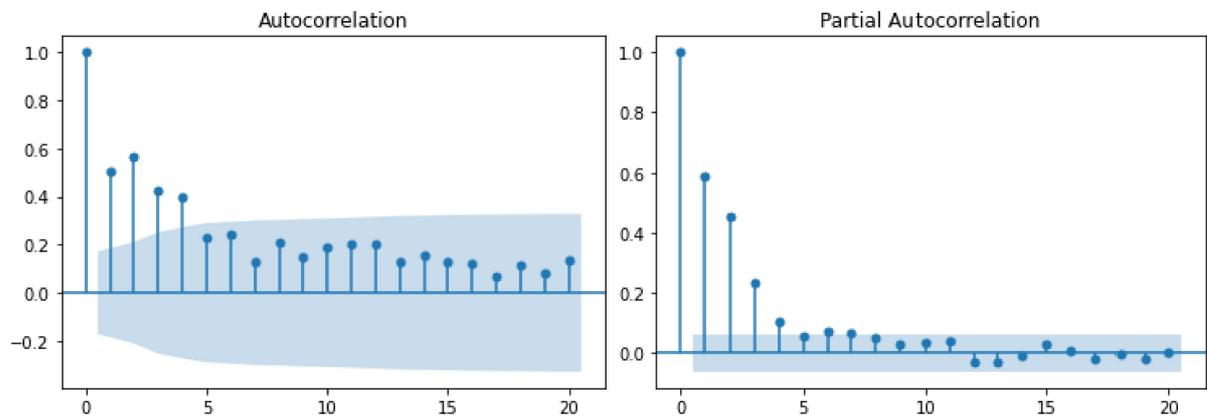
Os algoritmos bioinspirados por sua vez tiveram seus parâmetros descritos de acordo com a Tabela 13 (versões clássicas dos algoritmos de otimização). O tamanho da população considerado por cada um deles foi de 16 indivíduos aleatoriamente distribuídos (mesmo tamanho em todos algoritmos), utilizando uma distribuição uniforme no espaço de busca com parâmetros mínimo e máximo definidos pelos limites inferior e superior, respectivamente dos intervalos das soluções candidatas mostradas na Tabela 12. Cada indivíduo é um vetor composto pelos hiperparâmetros para cada modelo de aprendizado de máquinas em análise, o comprimento deste é determinado pelo número de hiperparâmetros a ser encontrados em cada modelo.

Com base na Tabela 12 da codificação algorítmica das soluções candidatas associadas a cada modelo de aprendizado de máquinas. A função objetivo do problema de minimização utilizada na busca com os algoritmos bioinspirados foi o RMSE, medido no conjunto de treinamento (conforme representado na Figura 8), utilizando um procedimento de validação cruzada de 5 *folds walk forward validation*, conforme a Figura 16. Após a conclusão da busca para cada algoritmo de busca bioinspirada, o desempenho do melhor modelo foi estimado usando o conjunto de teste, que é um subconjunto dos dados históricos. Os experimentos foram repetidos em 30 vezes independentes, com diferentes sementes aleatórias.

Esses experimentos foram subdivididos em duas etapas. A primeira é composta por quatro experimentos definidos em função da aplicação de cada metaheurística para todos os modelos de aprendizado de máquinas em análise. A avaliação da performance dos modelos foi feita dentro e entre as metaheurísticas, através da comparação de estatísticas de desempenho, gráficos e testes de hipóteses estatísticas. A segunda etapa constituída por um experimento, que consistiu na aplicação dos modelos de seleção de variáveis: LASSO e PMI ao melhor modelo resultante da primeira etapa de experimentação, para simplificação do modelo e seleção eficiente de conjuntos de variáveis com maior poder preditivo. Esses métodos de seleção de variáveis foram usados apenas com uma das metaheurísticas analisadas (a metaheurística superior).

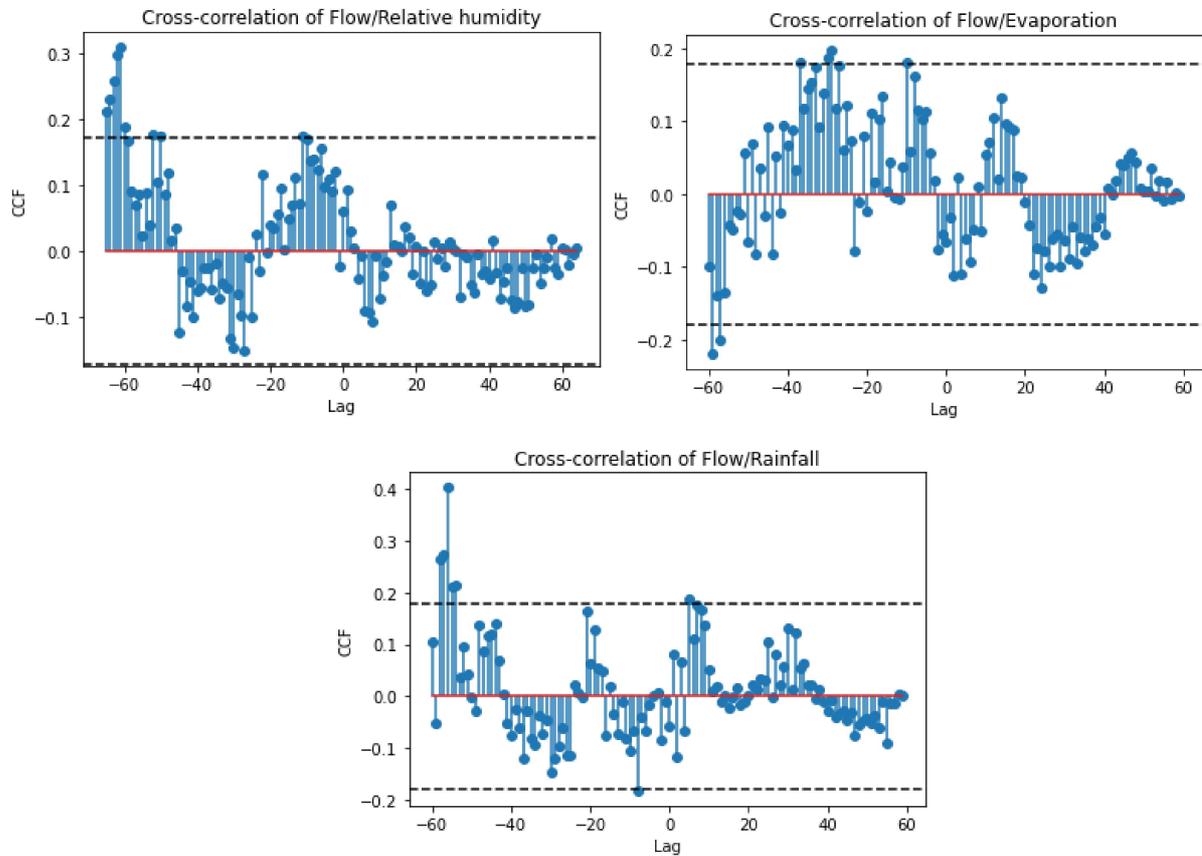
A estrutura de modelagem aplicada foi desenvolvida em linguagem de programação Python 3.9 (BROWNLEE, 2018). Os modelos de aprendizado de máquinas utilizados estão disponíveis nas bibliotecas Scikit-learn (PEDREGOSA et al., 2011) e Pyearth (MEISENBACHER et al., 2021) e os algoritmos de busca inspirados na natureza presentes na biblioteca PyGMO (OSABA et al., 2021). A visualização dos resultados foi feita mediante aplicação da biblioteca Matplotlib ou Seaborn (ROUGIER, 2012). Os experimentos foram rodados em um processador Samsung Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz, com memória ram de 24GB, ssd 256GB e placa gráfica MX110 e serão descritos a seguir.

Figura 17 - Funções de autocorrelação e autocorrelação parcial. As defasagens que se encontram dentro da parte sombreada são consideradas estatisticamente não-significativas.



Fonte: Elaborada pelo autor (2023)

Figura 18 - Funções de correlação cruzada entre vazão e precipitação, evaporação ou umidade relativa. As defasagens que se encontram entre as linhas tracejadas são consideradas estatisticamente não-significativas.



Fonte: Elaborada pelo autor (2023)

Tabela 12 – Codificação de soluções candidatas. A coluna IP indica o parâmetro interno na codificação dos algoritmos bioinspirados.

Estimador	IP	Descrição	Configurações / Intervalo
EN	$\theta_1$	Penalty term, $\alpha$	$[10^{-6}, 2]$
	$\theta_2$	$L_1$ -ratio parameter, $\rho$	$[0,1]$
ELM	$\theta_1$	No. neurons in the hidden layer, $L$	$[1, 500]$
	$\theta_2$	Regularization parameter $C$	$[0.0001, 10000]$
	$\theta_3$	Activation function $G$	1: Identity; 2: Sigmoid; 3: Hyperbolic Tangent; 4: Gaussian; 5: Swish; 6: ReLU;
SVR	$\theta_1$	Loss parameter, $\varepsilon$	$[10^{-5}, 100]$
	$\theta_2$	Regularization parameter, $C$	$[1, 10000]$
	$\theta_3$	Bandwidth parameter, $\gamma$	$[0.001, 10]$
MARS	$\theta_1$	Degree of piecewise polynomials, $q$	$[0,3]$
	$\theta_2$	Penalty factor, $\gamma$	$[1, 9]$
	$\theta_3$	Maximum number of terms, $M$	$[1, 500]$
XGB	$\theta_1$	Learning rate, $\eta$	$[10^{-6}, 1]$
	$\theta_2$	No. weak estimators, $M_{est}$	$[10, 500]$
	$\theta_3$	Maximum depth, $m_{depth}$	$[1, 20]$
	$\theta_4$	Regularization parameter, $\lambda_{reg}$	$[0, 100]$

Fonte: Elaborada pelo autor (2023)

Tabela 13 – Descrição dos parâmetros específicos dos algoritmos de otimização

Algoritmo	Descrição de parâmetros
PSO	$\omega = 0.7298$ , $c_1 = c_2 = 2.05$ (default)
GA	$cr = 0.95$ , $m = 0.2$ , crossover='single', mutation='uniform' (CARVALHO, 2017)
DE	$F = 0.9$ , $CR = 0.7$ , variant=1 (ARAUJO et al., 2016)
GWO	Nenhum específico no GWO

Fonte: Elaborada pelo autor (2023)

## 5.2 RESULTADOS E DISCUSSÃO

Esta seção relata os resultados obtidos nos experimentos computacionais realizados. Faz-se uma análise descritiva dos dados, a partir de medidas estatísticas de resumo de dados. Apresentam-se os resultados dos experimentos computacionais referentes à aplicação dos modelos de aprendizado otimizados propostos, sem e com utilização de métodos de seleção de variáveis. Uma análise quantitativa e qualitativa é conduzida para avaliação da

performance dos modelos perante medidas de desempenho e testes de hipóteses estatísticas. E, finalmente, uma análise de incerteza das previsões é utilizada para o estudo dos níveis de confiabilidade das mesmas.

### 5.2.1 ANÁLISE DESCRITIVA

O conjunto de dados exibido na Figura 8 é resumido nas Tabelas 14 e 15, que contém de informações estatísticas do conjunto de treino e teste, respectivamente.

As comparações entre as médias de cada variável no conjunto de dados de treino na Tabela 14 e as médias no conjunto de teste mostradas na Tabela 15 sugerem muita semelhança entre elas, o que sugere a existência de uma tendência linear estatisticamente não-significativa no tempo para cada variável analisada; porém uma análise formal de regressão linear entre as defasagens para cada variável deve ser aplicada.

Ademais, pode-se observar ainda em cada variável mostrada na Figura 8, um comportamento que se repete ao longo do tempo, ou seja uma flutuação periódica, o que significa que as séries em análise são sazonais. Portanto, dado que a sazonalidade é uma forma de não-estacionariedade, as séries são também não-estacionárias.

Tabela 14 – Informações estatísticas do conjunto de treinamento. A precipitação é indicada por  $R$ , afluente do vazão natural da barragem de Cahora Bassa por  $Q$ ,  $H$  é a umidade relativa e a evaporação por  $E$

	mean	std	min	25%	50%	75%	max
$R$ (mm)	2.241856	8.507345	0.000	0.000	0.000	0.000	125.800
$E$ (mm)	5.271375	2.316330	0.100	3.700	5.000	6.600	17.700
$H$ (%)	69.604497	16.750470	15.000	58.000	71.000	82.000	100.000
$Q$ ( $10^3$ m <sup>3</sup> /s)	2.162011	1.409618	0.009	1.295	1.765	2.637	10.364

Fonte: Elaborada pelo autor (2023)

Tabela 15 – Informação estatística do conjunto de teste.

	mean	std	min	25%	50%	75%	max
$R$ (mm)	1.851800	7.772177	0.000	0.0000	0.0000	0.00000	90.400
$E$ (mm)	5.805712	2.583359	0.000	4.0000	5.4000	7.60000	16.500
$H$ (%)	65.964006	17.622829	0.000	54.0000	68.0000	80.00000	99.000
$Q$ ( $10^3$ m <sup>3</sup> /s)	1.920284	0.923374	0.022	1.3555	1.7185	2.19125	10.305

Fonte: Elaborada pelo autor (2023)

## 5.3 RESULTADOS E DISCUSSÃO DAS SIMULAÇÕES

As vazões no reservatório de Cahora-Bassa foram previstas com 1, 3, 5, 7 dias à frente, com a utilização de modelos de aprendizado de máquinas EN, ELM, SVR, MARS e XGB, com seus parâmetros estimados por metaheurísticas GWO, DE, GA e PSO. Um total de oitenta e oito (88) modelos, isto é, 80 [(5 Modelos ML)x(4 Metaheurísticas)x (4

Horizontes de Previsão)] e 8 [(2 Modelos de seleção)x(4 Horizontes de Previsão)] foram desenvolvidos, ou seja vinte modelos para cada metaheurística, e dois para cada método de seleção, para cada horizonte de previsão.

Importa referir que, os modelos DE-SVR (SVR com parâmetros otimizados por DE) e GA-SVR (SVR com parâmetros otimizados por GA) foram ajustados considerando as metades dos intervalos das soluções candidatas mostradas na Tabela 12(Os limites superiores dos intervalos foram para as metades) devido à falhas na convergência durante as execuções com intervalo inteiro.

Para avaliar o desempenho destes modelos, oito diferentes medidas de desempenho e dois testes de hipóteses estatísticas foram calculadas, para medir sua robustez em termos de erro e o grau de precisão. Esta análise comparativa também foi realizada a fim de identificar o melhor modelo entre os modelos propostos. A execução bem sucedida será aquela cuja solução é conhecida, e encontrada mediante um critério de parada utilizando um número máximo permitido de avaliações. Os melhores resultados são destacados em negrito

### 5.3.1 ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS POR GWO

A Tabela 16 apresenta as estatísticas descritivas (a média e o desvio-padrão entre parêntesis) das medidas de desempenho produzidas pelos modelos de previsão propostos com seus parâmetros otimizados por GWO.

A primeira coluna mostra o horizonte de previsão, a segunda apresenta os modelos de aprendizado de máquina em análise e as colunas restantes, as medidas de desempenho. Os resultados mostram que os modelos apresentaram bom desempenho para todas as medidas em todos os horizontes de tempo em análise; porém, o modelo *Extreme Gradient Boosting(XGB)* híbrido obteve excelente performance em comparação com outros. Este modelo superou consistentemente outros modelos em todas as medidas de desempenho, nos horizontes  $t + 1$ ,  $t + 3$ , e  $t + 7$ , exceto para  $t + 5$ , onde *Multivariate Adaptive Regression Splines (MARS)* produziu o menor erro médio absoluto (MAE) e Erro percentual absoluto médio (MAPE). Embora o modelo XGB tenha superado os outros modelos de aprendizado pode se observar ainda que o modelo MARS obteve bons resultados em relação com ELM, EN e SVR, respectivamente.

A Figura 19 exhibe as distribuições de cada medida de desempenho para cada modelo nos diferentes horizontes temporais em análise. Os gráficos violinos permitem realizar uma análise qualitativa comparativa do modelo baseado na natureza da distribuição dos dados, estudo da influência dos valores extremos, apresenta mais detalhes em relação ao boxplot. Nesta figura é possível observar qualitativamente que o modelo XGB produziu melhor desempenho em relação aos outros modelos, como observado na tabela 16. Além disso, pode-se observar que as distribuições das medidas de desempenho são assimétricas para

cima ou para baixo, indicando alguma incerteza dos modelos em estimar valores extremos da vazão. O modelo XGB apresentou relativamente níveis baixos de incerteza na estimativa dos mesmos, indicando sua habilidade em estimar dados extremos. Em contraste, SVR e MARS tiveram níveis altos de incerteza na estimativa das observações extremas nas suas distribuições. Pode-se observar ainda nesta figura, um padrão de redução do desempenho dos modelos com o aumento do horizonte de previsão.

A Figura 20 mostra os gráficos das melhores soluções para cada modelo, de acordo com o RMSE pode-se observar em geral que os modelos obtiveram valores excelentes de RMSE, variando entre 0.071 e 0.171, o modelo XGB obteve o menor RMSE de  $0.071m^3/s$  para  $t + 3$ . Podemos destacar também melhores RMSE nos outros modelos: MARS com  $0.077m^3/s$  para  $t + 3$ , SVR com  $0.089m^3/s$  para previsão  $t + 1$  e finalmente, o ELM e EN ambos com  $0.098m^3/s$  também para previsão  $t + 1$ . Outras correspondentes medidas de qualidade de ajustamento não dimensionais e limitadas também podem ser visualizadas como KGE e WI sendo XGB com 0.978 e 0.991, MARS com 0.963 e 0.989, ELM com 0.956 e 0.982, EN com 0.957 e 0.982. O RMSE aumenta à medida que aumenta o horizonte de previsão, portanto, os melhores ajustes foram obtidos nos horizontes menos distantes, isto é, para  $t + 1$  e  $t + 3$ .

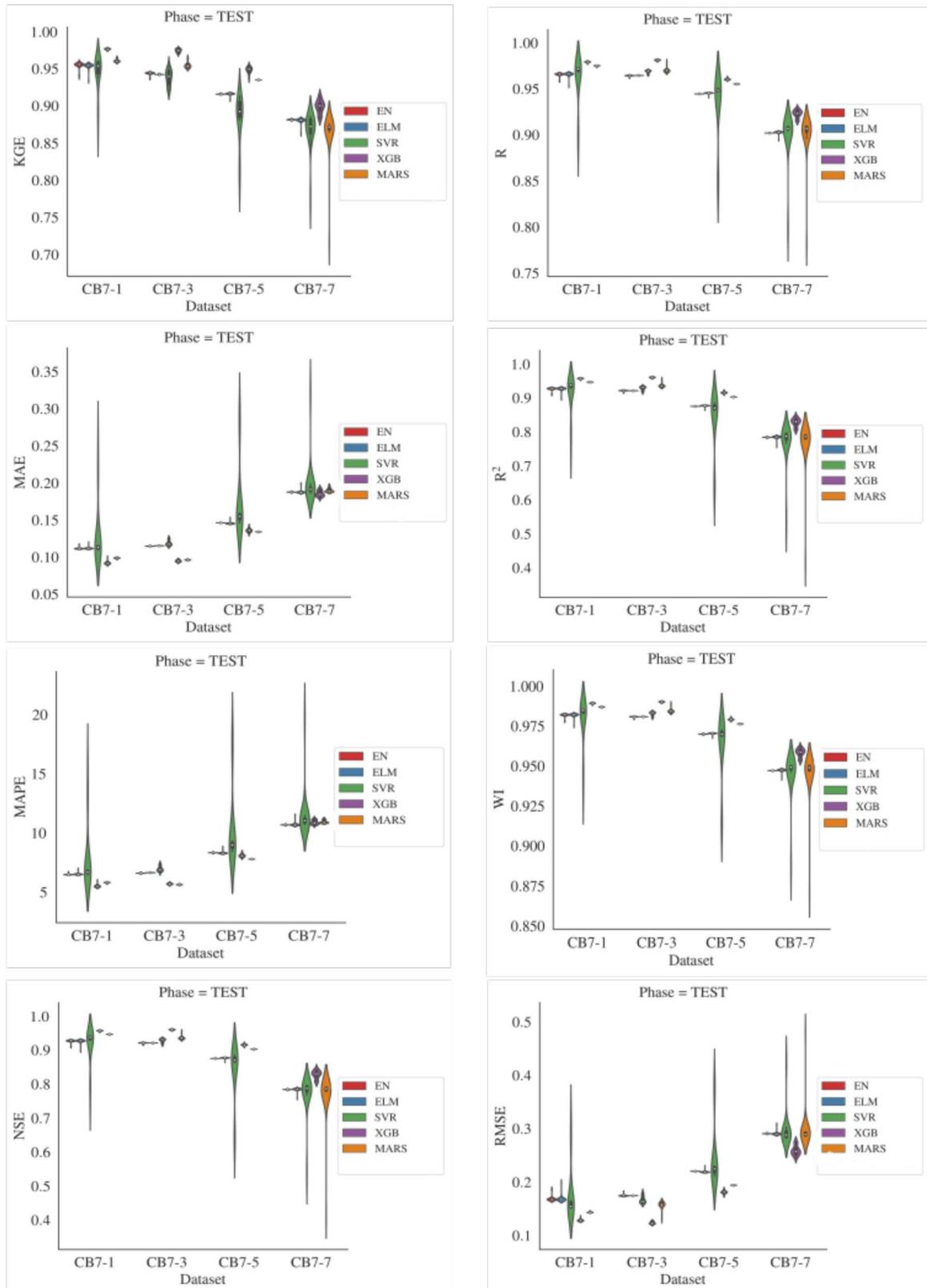
Em outra análise pode-se observar ainda nestes gráficos através da comparação entre dados observados e os dados previstos pelos modelos, que estes ficam distribuídos nas proximidades da reta ideal, o que indica boa aproximação entre dados observados e previstos pelos modelos, porém com o aumento do tempo de previsão, esta aproximação ou aderência vai perdendo a sua qualidade. A melhor aproximação foi obtida pelo modelo XGB com o horizonte de previsão  $t + 3$ .

Tabela 16 – **Média e desvio padrão das medidas de desempenho.** Desvios padrão entre parênteses. Seguindo o modelo apresentado na Eq. 4.1, a primeira linha indica o número de dias à frente que o modelo híbrido realizou a previsão. A segunda coluna exhibe os modelos de aprendizado de máquina integrados à abordagem híbrida com GWO. As colunas restantes resumem as métricas alcançadas pelo modelo. Um total de 30 execuções foram realizadas para cada entrada na tabela.

Dataset	Estimator	R	R <sup>2</sup>	WI	RMSE	MAE	MAPE	NSE	KGE
CB7-1	ELM	0.966 (0.002)	0.928 (0.005)	0.982 (0.001)	0.169 (0.006)	0.112 (0.002)	6.56 (0.093)	0.928 (0.005)	0.954 (0.004)
	EN	0.966 (0.002)	0.928 (0.004)	0.982 (0.00)	0.169 (0.005)	0.112 (0.001)	6.52 (0.052)	0.928 (0.004)	0.955 (0.004)
	MARS	0.975 (0.00)	0.948 (0.00)	0.987 (0.00)	0.144 (0.00)	0.099 (0.00)	5.83 (0.040)	0.948 (0.00)	0.961 (0.002)
	SVR	0.961 (0.026)	0.914 (0.059)	0.979 (0.016)	0.177 (0.050)	0.130 (0.042)	7.81 (2.72)	0.914 (0.059)	0.944 (0.027)
	XGB	<b>0.979</b> (0.00)	<b>0.958</b> (0.001)	<b>0.989</b> (0.00)	<b>0.129</b> (0.002)	<b>0.092</b> (0.002)	<b>5.54</b> (0.119)	<b>0.958</b> (0.001)	<b>0.977</b> (0.00)
CB7-3	ELM	0.965 (0.00)	0.922 (0.00)	0.981 (0.00)	0.175 (0.00)	0.116 (0.00)	6.67 (0.009)	0.922 (0.00)	0.943 (0.00)
	EN	0.964 (0.00)	0.922 (0.002)	0.981 (0.00)	0.175 (0.002)	0.115 (0.00)	6.63 (0.011)	0.922 (0.002)	0.944 (0.002)
	MARS	0.971 (0.003)	0.938 (0.005)	0.985 (0.001)	0.157 (0.007)	0.096 (0.00)	<b>5.66</b> (0.026)	0.938 (0.005)	0.955 (0.004)
	SVR	0.969 (0.002)	0.929 (0.006)	0.983 (0.001)	0.167 (0.006)	0.119 (0.003)	6.99 (0.216)	0.929 (0.006)	0.939 (0.010)
	XGB	<b>0.981</b> (0.00)	<b>0.961</b> (0.001)	<b>0.990</b> (0.00)	<b>0.124</b> (0.002)	<b>0.095</b> (0.001)	5.72 (0.069)	<b>0.961</b> (0.001)	<b>0.975</b> (0.002)
CB7-5	ELM	0.946 (0.00)	0.878 (0.002)	0.971 (0.00)	0.220 (0.002)	0.146 (0.001)	8.35 (0.106)	0.878 (0.002)	0.916 (0.002)
	EN	0.945 (0.00)	0.877 (0.00)	0.970 (0.00)	0.221 (0.00)	0.147 (0.00)	8.36 (0.004)	0.877 (0.00)	0.916 (0.00)
	MARS	0.956 (0.00)	0.904 (0.00)	0.977 (0.00)	0.194 (0.00)	<b>0.134</b> (0.00)	<b>7.83</b> (0.008)	0.904 (0.00)	0.936 (0.00)
	SVR	0.934 (0.036)	0.840 (0.084)	0.962 (0.020)	0.245 (0.056)	0.174 (0.047)	10.26 (3.14)	0.840 (0.084)	0.884 (0.032)
	XGB	<b>0.961</b> (0.00)	<b>0.917</b> (0.003)	<b>0.979</b> (0.00)	<b>0.181</b> (0.003)	0.136 (0.002)	8.12 (0.127)	<b>0.917</b> (0.003)	0.949 (0.004)
CB7-7	ELM	0.903 (0.001)	0.785 (0.005)	0.948 (0.00)	0.291 (0.003)	0.188 (0.002)	10.74 (0.142)	0.785 (0.005)	0.881 (0.003)
	EN	0.902 (0.00)	0.785 (0.00)	0.947 (0.00)	0.291 (0.00)	0.188 (0.00)	<b>10.72</b> (0.008)	0.785 (0.00)	0.882 (0.00)
	MARS	0.902 (0.023)	0.772 (0.067)	0.946 (0.014)	0.298 (0.034)	0.190 (0.003)	10.98 (0.129)	0.772 (0.067)	0.866 (0.028)
	SVR	0.900 (0.027)	0.771 (0.063)	0.945 (0.016)	0.298 (0.035)	0.200 (0.033)	11.63 (2.17)	0.771 (0.063)	0.867 (0.027)
	XGB	<b>0.923</b> (0.004)	<b>0.829</b> (0.012)	<b>0.959</b> (0.003)	<b>0.260</b> (0.009)	<b>0.186</b> (0.004)	10.93 (0.180)	<b>0.829</b> (0.012)	<b>0.899</b> (0.009)

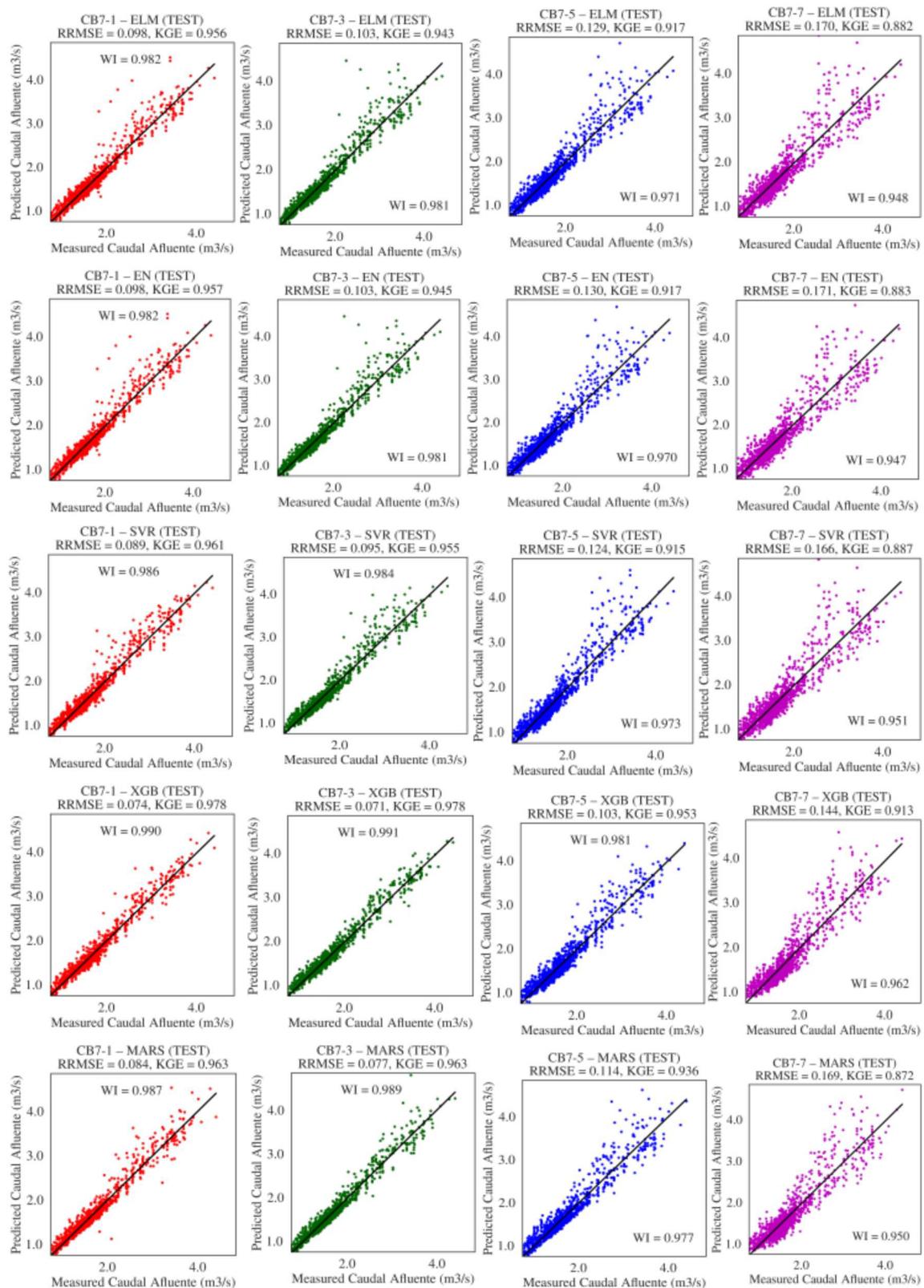
Fonte: Elaborada pelo autor (2023)

Figura 19 - Gráficos de Violino para GWO.



Fonte: Elaborada pelo autor (2023)

Figura 20 - Melhor solução de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente para GWO.



Fonte: Elaborada pelo autor (2023)

### 5.3.2 ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELO DE

A Tabela 17 apresenta uma análise quantitativa do desempenho dos modelos, médias e correspondentes desvios padrões das medidas de desempenho dos modelos otimizados por DE nos diferentes horizontes temporais são mostradas.

Os resultados desta indicam em geral que os modelos alcançaram bons desempenhos, porém O modelo SVR obteve melhor performance em comparação com outros em todas medidas de desempenho, exceto MAE e KGE onde o modelo XGB exibiu melhores resultados, para o horizonte de previsão  $t + 1$ , enquanto que, para os restantes horizontes  $t + 3$ ,  $t + 5$  e  $t + 7$ , XGB apresentou melhores resultados em quase todas medidas, menos para MAE para  $t + 3$  e  $t + 5$  onde o modelo MARS apresentou menor erro médio absoluto. Importa referir que, no caso em que XGB não superou os outros, este apresentou um resultados muito próximos do melhor, assegurando a sua superioridade e competitividade em relação aos outros.

A Figura 21 exibe os gráficos de violinos que representam as distribuições das medidas de desempenho em diferentes horizontes temporais. Quanto a natureza pode-se observar que, essas distribuições possuem assimétrias positivas ou negativas com alguma incerteza associada à estimativa dos valores extremos (outliers). O MARS foi o modelo que qualitativamente apresentou níveis altos de incerteza na estimação dos mesmos. Ainda nesta figura pode-se observar um padrão de queda da performance dos modelos à medida que aumenta o horizonte de tempo.

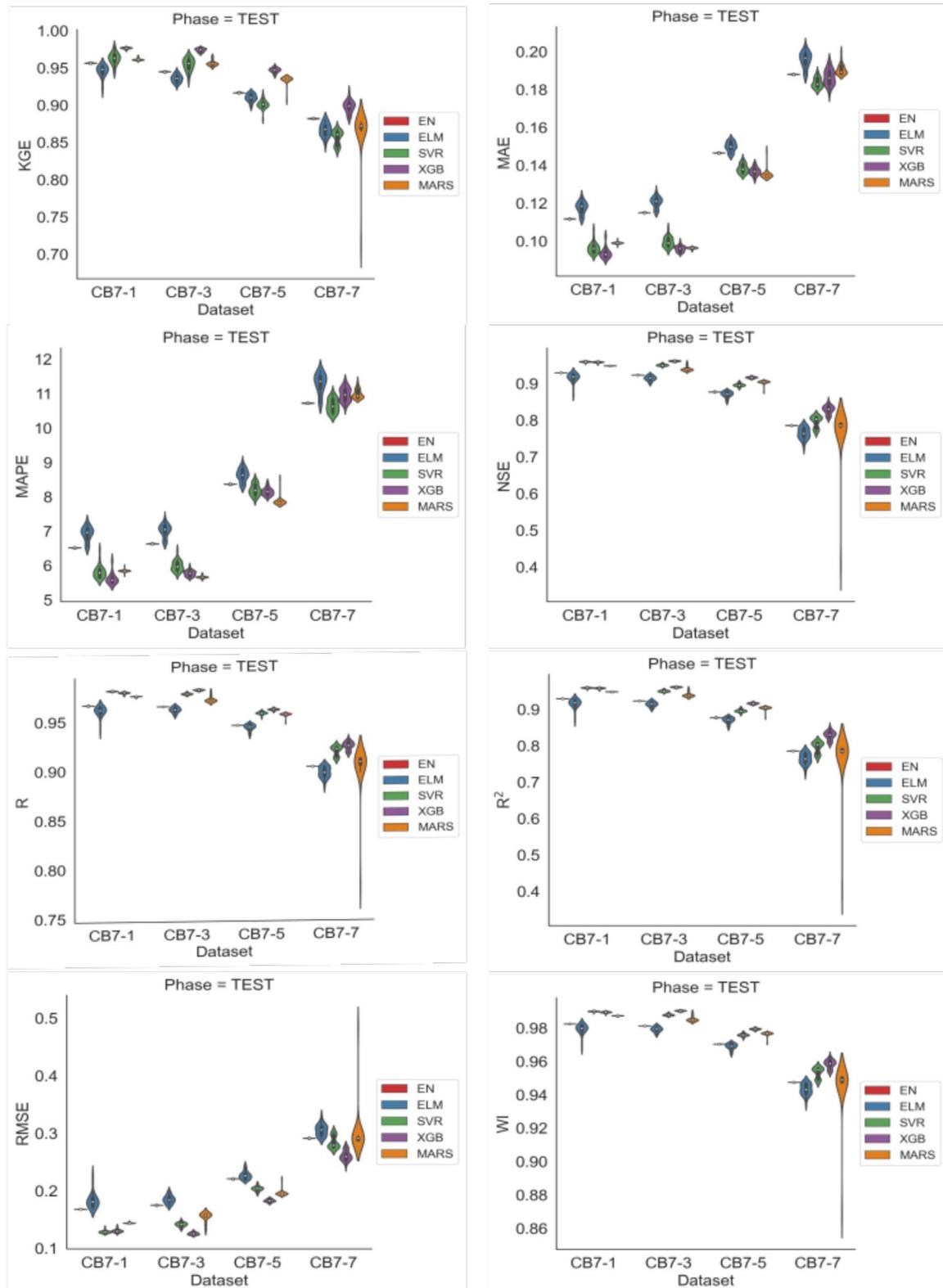
A Figura 22 mostra os gráficos das melhores soluções para cada modelo de acordo com o RMSE, em geral observa-se que os modelos obtiveram valores de RMSE muito próximos de zero, variando entre 0.071 e 0.171. O XGB obteve o menor RMSE de  $0.071m^3/s$  para  $t + 3$  seguido do SVR com  $0.073m^3/s$  para  $t + 1$ , MARS com 0.077 para previsão  $t + 3$  e finalmente, o ELM e EN ambos com  $0.098m^3/s$  para previsão  $t + 1$ . Outras correspondentes medidas de concordância entre dados observados e previstos também podem ser observadas como KGE e WI, onde o modelo SVR obteve 0.966 e 0.990, XGB 0.977 e 0.991, MARS 0.963 e 0.989, EN 0.957 e 0.982, ELM 0.955 e 0.982, respectivamente. Por outro lado, podemos observar ainda nestes gráficos a partir da comparação entre dados observados e os dados previstos pelos modelos que estes ficam distribuídos nas proximidades da reta ideal, significando uma boa aproximação entre dados observados e previstos, a melhor aproximação foi obtida pelo modelo XGB com o horizonte de previsão  $t + 3$ .

Tabela 17 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados com DE

DA	Estimator	R	R <sup>2</sup>	WI	RMSE	MAE	MAPE	NSE	KGE
1	ELM	0.961 (0.006)	0.915 (0.013)	0.979 (0.003)	0.182 (0.013)	0.118 (0.003)	6.91 (0.210)	0.915 (0.013)	0.946 (0.008)
	EN	0.966 (0.00)	0.929 (0.00)	0.982 (0.00)	0.168 (0.00)	0.112 (0.00)	6.51 (0.006)	0.929 (0.00)	0.956 (0.00)
	MARS	0.975 (0.00)	0.948 (0.00)	0.987 (0.00)	0.144 (0.00)	0.099 (0.00)	5.84 (0.037)	0.948 (0.00)	0.961 (0.002)
	SVR	<b>0.981</b> (0.00)	<b>0.958</b> (0.002)	<b>0.990</b> (0.00)	<b>0.128</b> (0.002)	0.096 (0.003)	5.80 (0.182)	<b>0.958</b> (0.002)	0.963 (0.008)
	XGB	0.979 (0.00)	0.957 (0.002)	0.989 (0.00)	0.130 (0.003)	<b>0.094</b> (0.003)	<b>5.61</b> (0.167)	0.957 (0.002)	<b>0.976</b> (0.001)
3	ELM	0.961 (0.003)	0.913 (0.007)	0.979 (0.002)	0.185 (0.007)	0.121 (0.003)	7.01 (0.185)	0.913 (0.007)	0.936 (0.005)
	EN	0.965 (0.00)	0.923 (0.00)	0.981 (0.00)	0.175 (0.00)	0.115 (0.00)	6.63 (0.006)	0.923 (0.00)	0.945 (0.00)
	MARS	0.971 (0.003)	0.939 (0.006)	0.985 (0.001)	0.155 (0.008)	<b>0.096</b> (0.00)	<b>5.66</b> (0.034)	0.939 (0.006)	0.956 (0.003)
	SVR	0.977 (0.00)	0.949 (0.003)	0.988 (0.00)	0.141 (0.004)	0.100 (0.003)	5.98 (0.154)	0.949 (0.003)	0.954 (0.008)
	XGB	<b>0.981</b> (0.00)	<b>0.960</b> (0.001)	<b>0.990</b> (0.00)	<b>0.125</b> (0.002)	0.096 (0.002)	5.77 (0.082)	<b>0.960</b> (0.001)	<b>0.974</b> (0.002)
5	ELM	0.943 (0.003)	0.870 (0.008)	0.969 (0.002)	0.227 (0.007)	0.150 (0.003)	8.62 (0.187)	0.870 (0.008)	0.910 (0.005)
	EN	0.945 (0.00)	0.877 (0.00)	0.970 (0.00)	0.221 (0.00)	0.147 (0.00)	8.36 (0.005)	0.877 (0.00)	0.917 (0.00)
	MARS	0.955 (0.002)	0.903 (0.005)	0.976 (0.001)	0.196 (0.005)	<b>0.135</b> (0.002)	<b>7.86</b> (0.122)	0.903 (0.005)	0.934 (0.005)
	SVR	0.957 (0.001)	0.895 (0.004)	0.976 (0.00)	0.203 (0.004)	0.138 (0.002)	8.20 (0.161)	0.895 (0.004)	0.901 (0.006)
	XGB	<b>0.960</b> (0.00)	<b>0.916</b> (0.002)	<b>0.979</b> (0.00)	<b>0.183</b> (0.002)	0.137 (0.002)	8.14 (0.107)	<b>0.916</b> (0.002)	<b>0.947</b> (0.003)
7	ELM	0.896 (0.006)	0.764 (0.016)	0.943 (0.003)	0.305 (0.010)	0.196 (0.004)	11.30 (0.288)	0.764 (0.016)	0.867 (0.009)
	EN	0.902 (0.00)	0.785 (0.00)	0.947 (0.00)	0.291 (0.00)	0.188 (0.00)	10.72 (0.006)	0.785 (0.00)	0.882 (0.00)
	MARS	0.902 (0.023)	0.773 (0.068)	0.946 (0.014)	0.297 (0.035)	0.190 (0.003)	10.99 (0.132)	0.773 (0.068)	0.867 (0.029)
	SVR	0.918 (0.005)	0.795 (0.015)	0.953 (0.003)	0.284 (0.010)	<b>0.184</b> (0.003)	<b>10.65</b> (0.196)	0.795 (0.015)	0.856 (0.009)
	XGB	<b>0.923</b> (0.004)	<b>0.828</b> (0.011)	<b>0.958</b> (0.003)	<b>0.261</b> (0.009)	0.187 (0.004)	10.97 (0.196)	0.828 (0.011)	<b>0.898</b> (0.008)

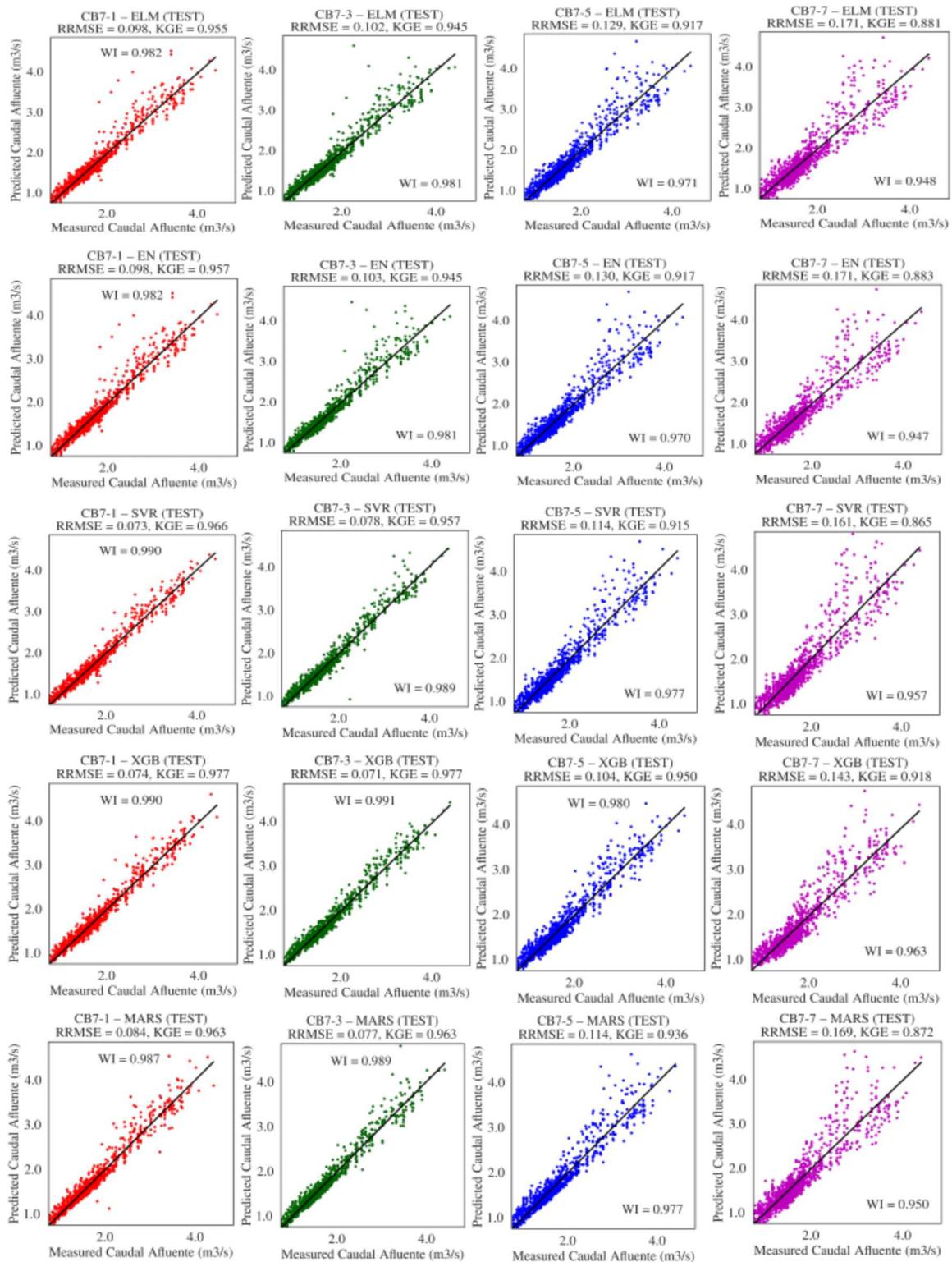
Fonte: Elaborada pelo autor (2023)

Figura 21 - Gráficos de Violino para DE.



Fonte: Elaborada pelo autor (2023)

Figura 22 - Melhor solução de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente para DE.



Fonte: Elaborada pelo autor (2023)

### 5.3.3 ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELOS GA

A Tabela 18 apresenta as estatísticas descritivas, média e o desvio padrão das medidas de desempenho produzidas pelos modelos de previsão cujos seus parâmetros foram otimizados por GA.

Os resultados indicaram bom desempenho dos modelos em todos os horizontes  $t + 1$ ,  $t + 3$ ,  $t + 5$  e  $t + 7$ . O modelo híbrido Extreme Gradient Boosting(XGB) superou outros modelos para todas as medidas e horizontes, exceto para  $t + 3$  e  $t + 5$  onde o MARS resultou com o menor MAE e para  $t + 7$  em que os modelos SVR e EN obtiveram o menor MAE e MAPE, respectivamente. Pode-se destacar ainda que o modelo MARS e SVR obteve bons resultados em comparação com ELM, EN.

A Figura 23 mostra as distribuições das medidas de desempenho para cada modelo nos diferentes horizontes temporais. Esta figura sinaliza uma queda de desempenho dos modelos em conformidade com aumento do horizonte de previsão. Pode se verificar ainda qualitativamente que o modelo XGB produziu melhor desempenho em relação aos outros modelos. Além disso, observam-se maiores assimetrias nas distribuições das medidas de desempenho, sinalizando alguma incerteza na estimativa dos valores extremos. Os modelos SVR e MARS tiveram níveis mais altos de incerteza de estimação de valores extremos nas suas distribuições em comparação com os outros, provocando maior variabilidade das mesmas.

A Figura 24 apresenta os gráficos das melhores soluções para cada modelo de acordo com a métrica RMSE. Em geral, observa-se os modelos obtiveram valores baixos de RMSE, porém, XGB obteve o menor RMSE de  $0.071m^3/s$  para  $t + 3$  seguido do SVR com  $0.073m^3/s$  para  $t + 1$ , MARS com 0.077 para previsão  $t + 3$  e, o ELM e EN ambos com  $0.098m^3/s$  também para previsão  $t + 1$ . Outras medidas de ajuste também podem ser visualizadas como KGE e WI tendo o modelo XGB obtido 0.978 e 0.991, MARS 0.963 e 0.989, SVR 0.973 e 0.990, EN 0.958 e 0.982, ELM 0.955 e 0.982, respectivamente.

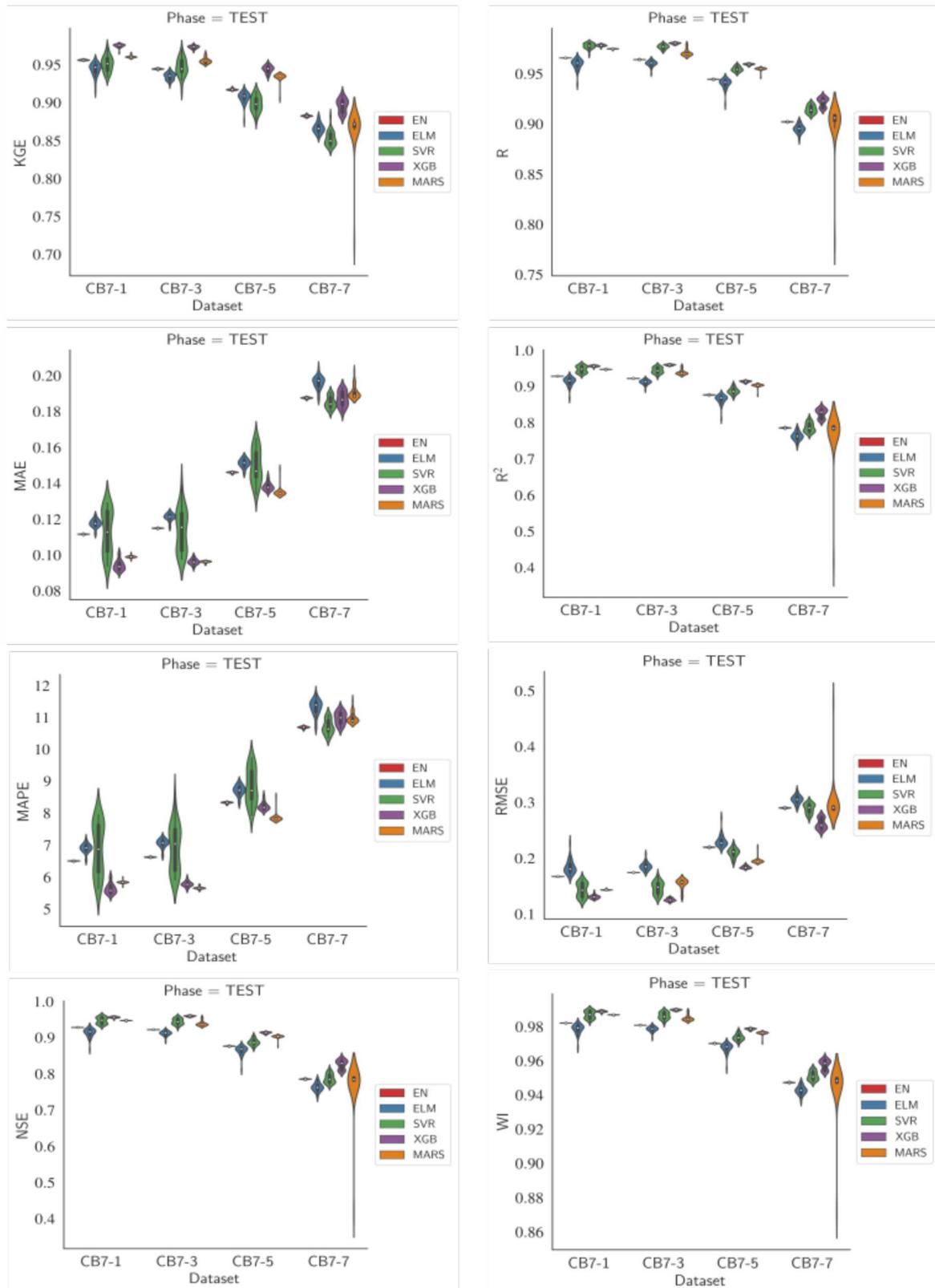
Por outro lado, pode-se visualizar ainda nestes gráficos comparando dados observados e os dados previstos pelos modelos que, estes ficam próximo da reta ideal, o que indica boa aproximação entre dados observados e previstos, isto é uma boa reprodução de todas as características dos dados observados através de modelos, XGB com  $t + 3$  de antecedência foi o modelo com a melhor aproximação.

Tabela 18 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados pelo GA

DA	Estimator	R	R <sup>2</sup>	WI	RMSE	MAE	MAPE	NSE	KGE
1	ELM	0.960 (0.005)	0.914 (0.012)	0.979 (0.003)	0.184 (0.012)	0.118 (0.002)	6.91 (0.140)	0.914 (0.012)	0.944 (0.008)
	EN	0.966 (0.00)	0.929 (0.00)	0.982 (0.00)	0.168 (0.00)	0.112 (0.00)	6.51 (0.008)	0.929 (0.00)	0.957 (0.00)
	MARS	0.975 (0.00)	0.948 (0.00)	0.987 (0.00)	0.144 (0.00)	0.099 (0.00)	5.84 (0.038)	0.948 (0.00)	0.961 (0.002)
	SVR	0.978 (0.003)	0.947 (0.009)	0.987 (0.002)	0.143 (0.012)	0.113 (0.012)	6.88 (0.790)	0.947 (0.009)	0.952 (0.010)
	XGB	<b>0.979</b> (0.00)	<b>0.956</b> (0.002)	<b>0.989</b> (0.00)	<b>0.131</b> (0.003)	<b>0.094</b> (0.003)	<b>5.64</b> (0.160)	<b>0.956</b> (0.002)	<b>0.976</b> (0.002)
3	ELM	0.961 (0.003)	0.912 (0.007)	0.979 (0.002)	0.186 (0.007)	0.121 (0.002)	7.06 (0.138)	0.912 (0.007)	0.935 (0.005)
	EN	0.965 (0.00)	0.923 (0.00)	0.981 (0.00)	0.175 (0.00)	0.115 (0.00)	6.63 (0.011)	0.923 (0.00)	0.945 (0.00)
	MARS	0.971 (0.003)	0.939 (0.007)	0.985 (0.002)	0.154 (0.009)	<b>0.096</b> (0.00)	<b>5.66</b> (0.039)	0.939 (0.007)	0.956 (0.004)
	SVR	0.977 (0.002)	0.945 (0.008)	0.987 (0.002)	0.147 (0.011)	0.114 (0.011)	6.89 (0.711)	0.945 (0.008)	0.947 (0.012)
	XGB	<b>0.981</b> (0.00)	<b>0.960</b> (0.002)	<b>0.990</b> (0.00)	<b>0.126</b> (0.003)	0.096 (0.002)	5.78 (0.085)	<b>0.960</b> (0.002)	<b>0.974</b> (0.002)
5	ELM	0.941 (0.005)	0.865 (0.013)	0.968 (0.003)	0.231 (0.010)	0.151 (0.002)	8.72 (0.161)	0.865 (0.013)	0.907 (0.008)
	EN	0.945 (0.00)	0.877 (0.00)	0.970 (0.00)	0.220 (0.00)	0.146 (0.00)	8.33 (0.026)	0.877 (0.00)	0.918 (0.001)
	MARS	0.955 (0.002)	0.903 (0.005)	0.976 (0.001)	0.195 (0.005)	<b>0.135</b> (0.002)	<b>7.85</b> (0.122)	0.903 (0.005)	0.934 (0.005)
	SVR	0.955 (0.002)	0.888 (0.009)	0.974 (0.002)	0.210 (0.008)	0.150 (0.009)	8.85 (0.532)	0.888 (0.009)	0.899 (0.010)
	XGB	<b>0.960</b> (0.00)	<b>0.914</b> (0.002)	<b>0.979</b> (0.00)	<b>0.184</b> (0.002)	0.138 (0.002)	8.22 (0.124)	<b>0.914</b> (0.002)	<b>0.945</b> (0.005)
7	ELM	0.896 (0.004)	0.763 (0.011)	0.943 (0.003)	0.306 (0.007)	0.196 (0.003)	11.32 (0.236)	0.763 (0.011)	0.867 (0.007)
	EN	0.903 (0.00)	0.786 (0.001)	0.948 (0.00)	0.290 (0.00)	0.188 (0.00)	<b>10.69</b> (0.032)	0.786 (0.001)	0.883 (0.001)
	MARS	0.901 (0.022)	0.771 (0.066)	0.946 (0.014)	0.299 (0.034)	0.191 (0.003)	10.99 (0.161)	0.771 (0.066)	0.866 (0.028)
	SVR	0.915 (0.004)	0.787 (0.013)	0.951 (0.003)	0.289 (0.009)	<b>0.185</b> (0.003)	10.71 (0.226)	0.787 (0.013)	0.853 (0.009)
	XGB	<b>0.922</b> (0.004)	<b>0.826</b> (0.013)	<b>0.958</b> (0.003)	<b>0.262</b> (0.010)	0.187 (0.004)	10.98 (0.193)	<b>0.826</b> (0.013)	<b>0.897</b> (0.009)

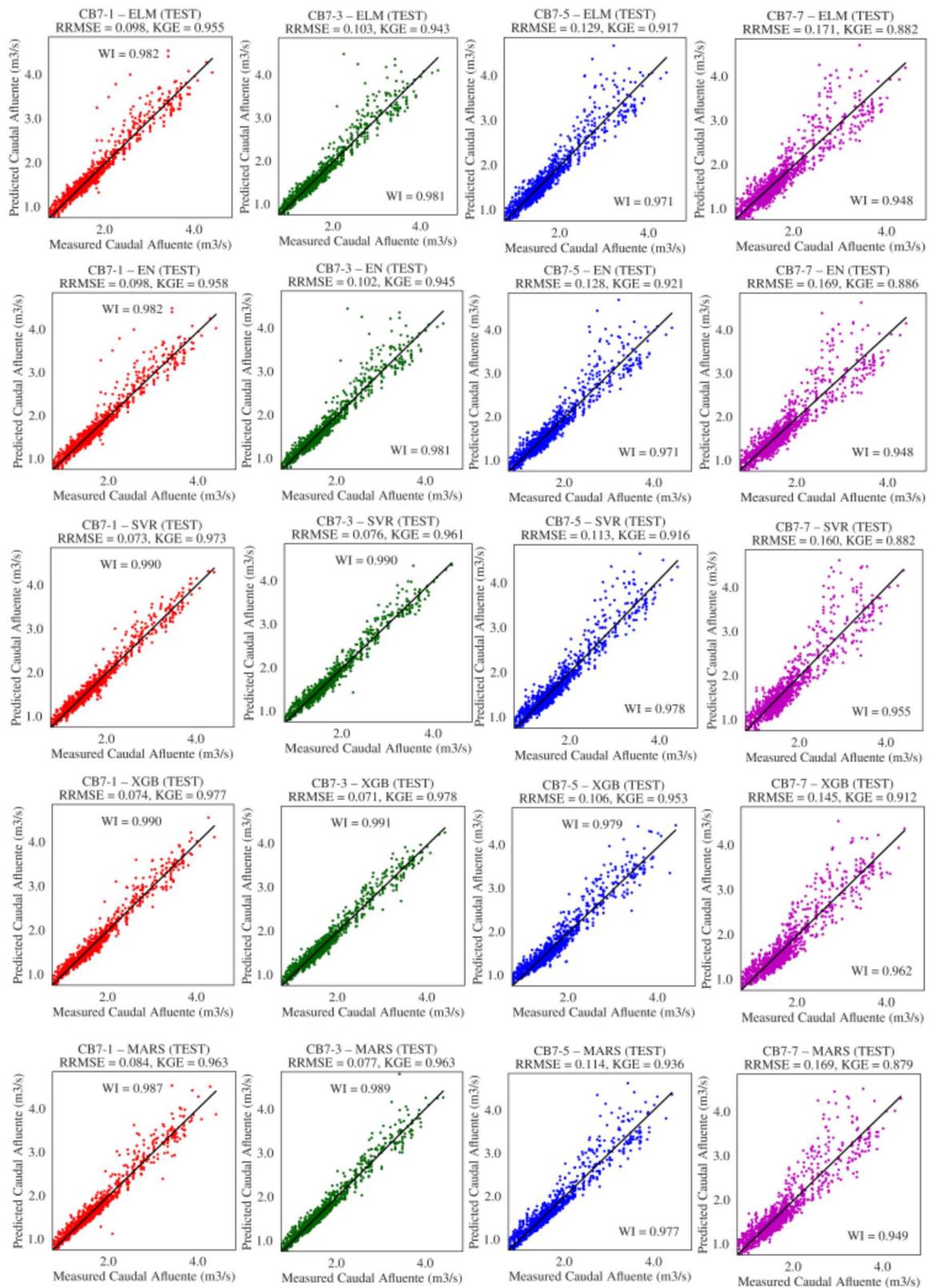
Fonte: Elaborada pelo autor (2023)

Figura 23 - Gráficos de Violino para GA.



Fonte: Elaborada pelo autor (2023)

Figura 24 - Melhor solução para cada modelo de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente com GA.



Fonte: Elaborada pelo autor (2023)

### 5.3.4 ANÁLISE DO DESEMPENHO DOS MODELOS OTIMIZADOS PELO PSO

A Tabela 19 demonstra as estatísticas descritivas, média e o desvio padrão das medidas de desempenho produzidas pelos modelos de previsão com seus parâmetros otimizados por PSO.

Os resultados demonstraram bom desempenho dos modelos em todos os horizontes temporais em análise. O modelo Extreme Gradient Boosting(XGB) híbrido foi o modelo com melhor performance em relação aos outros em todas as medidas de desempenho para o horizonte  $t + 1$ . Para os restantes  $t + 3$ ,  $t + 5$  e  $t + 7$ , este modelo obteve também melhor desempenho em quase todas as medidas, menos para  $t + 3$  e  $t + 5$  onde o MARS apresentou menores resultados de MAPE e MAE, e para  $t + 7$  onde EN apresentou o menor MAPE.

A Figura 25 exibe as distribuições de cada medida de desempenho para cada modelo nos diferentes horizontes temporais. Na figura observar-se um relativo equilíbrio das medidas de desempenho dos modelos à medida em que aumenta o horizonte de previsão com uma tendência de redução. Em outra análise, assimetrias podem ser observadas nas distribuições das medidas de desempenho, indicando a existência de alguma incerteza na estimação dos valores extremos. O SVR foi o modelo com níveis mais altos de incerteza no ajuste das observações extremas nas suas distribuições, seguido do MARS.

A Figura 26 apresenta os gráficos das melhores soluções para cada modelo de acordo com o RMSE, pode-se verificar que os modelos obtiveram valores pequenos de RMSE situando-se entre 0.071 e 0.171, o modelo XGB obteve o menor RMSE de  $0.071m^3/s$ . Importa destacar também menores valores de RMSE nos outros modelos: MARS com  $0.077m^3/s$  ambos para  $t + 3$ , SVR com 0.089 para previsão  $t + 1$  e, ELM e EN ambos com  $0.098m^3/s$  também para previsão  $t + 1$ . Outras medidas de ajuste também podem ser consideradas KGE e WI tendo o modelo XGB obtido 0.977 e 0.991, MARS 0.963 e 0.989, SVR 0.964 e 0.986, EN 0.957 e 0.982, ELM 0.956 e 0.982, respectivamente. Portanto, pode ser observado que os horizontes  $t + 1$  e  $t + 3$  têm melhores qualidades de ajustamento do que para  $t + 5$  e  $t + 7$ .

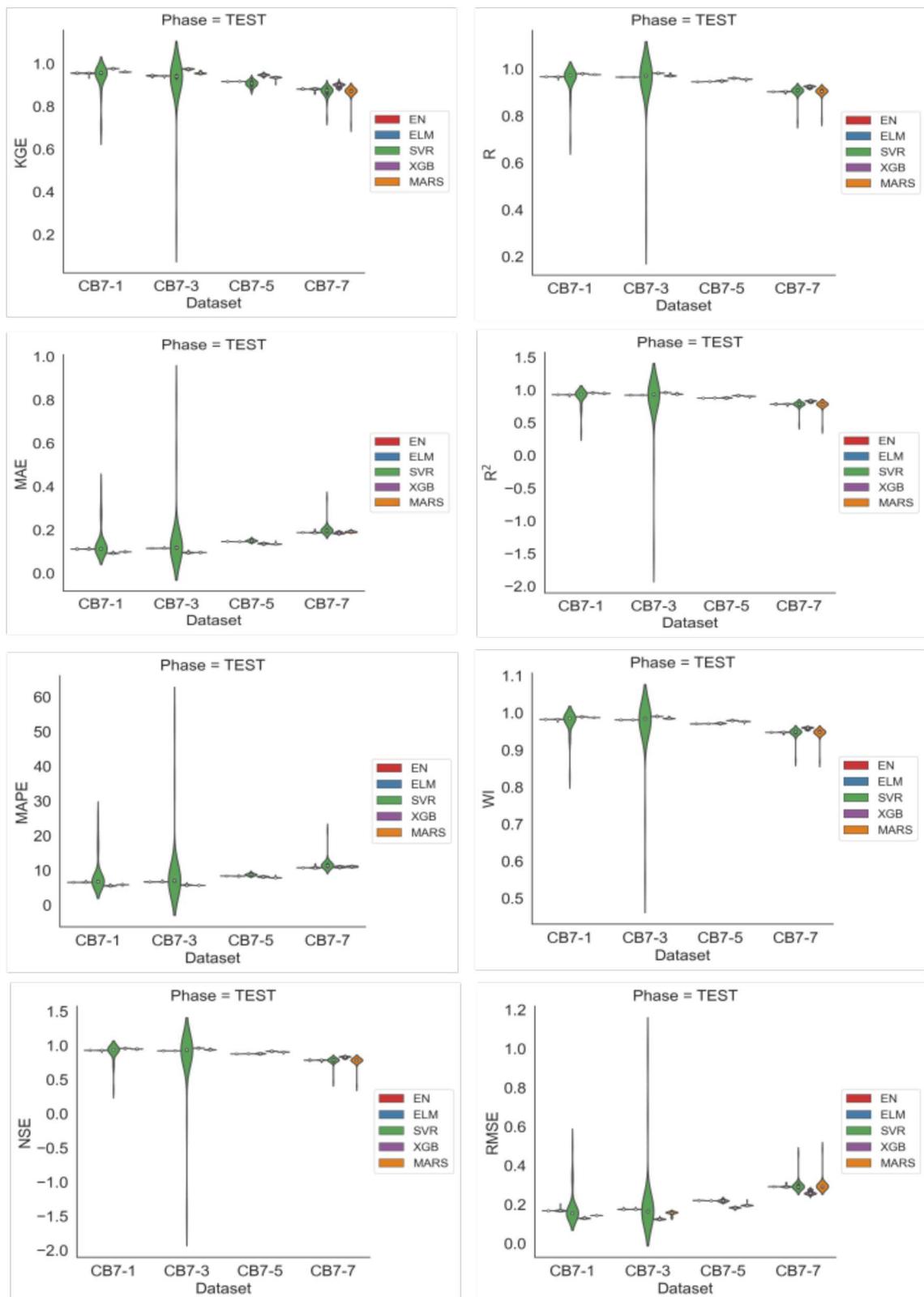
Em outra análise, observar-se ainda nesta figura uma boa aproximação entre valores observados em relação aos previstos pelos modelos, pois comparando os dados observados e os dados previstos pelos modelos frente a reta ideal verifica-se que estes ficam ao longo da mesma, o XGB com  $t + 3$  de antecedência foi o modelo com a maior aderência.

Tabela 19 – Estatísticas descritivas (médias e Desvios padrão) das medidas de desempenho dos modelos otimizados com PSO

DA	Estimator	R	R <sup>2</sup>	WI	RMSE	MAE	MAPE	NSE	KGE
1	ELM	0.966 (0.002)	0.928 (0.005)	0.982 (0.001)	0.169 (0.006)	0.112 (0.002)	6.56 (0.116)	0.928 (0.005)	0.954 (0.004)
	EN	0.966 (0.00)	0.929 (0.00)	0.982 (0.00)	0.168 (0.00)	0.112 (0.00)	6.52 (0.013)	0.929 (0.00)	0.956 (0.00)
	MARS	0.975 (0.00)	0.948 (0.00)	0.987 (0.00)	0.144 (0.00)	0.099 (0.00)	5.83 (0.029)	0.948 (0.00)	0.961 (0.002)
	SVR	0.955 (0.055)	0.897 (0.122)	0.975 (0.032)	0.184 (0.081)	0.139 (0.068)	8.43 (4.52)	0.897 (0.122)	0.934 (0.066)
	XGB	<b>0.979</b> (0.00)	<b>0.958</b> (0.001)	<b>0.989</b> (0.00)	<b>0.129</b> (0.002)	<b>0.093</b> (0.002)	<b>5.57</b> (0.127)	<b>0.958</b> (0.001)	<b>0.976</b> (0.001)
3	ELM	0.965 (0.00)	0.922 (0.002)	0.981 (0.00)	0.175 (0.002)	0.116 (0.002)	6.70 (0.112)	0.922 (0.002)	0.942 (0.002)
	EN	0.965 (0.00)	0.922 (0.002)	0.981 (0.00)	0.175 (0.002)	0.115 (0.00)	6.65 (0.026)	0.922 (0.002)	0.943 (0.002)
	MARS	0.971 (0.003)	0.939 (0.007)	0.985 (0.002)	0.155 (0.009)	<b>0.096</b> (0.00)	<b>5.66</b> (0.035)	0.939 (0.007)	0.956 (0.004)
	SVR	0.932 (0.142)	0.818 (0.460)	0.960 (0.090)	0.208 (0.169)	0.156 (0.143)	9.43 (9.50)	0.818 (0.460)	0.901 (0.148)
	XGB	<b>0.981</b> (0.00)	<b>0.960</b> (0.002)	<b>0.990</b> (0.00)	<b>0.125</b> (0.003)	0.096 (0.002)	5.75 (0.125)	<b>0.960</b> (0.002)	<b>0.974</b> (0.003)
5	ELM	0.946 (0.00)	0.878 (0.00)	0.971 (0.00)	0.219 (0.00)	0.146 (0.00)	8.32 (0.007)	0.878 (0.00)	0.917 (0.00)
	EN	0.945 (0.00)	0.876 (0.00)	0.970 (0.00)	0.221 (0.00)	0.147 (0.00)	8.36 (0.002)	0.876 (0.00)	0.916 (0.00)
	MARS	0.955 (0.002)	0.903 (0.005)	0.976 (0.001)	0.195 (0.005)	<b>0.135</b> (0.002)	<b>7.85</b> (0.123)	0.903 (0.005)	0.934 (0.005)
	SVR	0.948 (0.002)	0.879 (0.006)	0.971 (0.001)	0.218 (0.006)	0.150 (0.005)	8.73 (0.316)	0.879 (0.006)	0.906 (0.014)
	XGB	<b>0.960</b> (0.001)	<b>0.916</b> (0.003)	<b>0.979</b> (0.00)	<b>0.182</b> (0.003)	0.137 (0.003)	8.15 (0.126)	<b>0.916</b> (0.003)	<b>0.947</b> (0.005)
7	ELM	0.902 (0.003)	0.783 (0.008)	0.947 (0.002)	0.292 (0.005)	0.189 (0.003)	10.81 (0.255)	0.783 (0.008)	0.880 (0.006)
	EN	0.902 (0.00)	0.785 (0.00)	0.947 (0.00)	0.291 (0.00)	0.188 (0.00)	<b>10.73</b> (0.011)	0.785 (0.00)	0.881 (0.00)
	MARS	0.901 (0.023)	0.770 (0.068)	0.945 (0.014)	0.299 (0.035)	0.192 (0.003)	11.05 (0.145)	0.770 (0.068)	0.867 (0.029)
	SVR	0.902 (0.025)	0.773 (0.059)	0.946 (0.014)	0.297 (0.031)	0.202 (0.027)	11.74 (1.85)	0.773 (0.059)	0.867 (0.028)
	XGB	<b>0.923</b> (0.004)	<b>0.828</b> (0.012)	<b>0.958</b> (0.003)	<b>0.261</b> (0.009)	<b>0.186</b> (0.004)	10.95 (0.173)	<b>0.828</b> (0.012)	<b>0.898</b> (0.009)

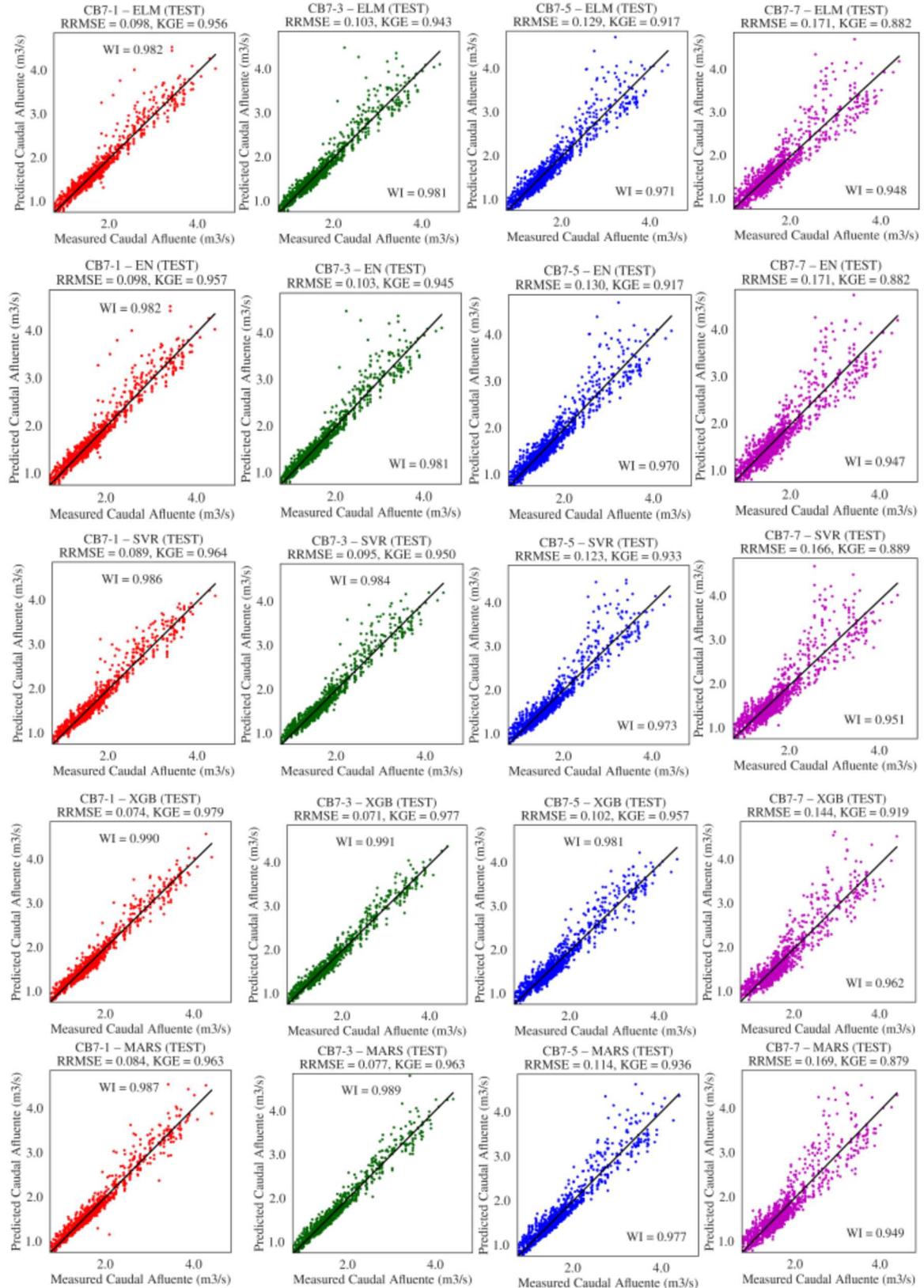
Fonte: Elaborada pelo autor (2023)

Figura 25 - Gráficos de Violino para PSO.



Fonte: Elaborada pelo autor (2023)

Figura 26 - Melhor solução para cada modelo de acordo com RMSE para fluxos de 1, 3, 5, 7 dias à frente com PSO.



Fonte: Elaborada pelo autor (2023)

### 5.3.5 ANÁLISE COMPARATIVA DOS RESULTADOS

Uma análise dos resultados obtidos com diferentes metaheurísticas utilizadas neste trabalho permitiu verificar quantitativamente que, de maneira geral, todos os modelos obtiveram boas performances em todas as medidas estatísticas de avaliação de desempenho, mesmo nos horizontes de previsão relativamente distantes. Portanto, a integração dos algoritmos evolutivos e/ou bioinspirados na otimização dos parâmetros dos modelos de aprendizado de máquinas permitiu a obtenção destes bons resultados na previsão multipassos à frente da vazão diária. Os algoritmos evolucionários/bioinspirados têm provado serem capazes de encontrar uma boa aproximação para problemas complexos, e também têm alcançado bons resultados na determinação de soluções de boa qualidade na resolução desses problemas (SANTOS, 2019), isto é, permitem a obtenção de *state-of-art results* (NGUYEN et al., 2022).

A Tabela 20 apresenta os resultados do teste estatístico ANOVA de um fator. A hipótese nula do teste Anova é que a média em cada métrica de avaliação é igual para todas as metaheurísticas ou modelos. Como pode ser visto, rejeita-se a hipótese nula para todas as métricas, porque os valores-p de cada uma delas são menores que o nível de significância de 0.05. Isso significa que, todas as métricas podem ser utilizadas como critério para avaliação do desempenho das diferentes metaheurísticas ou modelos de previsão.

Tabela 20 –  $p$ -valores do teste ANOVA para cada medida de Desempenho

Metric	R	R <sup>2</sup>	RMSE	MAPE	MAE	WI	NSE	KGE
$p$ -value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Fonte: Elaborada pelo autor (2023)

A Tabela 21, por sua vez, mostra os resultados de um teste de Tukey para comparações múltiplas de pares de médias das metaheurísticas. Esta tabela apresenta na primeira coluna os pares de metaheurísticas, na segunda a diferença média, na terceira o correspondente valor-p, na quarta e quinta os limites mínimos e máximos dos intervalos de confiança correspondentes as diferenças das médias dos pares de metaheurísticas e, por último a decisão tomada. A hipótese nula é de que as médias em cada par de metaheurísticas são iguais. Como se pode observar, rejeita-se esta hipótese nula, pois alguns pares apresentam valores-p menores que o nível de significância de 0.05. Portanto, algumas médias de uma metaheurística são diferentes das outras. Além disso, pode-se observar ainda que, a metaheurística GWO mostrou-se superior que as outras metaheurísticas.

Tabela 21 – Teste de Comparações Múltiplas de Médias das metaheurísticas - Tukey HSD, Significância=0.05

Pair	meandiff	P-adj	Lower	Upper	Reject
<b>DE-GWO</b>	<b>0.007</b>	<b>0.017</b>	<b>0.001</b>	<b>0.013</b>	<b>True</b>
DE-PSO	0.004	0.268	-0.002	0.011	False
DE-SGA	-0.001	0.900	-0.007	0.005	False
GWO-PSO	-0.003	0.664	-0.009	0.004	False
<b>GWO-SGA</b>	<b>-0.008</b>	<b>0.005</b>	<b>-0.014</b>	<b>-0.002</b>	<b>True</b>
PSO-SGA	-0.005	0.130	-0.012	0.001	False

Fonte: Elaborada pelo autor (2023)

O GWO é um algoritmo usado com sucesso no campo da engenharia e que pode produzir desempenho de alto nível em comparação com outras metaheurísticas conhecidas (MIRJALILI, 2015), e mostrou ser superior em vários trabalhos sobre previsão de vazão, com particular destaque os seguintes: Tikhamarine et al. (2019), Riahi-Madvar et al. (2021), Martinho et al. (2022), Tikhamarine et al. (2020), Kilinc e Yurtsever (2022), Dehghani et al. (2019), entre outros.

A superioridade deste algoritmo, pode estar relacionada com sua capacidade de exploração e refinamento no espaço de busca de soluções, duas propriedades importantes no estudo do poder dos algoritmos evolutivos/bioinspirados para solucionar os problemas, que é muito intensa, evitando ótimos locais (MIRJALILI et al., 2014).

Importa referir que DE e GA tiveram diferenças não- significativas. Esta constatação foi observada por Nguyen et al. (2021). Ao comparar as performances do modelo *Extreme gradient boosting* em relação à dois algoritmos evolutivos: Algoritmos genéticos e Evolução Diferencial, ou seja, GA-XGB e DE-XGB, verificaram que estes modelos apresentaram também resultados similares.

A comparação entre os modelos de aprendizado de máquinas em análise neste trabalho é por sua vez feita na Tabela 22, usando o teste de comparações múltiplas de médias ou teste Tukey ( $\alpha = 0.05$ ) dos pares dos modelos. A hipótese nula é de que as médias em cada par de modelos são iguais. O resultado deste teste mostra que rejeita-se a hipótese nula, pois como se pode observar alguns pares apresentam valores-p menores que o nível de significância de 0.05. Portanto, algumas médias de um modelo são diferentes das médias dos outros modelos. Isto indica ainda que o modelo *Extreme gradient boosting* (XGB) superou os outros modelos, de acordo com essas métricas de avaliação em todas as metaheurísticas e todos horizontes de previsão. O *Elastic net* foi o modelo com resultados baixos.

Tabela 22 – Teste de Comparações Múltiplas de Médias dos modelos - Tukey HSD, Significância=0.05

Pair	meandiff	p-adj	Lower	Upper	Reject
ELM-EN	0.004	0.753	-0.005	0.012	False
ELM-MARS	0.013	0.001	0.004	0.021	True
ELM-SVR	0.008	0.068	-0.000	0.017	False
<b>ELM-XGB</b>	<b>0.030</b>	<b>0.001</b>	<b>0.021</b>	<b>0.038</b>	<b>True</b>
EN-MARS	0.009	0.046	0.000	0.018	True
EN-SVR	0.005	0.572	-0.004	0.013	False
<b>EN-XGB</b>	<b>0.026</b>	<b>0.001</b>	<b>0.017</b>	<b>0.035</b>	<b>True</b>
MARS-SVR	-0.004	0.670	-0.013	0.005	False
<b>MARS-XGB</b>	<b>0.017</b>	<b>0.001</b>	<b>0.008</b>	<b>0.026</b>	<b>True</b>
<b>SVR-XGB</b>	<b>0.021</b>	<b>0.001</b>	<b>0.013</b>	<b>0.030</b>	<b>True</b>

Fonte: Elaborada pelo autor (2023)

A Tabela 23 mostra de maneira ilustrativa todos os modelos híbridos gerados por metaheurísticas e modelos analisados. Um total de vinte modelos híbridos são obtidos e comparados entre si, para obter o melhor.

Da combinação dos resultados analisados separadamente nas Tabelas 21 e 22, dos testes de Tukey para comparação entre metaheurísticas e entre modelos, respectivamente, permitiu concluir que o modelo *Extreme gradient boosting* assistido por *grey wolf optimization* (GWO-XGB) é o modelo superior à todos os outros modelos desenvolvidos, em termos de desempenho.

Tabela 23 – Comparações de modelos híbridos gerados por combinações de modelos e Metaheurísticas em análise

ELM	(DE-ELM)	(PSO-ELM)	(SGA-ELM)	GWO-ELM
EN	(DE-EN)	(PSO-EN)	(SGA-EN)	GWO-EN)
MARS	(DE-MARS)	(PSO-MARS)	(SGA-MARS)	(GWO-MARS)
SVR	(DE-SVR)	(PSO-SVR)	(SGA-SVR)	(GWO-SVR)
XGB	(DE-XGB)	(PSO-XGB)	(SGA-XGB)	(GWO-XGB)
	DE	PSO	SGA	GWO

Fonte: Elaborada pelo autor (2023)

O modelo XGB já demonstrou superioridade dos seus resultados em comparação com outros modelos na resolução de desafios por meio de aprendizado de máquinas em competições em diferentes plataformas como *Kaggle* e *KDD Cup*, também tem sido utilizado em várias aplicações de ponta na indústria (NGUYEN et al., 2022), e para executar tarefas de classificação e regressão, que foram validadas em muitos casos, como previsão de vendas em lojas, previsão de comportamento do cliente, previsão de cliques em anúncios, previsão de perigo, previsão de texto da web, classificação de *malware* (ISLAM et al., 2021).

Em hidrologia, este modelo mostrou ser superior que *Random Forest (RF)* (SAHOUR et al., 2021; NGUYEN et al., 2021), *support vector machine (SVM)* (NI et al., 2020c; YU et al., 2020), *Classification and regression trees (CART)* (NGUYEN et al., 2021), Redes Neurais Artificiais (OSMAN et al., 2021), Redes Neurais recorrentes (HEINEN, 2018) em problemas de previsão simples e multipassos de vazão e, devido a sua grande performance este foi também aplicado como alternativo à previsão de cheias (VENKATESAN; MAHINDRAKAR, 2019).

É interessante observar que SVR e MARS alcançaram resultados médios competitivos, considerando as métricas analisadas. No entanto, os níveis de incerteza muito altos no ajuste dos valores extremos, que são tratados como *outliers* pelos modelos deterioraram seu desempenho. Embora os recursos de modelagem SVR e MARS tenham apresentado desempenho inferior ao XGB, a busca evolutiva desempenhou um papel essencial para encontrar os parâmetros internos adequados que levam a uma boa modelagem da vazão.

Os modelos mostraram também qualitativamente uma boa aderência ou aproximação entre os dados observados e os estimados, isto é, os modelos foram capazes de reproduzir as características da série de dados observados, como mudança de nível nos períodos críticos com vazões mais baixas e vazões mais altas, tendência, sazonalidade e qualquer outra característica oculta com excelente qualidade. Portanto, estes modelos podem fornecer subsídios na tomada de decisão com antecedência nas operações do reservatório.

Porém, este desempenho se deteriora à medida que o horizonte de previsão aumenta, isto é, horizontes menos distantes obtêm melhores resultados do que os relativamente mais distantes. Este comportamento é esperado porque o aumento do horizonte de previsão resulta em maior complexidade para a relação de entrada-saída envolvendo variáveis ambientais e vazão. Além disso, o avanço no horizonte de previsão traz mais incerteza para o valor futuro da vazão. Nesse contexto, o aumento da não-linearidade e incerteza torna as previsões com modelos de aprendizado de máquina mais desafiadoras refletido nas métricas de desempenho.

Outro aspecto observado que também deteriora o desempenho dos modelos é a existência de alguma incerteza associada ao ajuste dos valores extremos, que são vistos como *outliers* pelos mesmos, que caracterizam o comportamento caótico e alta estocasticidade da vazão (JIANG et al., 2020).

Portanto, verificou-se uma variabilidade em dados de séries temporais modelados, essa variação é muito mais alta nos picos. As estimativas dos modelos aos eventos extremos ou fluxos extremos é difícil. Contudo, ressalta-se a importância da identificação precisa de vazões extremas na tomada de decisão das operações de barragens. Devido às características geográficas do leito do rio a jusante da barragem, o vale se alarga e o rio desenvolve uma estreita planície de inundação (MCCARTNEY et al., 2018).

Como resultado, vazões extremas podem fazer com que o fluxo do rio mude resultando em inundações e desastres repentinamente. Previsões precisas de vazão extrema auxiliam na regulação segura da vazão do rio, permitindo que as demandas de produção de energia sejam atendidas. Portanto, uma análise de incerteza para avaliar os níveis de confiabilidade da previsão em relação as variáveis foi aplicada, mais detalhes podem ser encontrados na seção 4.6.

### 5.3.5.1 ANÁLISE DAS DISTRIBUIÇÕES DOS PARÂMETROS INTERNOS DOS MODELOS DE APRENDIZADO DE MÁQUINAS PROPOSTOS

Como foi observado na tabela 21 do teste de Tukey, o algoritmo GWO obteve resultados superiores em termos de performance em comparação com outros algoritmos (PSO, DE, GA). Nesta seção, procura-se verificar através de uma análise das distribuições dos parâmetros dos modelos ML a convergência das metaheurísticas analisadas. Esta análise é feita apenas para previsão  $Q_{t+7}$ , pois apresenta o maior desafio para os modelos híbridos de aprendizado de máquina.

Para tal, considera-se nesta análise apenas o modelo XGB para todas as metaheurísticas (GA, DE, PSO e GWO), uma vez que, de acordo com os resultados exibidos na tabela 22, este superou os outros modelos de aprendizado de máquina na previsão da vazão. Os parâmetros internos do modelo XGB determinados pelos algoritmos inspirados na natureza foram coletados em cada execução, e suas distribuições são mostradas nas figuras 27, 28, 29 e 30 para os algoritmos GA, DE, GWO e PSO, respectivamente. São eles: o número de estimadores  $M_{est}$ , a taxa de aprendizado  $LR$ , o parâmetro de regularização  $\lambda$  e a profundidade máxima  $m_{depth}$ .

Pode-se observar que, de um total de 20 alternativas disponíveis para o parâmetro profundidade máxima  $m_{depth}$  (1 a 20) (tabela 12, apenas quatro foram observadas nas soluções finais variando de 2 a 5 para GA e GWO e, cinco variando de 2 a 6 foram observadas para DE e PSO. Primeiros valores deste parametro variando de 2 a 4 foram os mais observados em todas as metaheurísticas nas soluções finais, sendo o  $m_{depth} = 4$  o mais frequente com 10 para GA, 11 para DE e, 13 para GWO e PSO, respectivamente do total de 30 execuções independentes. A utilização árvores de decisão com profundidades pequenas ou seja, aprendizes fracos nas soluções finais reforça a propriedade de *ensemble* do modelo XGB, que é a de utilizar aprendizes fracos para obter um preditor forte.

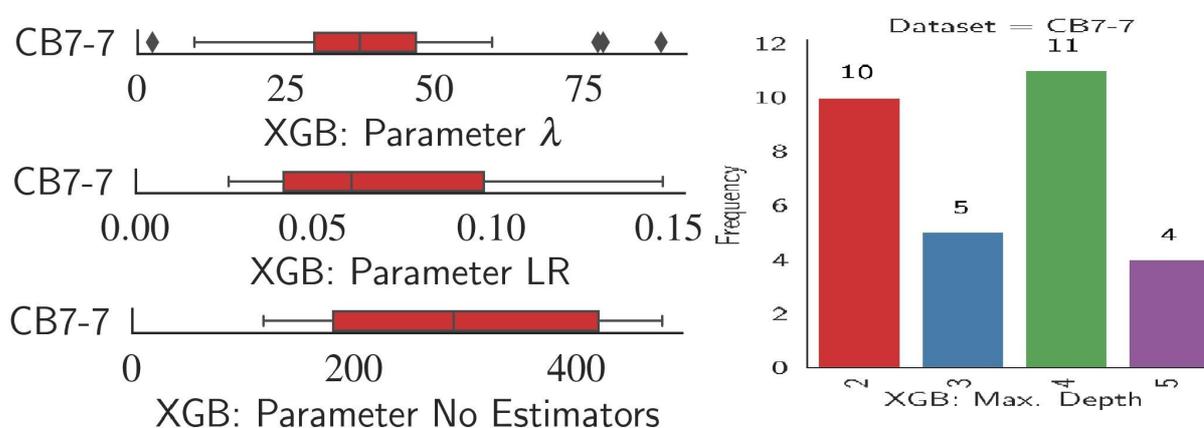
O parâmetro de regularização  $\lambda$  encontra-se no intervalo [25-50] para as metaheurísticas GA, DE e PSO e, [20-40] para GWO, com a presença de alguns *outliers* nos três primeiros algoritmos mencionados. Conforme mostrado na faixa interquartil do boxplot, o número de estimadores fracos  $M_{est}$  variou de aproximadamente 200 a 400 estimadores em todas as metaheurísticas analisadas.

A distribuição da taxa de aprendizado foi não superior a 0.10 em todas as meta-

heurísticas. Este parâmetro apresentou uma variação pequena com resultados semelhantes. O algoritmo GA teve amplitude interquartilica relativamente maior em relação aos outros.

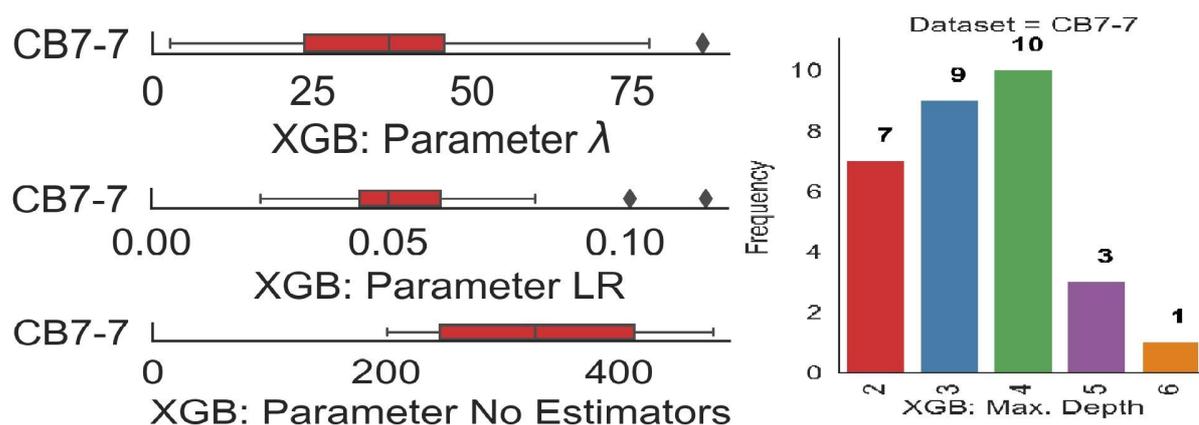
Os resultados semelhantes obtidos pelas metaheurísticas nas distribuições dos parametros dos modelo de ML, podem indicar que, estas obtiveram uma parametrização semelhante e, por conseguinte, tendem a convergir para o mesmo ótimo global, porém uma com maior precisão que a outra, sendo GWO com a metaheurística superior. No entanto, entende-se que, uma análise aprofundada julgada complexa, e que não é do escopo deste trabalho seja ainda necessária para provar esta afirmação.

Figura 27 - Distribuição dos Parâmetros de XGB com GA.



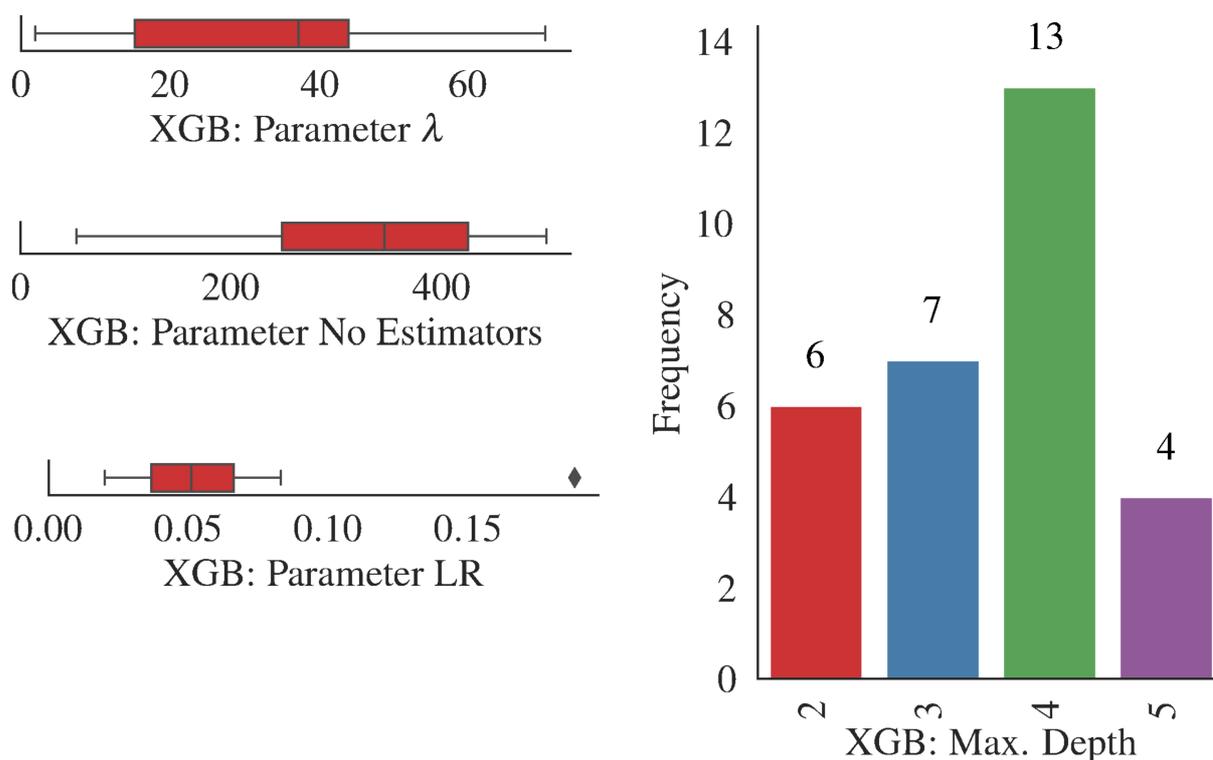
Fonte: Elaborada pelo autor (2023)

Figura 28 - Distribuição dos Parâmetros de XGB com DE.



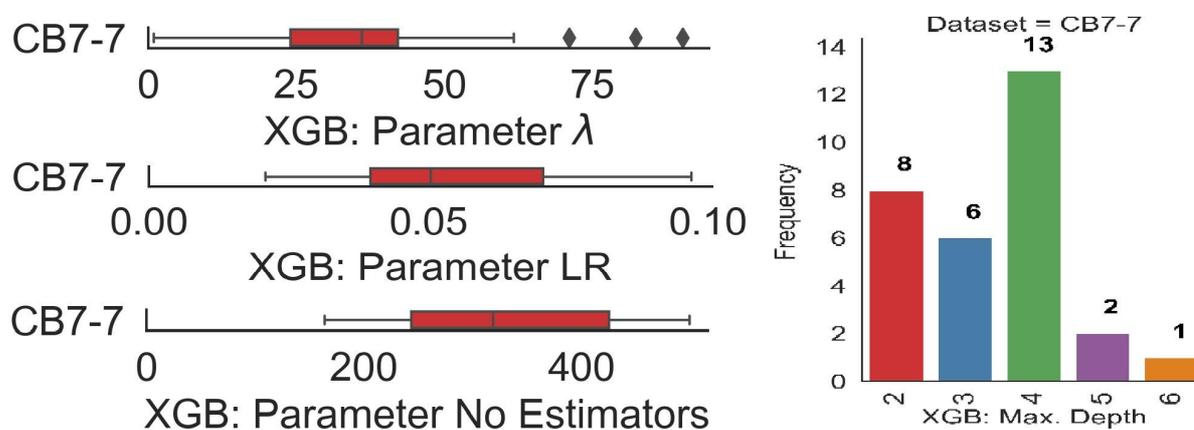
Fonte: Elaborada pelo autor (2023)

Figura 29 - Distribuição dos Parâmetros de XGB com GWO.



Fonte: Elaborada pelo autor (2023)

Figura 30 - Distribuição dos Parâmetros de XGB com PSO.



Fonte: Elaborada pelo autor (2023)

#### 5.4 RESULTADOS E DISCUSSÃO DOS MODELOS COM SELEÇÃO DE VARIÁVEIS

Muitos modelos desenvolvidos na literatura são majoritariamente para previsão um passo à frente ou simples. Porém, os resultados mostram que para as previsões um passo à frente é difícil discriminar inequivocamente a favor de um modelo em relação aos outros, dado que, os modelos têm obtido resultados satisfatórios. Para a tarefa de previsão de vários passos à frente, o efeito das componentes estocásticas de séries temporais torna-se mais significativo com o aumento do tempo de previsão, o que dificulta para identificar o

número de defasagens significativas da(s) variável(eis) que afetam a previsão do processo (REZAIE-BALF et al., 2019).

Apesar deste processo de identificação ou seleção de variáveis ser uma tarefa difícil, ressalta-se a sua importância, pois mesmo que resultados satisfatórios sejam encontrados resultantes da hibridação dos modelos de aprendizado de máquina através integração de algoritmos bioinspirados ou inspirados na natureza no ajuste de hiperparâmetros, como foi observado na seção 5.3.5; esses modelos frequentemente apresentam um custo computacional relativamente alto, requerendo uma simplificação para redução do mesmo, o que pode tornar a aprendizagem dos modelos muito mais rápida.

Por outro lado, modelos aprendizado de máquina intensivos de dados também não são recomendados na modelagem hidrológica, pois são considerados hidrológicamente não robustos (MUSHTAQ et al., 2022). Portanto, não só existe o interesse de melhorar a performance dos modelos de aprendizado de máquina, mas também estudar formas eficientes de simplificá-los sem perdê-lá. Como resultado, métodos de seleção de variáveis são vitais e examinados para fornecer aos modelos ML com dados de entrada-saída mais informativos para melhorar o desempenho, diminuir o esforço computacional e promover um melhor entendimento do processo gerador dos dados (VILLANUEVA, 2006).

Métodos de seleção ou combinação de subconjuntos de variáveis de entrada como: *Least Absolute Shrinkage and Selection Operator (LASSO)* e Redes Neurais Artificiais otimizadas por Enxame de Partículas (PSO-ANN) (ALIZADEH et al., 2020), métodos de seleção de recursos que consideram a incerteza nas variáveis de entrada (QUILTY et al., 2016), *Extreme Gradient Boosting (XGB)* (PHAM et al., 2019), método de regressão passo a passo (TONGAL; BOOIJ, 2018), método *SHapley Additive exPlanations (SHAP)* (MUSHTAQ et al., 2022), entre outros têm sido utilizados. Porém, a utilização destes para simulação da vazão e outras variáveis hidrológicas é recente, constituindo um novo desenvolvimento na área.

Quando os modelos de aprendizado de máquina foram ajustados para previsão de vazão, vários autores relataram claramente que a seleção das variáveis de entradas relevantes foi uma tarefa desafiadora e, nos principais estudos relatados, a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) foram apenas as técnicas mais relatadas para seleção de variáveis de entrada (IVS) acompanhadas pela matriz de correlação autônoma (HUSSAIN; KHAN, 2020b; TIKHAMARINE et al., 2020; HADI et al., 2019; ADNAN et al., 2021; AL-SUDANI et al., 2019b; KISI et al., 2022).

No entanto, esses métodos não são muito eficazes para o caso multivariado de análise de séries temporais (ou univariado com um número maior de atrasos), pois dado um número  $L$  de variáveis de entrada existem  $2^L - 1$  subconjuntos de entradas possíveis (PUMA-VILLANUEVA et al., 2006; SIQUEIRA et al., 2020). O papel dos métodos de seleção é escolher o subconjunto que permite a melhor previsão dos valores futuros de uma

série temporal.

Esta seção analisa a capacidade de seleção de dois métodos: *Least Absolute Shrinkage and Selection Operator (LASSO)* e *Partial Mutual Information variable selection (PMI)*.

Como foi referido na seção 5.3.5 o modelo XGB superou todos os outros em todos horizontes de previsão. Sem perder a generalidade, os métodos de seleção de variáveis LASSO e PMI foram utilizados apenas para este modelo.

O procedimento que incorpora o modelo LASSO ou PMI na abordagem XGB funciona da seguinte forma, primeiro, o LASSO ou PMI são aplicados ao conjunto de dados original e algumas variáveis são selecionadas para serem eliminadas. Em seguida, a eliminação de variáveis é feita identificando variáveis com coeficientes nulos ou removendo dependências redundantes entre todas as entradas, considerando tanto a significância da entrada quanto a dependência da entrada.

Este procedimento reduz o número de variáveis no conjunto de dados, permitindo a construção de um modelo simples. E finalmente, o XGB é ajustado a esses dados reduzidos resultando em um modelo com precisão semelhante, porém menos complexo. A filtragem é realizada com base nos valores do regularizador LASSO ou de PMI. Somente variáveis com valores do regularizador LASSO  $\alpha_{L_1}$  num intervalo definido ou de PMI acima de um limite  $C_{PMI}$  definido (tabela 24) são selecionadas como variáveis de entrada do modelo.

Essas técnicas de seleção de variáveis foram utilizadas com GWO (melhor algoritmo bioinspirado obtido). A Tabela 24 apresenta a codificação algorítmica das soluções e parâmetros candidatos associados a cada método de aprendizado de máquina. Os modelos EN, SVR, ELM, MARS e XGB foram mantidos nas mesmas condições que a codificação anterior mostrada na tabela 12. A função objetivo que é usada no problema de otimização formulado a partir da busca com GWO é a minimização do RMSE determinado pelo conjunto de treinamento (conforme representado na Figura 8), usando um procedimento de validação cruzada *5-fold walk forward*. Os experimentos foram repetidos em 30 execuções independentes com diferentes sementes aleatórias.

Tabela 24 – Codificação de soluções potenciais. A coluna IP representa o parâmetro interno na codificação GWO.

Estimator	IP	Summary	Selection/Range
EN	$\theta_1$	Penalty term, $\alpha$	$[10^{-6}, 2]$
	$\theta_2$	$L_1$ -ratio parameter, $\rho$	$[0, 1]$
ELM	$\theta_1$	No. hidden layer neurons, $L$	$[1, 500]$
	$\theta_2$	Regularization parameter, $C$	$[0.0001, 1000]$
	$\theta_3$	Activation function $G$	1: Identity; 2: Sigmoid; 3: Hyperbolic Tangent; 4: Gaussian; 5: Swish; 6: ReLU;
SVR	$\theta_1$	Loss parameter, $\varepsilon$	$[10^{-5}, 100]$
	$\theta_2$	Regularization parameter, $C$	$[1, 1000]$
	$\theta_3$	Bandwidth parameter, $\gamma$	$[0.001, 10]$
MARS	$\theta_1$	Degree of piecewise polynomials, $q$	$[0, 3]$
	$\theta_2$	Penalty factor, $\gamma$	$[1, 9]$
	$\theta_3$	Number of terms, $M$	$[1, 500]$
XGB	$\theta_1$	Learning rate, $\eta$	$[10^{-6}, 1]$
	$\theta_2$	No. weak estimators, $M_{est}$	$[10, 500]$
	$\theta_3$	Maximum depth, $m_{depth}$	$[1, 20]$
	$\theta_4$	Regularization parameter, $\lambda_{reg}$	$[0, 100]$
XGB-LASSO	$\theta_1$	Learning rate, $\eta$	$[10^{-6}, 1]$
	$\theta_2$	No. weak estimators, $M_{est}$	$[10, 500]$
	$\theta_3$	Maximum depth, $m_{depth}$	$[1, 20]$
	$\theta_4$	Regularization parameter, $\lambda_{reg}$	$[0, 100]$
	$\theta_5$	LASSO regularizer, $\alpha_{L_1}$	$[0, 100]$
XGB-PMI	$\theta_1$	Learning rate, $\eta$	$[10^{-6}, 1]$
	$\theta_2$	No. weak estimators, $M_{est}$	$[10, 500]$
	$\theta_3$	Maximum depth, $m_{depth}$	$[1, 20]$
	$\theta_4$	Regularization parameter, $\lambda_{reg}$	$[0, 100]$
	$\theta_5$	Mutual information minimal coefficient, $C_{PMI}$	$[0.5, 1]$

Fonte: Elaborada pelo autor (2023)

As estatísticas descritivas das métricas de desempenho produzidas pelos modelos de previsão propostos são exibidas na Tabela 25.

Os resultados mostram que o modelo híbrido integrado com Extreme Gradient Boosting (XGB), o modelo híbrido integrado com Extreme Gradient Boosting assistido por LASSO (XGB-LASSO) e o modelo híbrido integrado com Extreme Gradient Boosting assistido por PMI (XGB-PMI) obtiveram os melhores desempenhos para todos os horizontes de previsão. O modelo XGB supera consistentemente outros modelos em todas as métricas para horizontes  $t+1$ ,  $t+3$ , e  $t+7$ , exceto para  $t+5$ , onde o *Multivariate Adaptive Regression Splines* (MARS) produziu o menor erro médio absoluto (MAE). A Tabela 25 mostra que XGB, XGB-LASSO e XGB-PMI apresentaram desempenho semelhante nas execuções independentes. Deve-se notar que o XGB-LASSO e o XGB-PMI produzem um desempenho semelhante, mas com menos variáveis, como visto na Figura 31. Esta figura

mostra o número médio de variáveis nos modelos finais obtidos com o XGB-LASSO e XGB-PMI em função do número de dias à frente. Os resultados foram calculados ao longo de 30 execuções independentes. As barras de erro mostram o desvio padrão.

O número de variáveis permaneceu menor que 9 para todos os horizontes no modelo XGB-PMI. No entanto, quando comparado com o padrão XGB (sem seleção de variáveis de entrada) pode-se observar que houve uma redução significativa do número de variáveis de 75% para 1, 3 e 5 dias à frente e 68% para 7 dias à frente.

Considerando XGB-LASSO, a média de variáveis é de 14 variáveis para um dia à frente e 13 variáveis para três dias à frente. O número médio de variáveis aumenta para 16 com cinco dias de antecedência e cresce para 21 variáveis com sete dias de antecedência. Em comparação com o XGB sem a seleção de recursos houve uma redução no número de variáveis de 50% para 1 dia à frente, 54% para 3 dias à frente, 43% para 5 dias à frente e 25% para 7 dias à frente.

Comparando por sua vez, o PMI com o LASSO pode-se observar que o PMI tem uma melhor capacidade de reduzir variáveis quando comparado com o LASSO, com uma diferença de 25%, 21%, 32% e 43% para os horizontes 1, 3, 5 e 7, respectivamente. O PMI pode determinar as entradas que melhor caracterizam a relação de entrada/saída subjacente, este fornece uma melhor representação do processo modelado quando comparado a alternativas lineares (QUILTY et al., 2016).

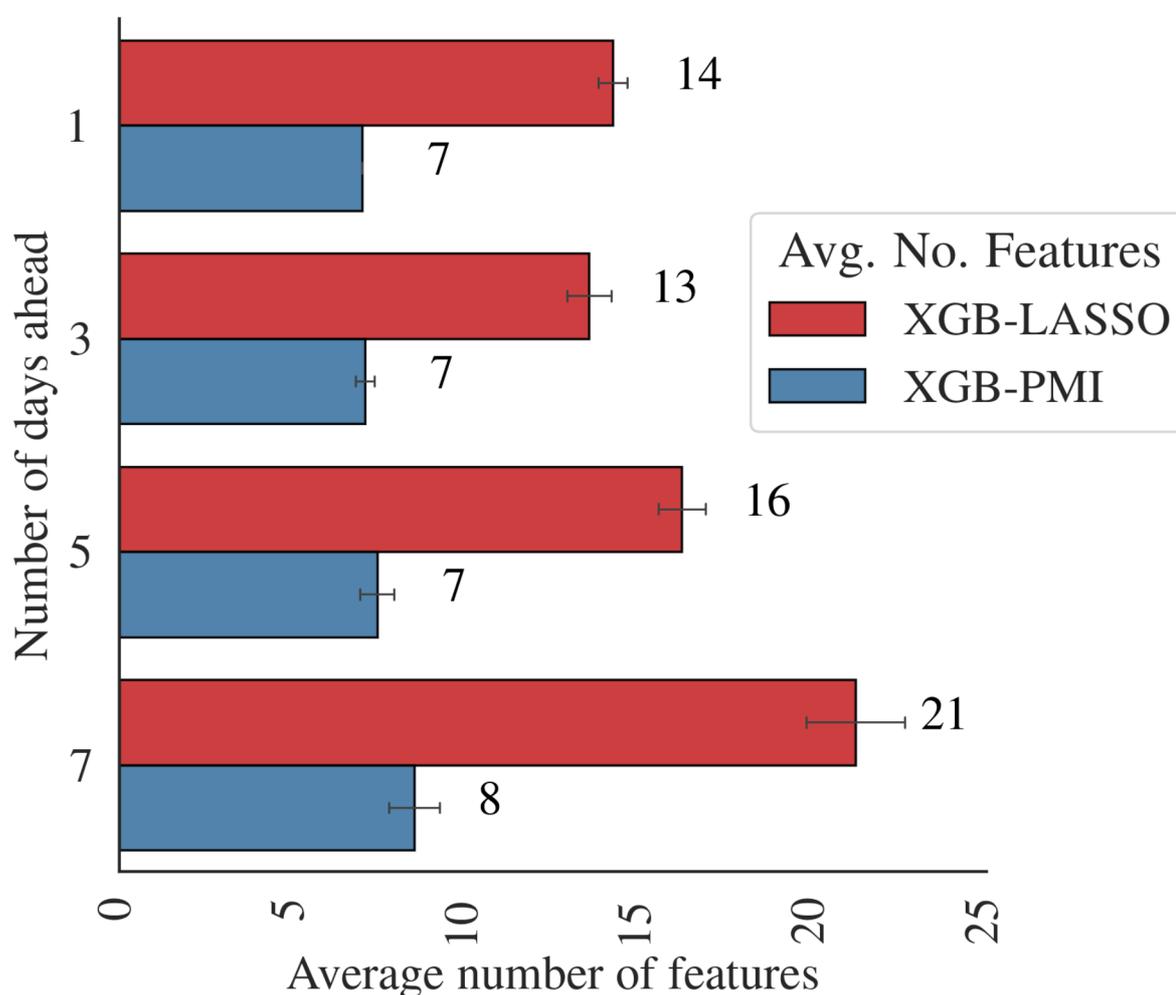
Pode-se observar que, em geral, o número médio de variáveis aumenta conforme aumenta o horizonte de previsão. Esse aumento está relacionado à necessidade de incluir mais informações para horizontes maiores, indicando maior dificuldade em representar a vazão. Como resultado, a seleção de modelos realizada pelo GWO tende a incluir mais informações, aumentando o número médio de variáveis nos modelos finais.

Tabela 25 – Métricas de média e desvio padrão. Desvios padrão entre parênteses. DA indica o número de dias antes do modelo híbrido realizar a previsão na primeira coluna. A segunda coluna exibe os modelos de aprendizado de máquina integrados à abordagem híbrida. O acrônimo métrico é mostrado na primeira linha. As colunas restantes resumem as métricas alcançadas pelo modelo. Um total de 30 execuções foram realizadas para cada entrada na Tabela

DA	Estimator	R	R <sup>2</sup>	WI	RMSE	MAE	MAPE	NSE	KGE
1	ELM	0.966 (0.002)	0.928 (0.005)	0.982 (0.001)	0.169 (0.006)	0.112 (0.002)	6.56 (0.093)	0.928 (0.005)	0.954 (0.004)
	EN	0.966 (0.002)	0.928 (0.004)	0.982 (0.00)	0.169 (0.005)	0.112 (0.001)	6.52 (0.052)	0.928 (0.004)	0.955 (0.004)
	MARS	0.975 (0.00)	0.948 (0.00)	0.987 (0.00)	0.144 (0.00)	0.099 (0.00)	5.83 (0.040)	0.948 (0.00)	0.961 (0.002)
	SVR	0.961 (0.026)	0.914 (0.059)	0.979 (0.016)	0.177 (0.050)	0.130 (0.042)	7.81 (2.72)	0.914 (0.059)	0.944 (0.027)
	XGB	<b>0.979</b> (0.00)	<b>0.958</b> (0.001)	<b>0.989</b> (0.00)	<b>0.129</b> (0.002)	<b>0.092</b> (0.002)	<b>5.54</b> (0.119)	<b>0.958</b> (0.001)	<b>0.977</b> (0.00)
	XGB-LASSO	0.978 (0.00)	0.956 (0.002)	0.989 (0.00)	0.132 (0.002)	0.094 (0.003)	5.64 (0.168)	0.956 (0.002)	0.975 (0.001)
	XGB-PMI	0.978 (0.001)	0.955 (0.003)	0.989 (0.00)	0.134 (0.004)	0.096 (0.005)	5.71 (0.270)	0.955 (0.003)	0.975 (0.002)
3	ELM	0.965 (0.00)	0.922 (0.00)	0.981 (0.00)	0.175 (0.00)	0.116 (0.00)	6.67 (0.009)	0.922 (0.00)	0.943 (0.00)
	EN	0.964 (0.00)	0.922 (0.002)	0.981 (0.00)	0.175 (0.002)	0.115 (0.00)	6.63 (0.011)	0.922 (0.002)	0.944 (0.002)
	MARS	0.971 (0.003)	0.938 (0.005)	0.985 (0.001)	0.157 (0.007)	0.096 (0.00)	5.66 (0.026)	0.938 (0.005)	0.955 (0.004)
	SVR	0.969 (0.002)	0.929 (0.006)	0.983 (0.001)	0.167 (0.006)	0.119 (0.003)	6.99 (0.216)	0.929 (0.006)	0.939 (0.010)
	XGB	<b>0.981</b> (0.00)	0.961 (0.001)	<b>0.990</b> (0.00)	0.124 (0.002)	0.095 (0.001)	5.72 (0.069)	0.961 (0.001)	0.975 (0.002)
	XGB-LASSO	0.981 (0.00)	0.961 (0.001)	0.990 (0.00)	0.124 (0.002)	0.095 (0.002)	5.72 (0.092)	0.961 (0.001)	0.974 (0.003)
	XGB-PMI	0.981 (0.00)	<b>0.962</b> (0.001)	0.990 (0.00)	<b>0.123</b> (0.002)	<b>0.093</b> (0.001)	<b>5.61</b> (0.064)	<b>0.962</b> (0.001)	<b>0.979</b> (0.00)
5	ELM	0.946 (0.00)	0.878 (0.002)	0.971 (0.00)	0.220 (0.002)	0.146 (0.001)	8.35 (0.106)	0.878 (0.002)	0.916 (0.002)
	EN	0.945 (0.00)	0.877 (0.00)	0.970 (0.00)	0.221 (0.00)	0.147 (0.00)	8.36 (0.004)	0.877 (0.00)	0.916 (0.00)
	MARS	0.956 (0.00)	0.904 (0.00)	0.977 (0.00)	0.194 (0.00)	<b>0.134</b> (0.00)	<b>7.83</b> (0.008)	0.904 (0.00)	0.936 (0.00)
	SVR	0.934 (0.036)	0.840 (0.084)	0.962 (0.020)	0.245 (0.056)	0.174 (0.047)	10.26 (3.14)	0.840 (0.084)	0.884 (0.032)
	XGB	<b>0.961</b> (0.00)	<b>0.917</b> (0.003)	<b>0.979</b> (0.00)	<b>0.181</b> (0.003)	0.136 (0.002)	8.12 (0.127)	<b>0.917</b> (0.003)	0.949 (0.004)
	XGB-LASSO	0.960 (0.001)	0.917 (0.003)	0.979 (0.00)	0.181 (0.003)	0.135 (0.002)	8.10 (0.071)	0.917 (0.003)	0.951 (0.003)
	XGB-PMI	0.957 (0.00)	0.913 (0.002)	0.978 (0.00)	0.186 (0.002)	0.138 (0.002)	8.20 (0.124)	0.913 (0.002)	<b>0.954</b> (0.002)
7	ELM	0.903 (0.001)	0.785 (0.005)	0.948 (0.00)	0.291 (0.003)	0.188 (0.002)	10.74 (0.142)	0.785 (0.005)	0.881 (0.003)
	EN	0.902 (0.00)	0.785 (0.00)	0.947 (0.00)	0.291 (0.00)	0.188 (0.00)	<b>10.72</b> (0.008)	0.785 (0.00)	0.882 (0.00)
	MARS	0.902 (0.023)	0.772 (0.067)	0.946 (0.014)	0.298 (0.034)	0.190 (0.003)	10.98 (0.129)	0.772 (0.067)	0.866 (0.028)
	SVR	0.900 (0.027)	0.771 (0.063)	0.945 (0.016)	0.298 (0.035)	0.200 (0.033)	11.63 (2.17)	0.771 (0.063)	0.867 (0.027)
	XGB	<b>0.923</b> (0.004)	0.829 (0.012)	<b>0.959</b> (0.003)	0.260 (0.009)	0.186 (0.004)	10.93 (0.180)	0.829 (0.012)	0.899 (0.009)
	XGB-LASSO	0.923 (0.005)	<b>0.830</b> (0.014)	0.959 (0.003)	<b>0.258</b> (0.011)	<b>0.185</b> (0.005)	10.91 (0.240)	<b>0.830</b> (0.014)	0.902 (0.011)
	XGB-PMI	0.918 (0.003)	0.826 (0.010)	0.957 (0.002)	0.262 (0.008)	0.188 (0.003)	11.08 (0.122)	0.826 (0.010)	<b>0.907</b> (0.008)

Fonte: Elaborada pelo autor (2023)

Figura 31 - Número médio de variáveis para XGB-LASSO e XGB-PMI. As barras de erro mostram o desvio padrão.



Fonte: Elaborada pelo autor (2023)

A Tabela 26 mostra as variáveis selecionadas para os melhores modelos XGB-LASSO e XGB-PMI com 1, 3, 5 e 7 dias de antecedência.

A primeira coluna indica o modelo, a segunda o horizonte de previsão, a terceira o número de variáveis selecionadas e a última coluna mostra as variáveis selecionadas pelo LASSO e PMI. Conforme o observado na terceira coluna, o número de variáveis nos modelos para cada horizonte de previsão é consistente com o número médio de variáveis mostrado na Figura 31 para todos os modelos. Para  $Q_{t+1}$  e  $Q_{t+3}$ , o modelo selecionou apenas informações de vazão e umidade. Para ambos os casos, todos os antecedentes das variáveis vazão e umidade foram selecionados para XGB-LASSO. À medida que o horizonte avança e a complexidade da previsão aumenta, mais variáveis foram selecionadas para construir o modelo XGB-LASSO.

Para o horizonte de 5 dias, a abordagem de seleção adicionou informações sobre a precipitação nas defasagens  $(t - 3)$  e  $(t - 6)$ . Para o horizonte de 7 dias foram incluídas as

informações pluviométricas de todas as defasagens. Além disso, informações de evaporação em  $(t - 1)$  e  $(t - 2)$  foram selecionadas. Pode-se destacar que para os horizontes 5 e 7 dias à frente, o modelo selecionou gradativamente variáveis do ciclo hidrológico para gerar estimativas precisas de vazão.

Para XGB-PMI, o modelo selecionou apenas informações de vazão e Umidade, onde para  $t + 1$  e  $t + 3$  foi considerada apenas a vazão e, para  $t + 5$  e  $t + 7$  além da vazão que é considerada nos dois horizontes de previsão foram adicionadas informações de umidade em  $(t - 2)$  para  $t + 5$  e, em  $(t - 1)$  e  $(t - 4)$  para  $t + 7$ , respectivamente.

Tabela 26 – Variáveis selecionadas para os melhores modelos XGB-LASSO e XGB-PMI para diferentes horizontes de previsão. A vazão natural é indicada por  $Q$ , a precipitação por  $R$ ,  $H$  é a umidade relativa e  $E$  é a evaporação.

Model	Horizon	No. Var.	Variables
XGB-LASSO	$Q_{t+1}$	14	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, H_t, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}, H_{t-5}, H_{t-6}$
	$Q_{t+3}$	13	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, H_t, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}, H_{t-6}$
	$Q_{t+5}$	16	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, R_{t-3}, R_{t-6}, H_t, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-4}, H_{t-5}, H_{t-6}$
	$Q_{t+7}$	22	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-5}, R_{t-6}, E_t, E_{t-1}, E_{t-2}, H_t, H_{t-1}, H_{t-2}, H_{t-3}, H_{t-5}, H_{t-6}$
XGB-PMI	$Q_{t+1}$	7	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}$
	$Q_{t+3}$	7	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}$
	$Q_{t+5}$	8	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, H_{t-2}$
	$Q_{t+7}$	9	$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, H_{t-1}, H_{t-4}$

Fonte: Elaborada pelo autor (2023)

Como ficou claro nos resultados, incluir um número alto de entradas não garante melhor precisão de previsão, porque o aumento do número de entradas pode aumentar a variação e adicionar mais complexidade ao modelo implementado, e isso deteriora a precisão do modelo na previsão.

A outra questão importante que vale a pena mencionar é que a combinação de entrada recomendada pela análise de correlação (ACF, PACF e CCFs) não deve ser diretamente aplicada aos métodos de aprendizado de máquinas. A utilização de métodos eficientes de seleção de variáveis é recomendada, por que fornecem recursos que são preditores mais apropriados para simular e prever o fluxo de água e, também permitem capturar relações ocultas que não foram detectadas utilizando a análise de correlação.

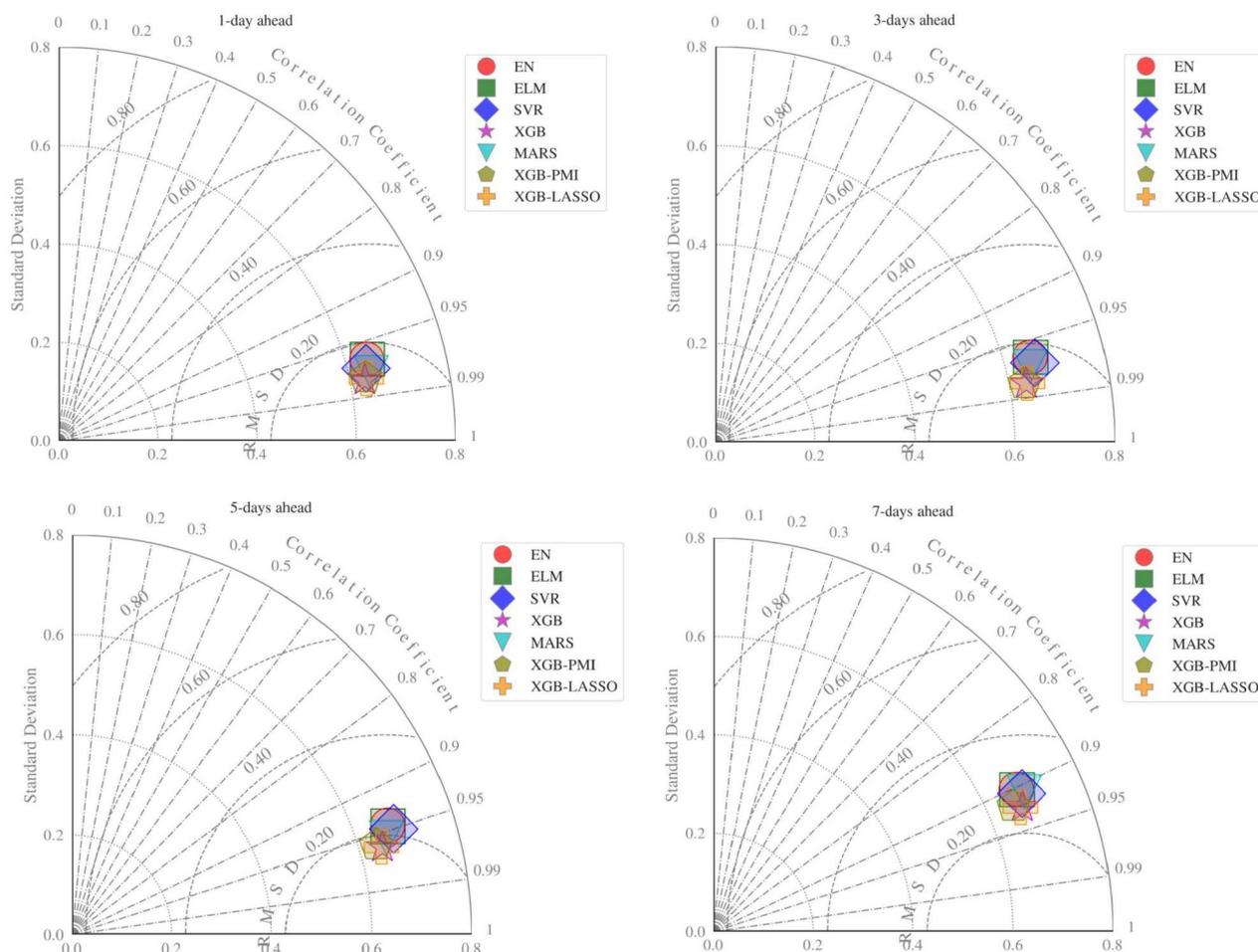
Portanto, pode-se observar que, alguns efeitos das variáveis hidrometeorológicas ( precipitação, Evaporação e umidade relativa) anteriormente considerados como não significativos utilizando a análise de correlação, foram adicionados como preditores sobre a variável de saída (vazão) usando métodos de seleção de entradas. A vazão seguida da umidade relativa são as principais variáveis selecionadas, no entanto, outras variáveis são adicionadas em horizontes distantes.

O PMI apresentou maior capacidade de seleção de variáveis que o LASSO. A vantagem deste modelo de seleção frente aos lineares é que é baseado apenas nas distribuições probabilidades dentro dos dados, não faz hipótese quanto à estrutura da dependência entre as variáveis e apresenta insensibilidade ao ruído e transformações de dados (MAY et al., 2011).

Mas, apesar deste resultado, a seleção realizada pelo modelo LASSO é ainda considerada eficiente, porque este modelo já também mostrou ser superior em relação a outros modelos de seleção como Redes neurais artificiais (ALIZADEH et al., 2020) e regressão passo a passo (clássica) (HASTIE et al., 2009a).

A Figura 32 exibe o diagrama de Taylor das previsões de vários passos à frente para o conjunto de teste ( $t + 1$ ,  $t + 3$ ,  $t + 5$  e  $t + 7$ ). O diagrama de Taylor é um diagrama que fornece uma representação visual de quão bem as previsões e observações se correspondem, com base na Raiz Quadrada da Média do Quadrado das Diferenças Centralizadas(RMSD), no coeficiente de correlação ( $R$ ) e no desvio padrão ( $\sigma$ ) (PEREIRA et al., 2014; YASEEN et al., 2020). Como demonstrado, os modelos geralmente geram resultados próximos aos dados observados, com valores maiores de  $R$ , menores de  $RMSD$  em todos os horizontes de previsão. Para os horizontes  $t + 1$  e  $t + 3$  os modelos obtiveram melhores níveis de concordância com os dados observados. Em contraste, para  $t + 5$  e  $t + 7$ , a qualidade da concordância com os dados observados é decrescente. Portanto, horizontes menos distantes obtêm melhores resultados do que os relativamente mais distantes. Além disso, os modelos baseados em XGB obtiveram concordância relativamente maior com os dados observados do que outros modelos.

Figura 32 - Diagramas de Taylor ilustrando modelos com várias etapas à frente:  $t + 1$ ,  $t + 3$ ,  $t + 5$  e  $t + 7$



Fonte: Elaborada pelo autor (2023)

A Tabela 27 apresenta estudos recentes que utilizaram abordagens híbridas para previsão de vazão de rios, aplicando métodos de aprendizado de máquinas com diferentes metaheurísticas para otimizar os parâmetros. A tabela visa validar os resultados obtidos nesta pesquisa com os resultados da literatura.

Considerando que este estudo é o primeiro a ser desenvolvido na área de modelagem e previsão utilizando modelos de inteligência computacional ou de aprendizado de máquinas sobre a bacia hidrográfica do rio Zambeze, a comparação foi feita com estudos similares em rios de outras regiões do mundo.

A comparação mostrou que o modelo aqui desenvolvido apresenta resultados competitivos com estratégias recentes aplicadas a problemas de previsão de vazão em rios de diferentes regiões no mundo, com diferentes condições hidroclimáticas e geomorfológicas.

Tabela 27 – Resumo de métodos híbridos computacionais para previsão de vazões fluviais.

Referência	Estudo de Caso	Método	RMSE	R	R <sup>2</sup>
(YASEEN et al., 2017)	Pahang	Adaptive Neuro-Fuzzy Inference Systems with Firefly Optimization Algorithm (ANFIS-FFA)	0.984	1.000	–
(ADNAN et al., 2019b)	Fujiang	Particle Swarm Optimization(PSO) with an Adaptive Neuro-Fuzzy Inference System (ANFIS-PSO)	163.0	–	–
(AL-SUDANI et al., 2019a)	Tigris	Differential Evolution model with Multivariate Adaptive Regression Spline (MARS-DE)	46.64	–	–
(TIKHAMARINE et al., 2020)	Nile	Support Vector Regression Based on Grey Wolf Optimizer (SVR-GWO)	2.0570	0.9363	–
(MALIK et al., 2020)	Ramganga	Support Vector Regression with Harris Hawks Optimization (SVR-HHO)	181.122	–	0.7764
(TIKHAMARINE et al., 2019)	Ain Bedra, Fer-matou	Wavelet Support Vector Regression with Grey Wolf Optimizer (WSVR-GWO)	0.1562	0.982	–
(ADNAN et al., 2021)	Mangla	Extreme Machine Learning with Particle Swarm Optimization and Grey Wolf Optimization (ELM-PSOGWO)	55.14	–	0.925
(DIFI et al., 2022)	Algeria	Extreme learning machine optimized by Bat algorithm(BAT-ELM)	–	–	0.927-0.973
(IKRAM et al., 2022)	Indus-Pakistan	Support vector regression optimized by Covariance Matrix Adaptation Evolution Strategy (CMAES-SVR)	0.266	–	–
(HAZNEDAR; KILINC, 2022)	Seyhan-Turkey	Adaptive neuro-fuzzy inference system with a genetic algorithm (GA-ANFIS)	–	–	0.9336
(AHMED et al., 2021)	Aswan Dam (HAD)-Nile	Multi-layer Perceptron with Nuclear Reaction Optimization(NRO-MLP)	2.35	0.924	–
(ADNAN et al., 2022)	–	Support vector regression with the modified mayfly optimization algorithm (SAMOA-SVR)	0.147	–	–

*Continua na próxima página*

Tabela 27 – Continuação da tabela

Referência	Estudo de Caso	Método	RMSE	R	R <sup>2</sup>
(KILINC, 2022)	Orontes-Turkey	Particle Swarm Optimization and Long Short-Term Memory Model (PSO-LSTM)	0.9969	–	0.9515
(HADI et al., 2019)	Goksu-Himmeti-Turkey	Extreme Learning Machine with Extreme gradient boosting feature selection model (XGBELM)	10.556	–	–
(LATIFOĞLU, 2022)	Skookumchuck-Canadá	Correlation based Feature Selection and Random Forest Method (CFS-RF)	45.6941	0.9307	0.8661
(SAMANATARAY; SAHOO, 2021)	Barak-India	Particle Swarm Optimization with an Adaptive Neuro-Fuzzy Inference System (ANFIS-PSO)	5.887	–	0.9668
(KHOSRAVI et al., 2022)	Korkorsar-Iran	Convolutional Neural Network with BAT metaheuristic algorithm (CNN-BAT)	–	–	0.92
(REN et al., 2020)	United States	K-nearest neighbor with Partial linear Correlation Feature Selection Method (KNN-PCI)	0.4805	0.922	–
(NGUYEN, 2022)	Khuc-Vietnam	Recurrent Gate Unit with gray wolf optimizer(GWO-GRU)	–	–	0.883
(MUSHTAQ et al., 2022)	Pakistan	Extreme Gradient Boosting with SHapley Additive exPlanations(XGB-SHAP)	–	–	0.90
(YASEEN et al., 2020)	Tigris-Iraq	Extreme Learning Machine Model with Salp Swarm Algorithm(SSA-ELM)	80.592	0.81	–
(KILINC, 2022)	Seyhan -Turkey	grey wolf algorithm (GWO)-based gated recurrent unit (GRU) (GWO-GRU)	–	–	0.9477
(SOUZA et al., 2022)	Zambeze-Mozambique	Particle Swarm Optimization(PSO) with Group Modeling Data Handling(GMDH) (PSO-GMDH)	0.733	–	0.80
Pesquisa Desenvolvida	Zambezi	Grey Wolf Optimization and Extreme Gradient Booting with Feature selection model	0.124	0.981	0.961

*Fim da tabela*

Fonte: Elaborada pelo autor (2023)

### 5.4.1 DISTRIBUIÇÃO DOS PARÂMETROS INTERNOS PARA O MÉTODO XGB

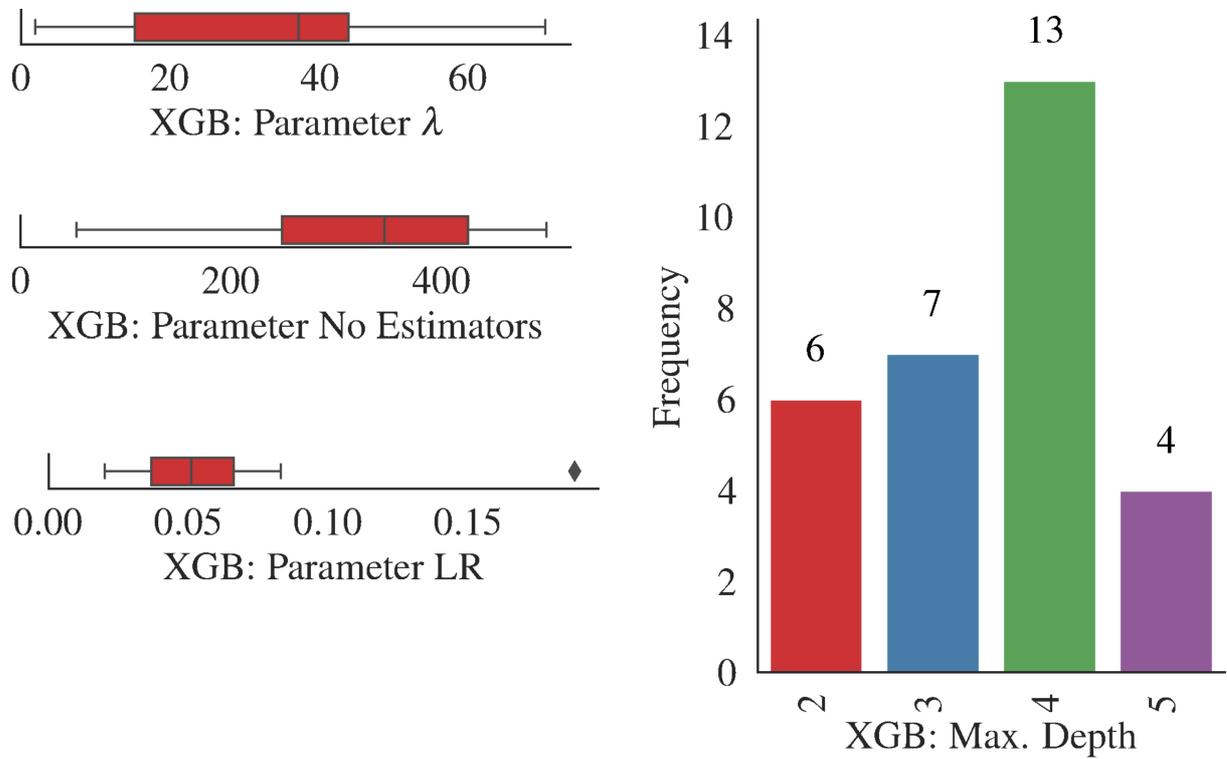
Conforme mostrado na Tabela 25, XGB, XGB-LASSO e XGB-PMI superam os outros métodos de aprendizado de máquina na previsão dos fluxos. Esta seção faz a análise da distribuição dos parâmetros de XGB, XGB-LASSO e XGB-PMI, restringindo-se apenas na análise de parâmetros da previsão  $Q_{t+7}$ . Os parâmetros internos encontrados pelo algoritmo GWO foram reunidos em cada execução, e suas distribuições são mostradas nas Figuras 33, 34 e 35 para XGB, XGB-LASSO e XGB-PMI, respectivamente. Figura 33 exhibe a distribuição de parâmetros XGB representando o número de estimadores  $M_{est}$ , a taxa de aprendizado  $LR$ , o parâmetro de regularização  $\lambda$  e a profundidade máxima  $m_{depth}$ . A partir desta figura, pode-se observar que apenas quatro foram observadas nas soluções finais variando de 2 a 5. De acordo com a Tabela 24, a profundidade máxima varia de 1 a 20, de um total de 20 alternativas estavam disponíveis. Conforme mostrado na Figura 33, o  $m_{depth} = 4$  foi selecionado em 13 das 30 execuções, enquanto  $m_{depth} = 2$  aparecem nas soluções finais em seis execuções,  $m_{depth} = 3$  em sete execuções e  $m_{depth} = 5$  em quatro execuções. Esse resultado reforça a característica de reforço do conjunto de aprendizes fracos: árvores de decisão com profundidades pequenas que combinadas geram um preditor preciso. O ajuste do parâmetro de regularização  $\lambda$  encontra-se no intervalo [20-40], e existem alguns *outliers*. No entanto, para todas as execuções  $\lambda < 60$ . O número de estimadores fracos  $M_{est}$  variou de aproximadamente 200 a 400 estimadores, conforme demonstrado na faixa interquartil do boxplot. Há uma compensação entre a profundidade de aprendizes fracos (árvores de decisão) e o número de estimadores empregados na abordagem aditiva de aumento de gradiente. À medida que a profundidade do aprendiz individual diminui, a quantidade de estimadores aumenta para manter a precisão do modelo. A distribuição da taxa de aprendizado mostra que o valor mediano do parâmetro para a taxa de aprendizado (LR) é 0.05, apresentando uma pequena variação conforme apresentado pelos valores interquartis.

A Figura 34 exhibe a distribuição dos parâmetros XGB-LASSO. A distribuição dos parâmetros do XGB-LASSO foi semelhante à distribuição do XGB para o parâmetro de regularização e o número de estimadores. No entanto, pode-se observar que os valores da taxa de aprendizado foram menores para o modelo XGB-LASSO devido ao menor número de variáveis utilizadas para representar o problema. O parâmetro de regularização L1 ( $\alpha_{L_1}$ ) foi ajustado ao menor valor possível, permitindo ao LASSO selecionar mais variáveis, o que se reflete na média de 22 variáveis observada sete dias à frente na Figura 31. Por outro lado, o modelo XGB-LASSO utilizou árvores de decisão mais profundas. Modelos com profundidade 4 ou 5 foram encontrados 22 de 30 vezes ao longo das execuções independentes.

A Figura 35 mostra a distribuição de parâmetros para o modelo XGB-PMI, que também era análogo aos modelos anteriores baseados em XGB. O parâmetro referente

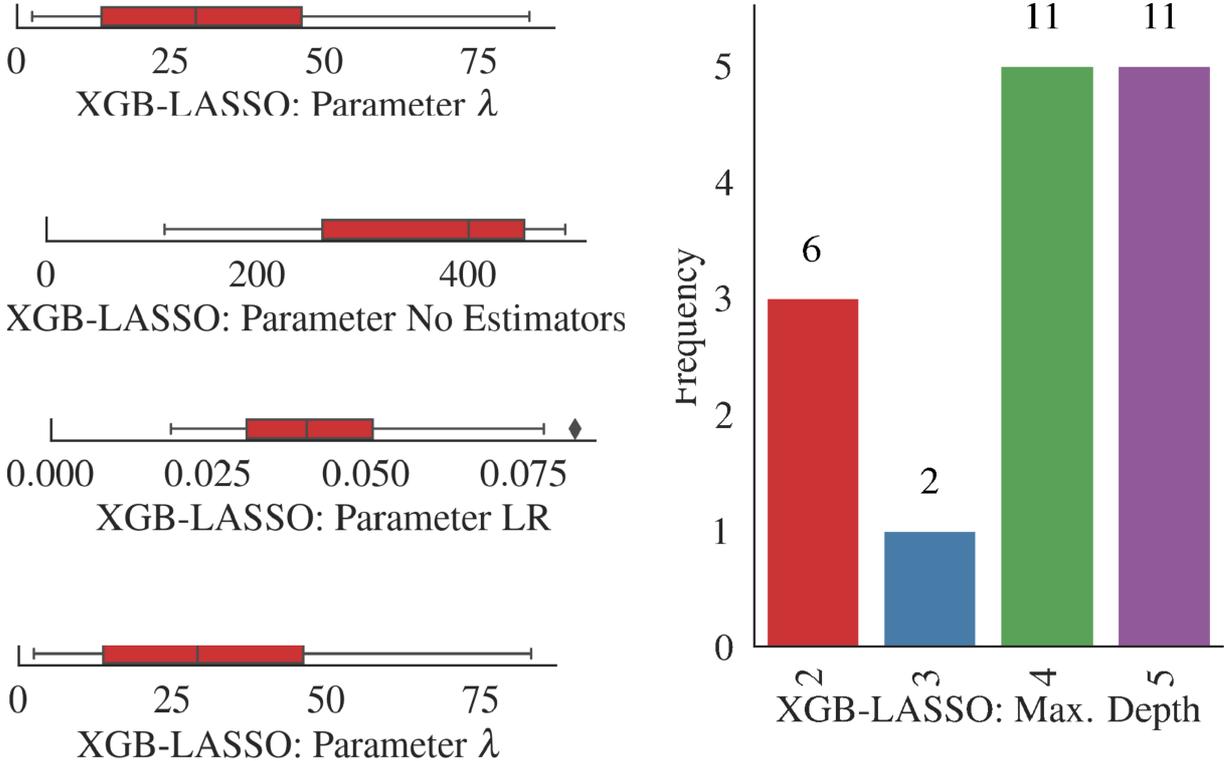
ao coeficiente de informação mútua  $C_{PMI}$  foi ajustado gulosamente, ou seja, selecionando apenas as variáveis com mais informação para a variável resposta e penalizando as com menos informação o que levou a uma redução significativa das variáveis.

Figura 33 - Distribuição de parâmetros do XGB em 30 execuções.



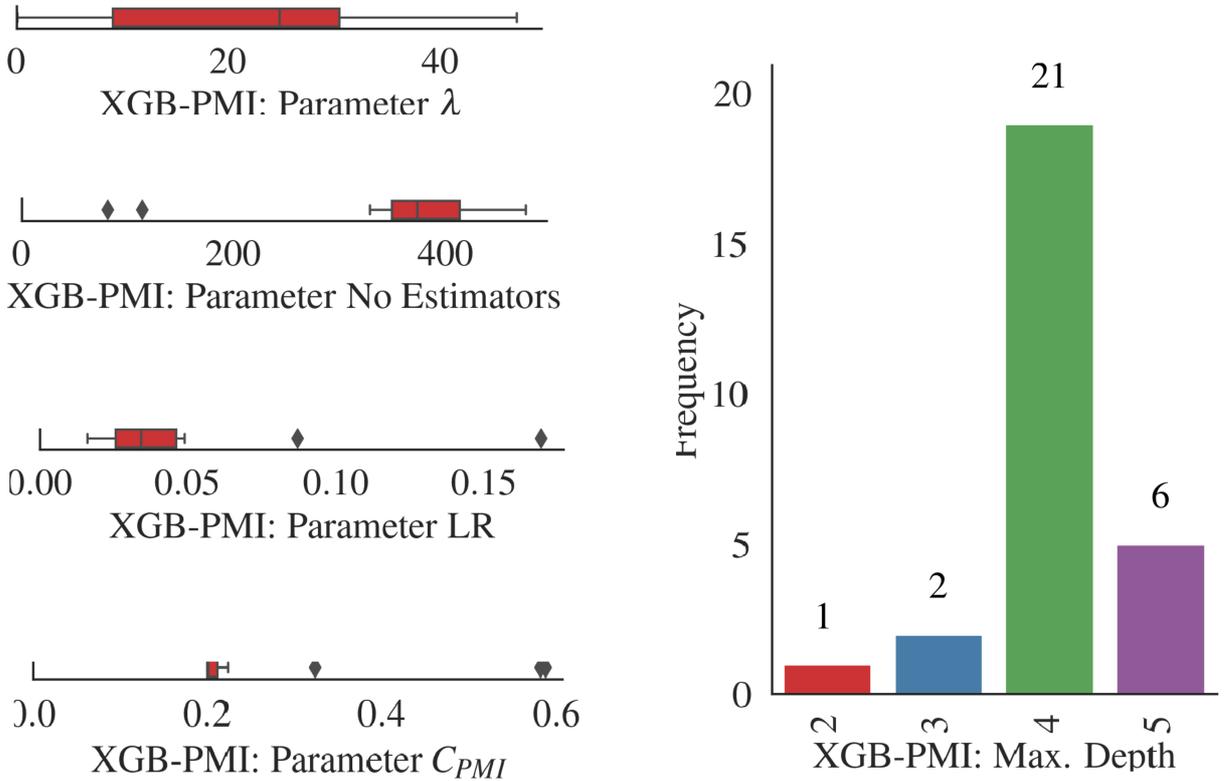
Fonte: Elaborada pelo autor (2023)

Figura 34 - Distribuição de parâmetros do XGB-LASSO em 30 execuções.



Fonte: Elaborada pelo autor (2023)

Figura 35 - Distribuição de parâmetros do XGB-PMI em 30 execuções.



Fonte: Elaborada pelo autor (2023)

## 5.5 ANÁLISE DE INCERTEZA

Como se sabe a utilidade de um modelo em geral, e em hidrologia em particular, não só se baseia na precisão das suas previsões, como também, na confiabilidade das mesmas. Nessa perspectiva, esta seção apresenta uma análise de incerteza das previsões. Dois métodos são utilizados, métodos analíticos ou estatísticos e o método de reamostragem de Montecarlo, detalhadamente abordados na seção 4.6.

A Tabela 28 apresenta os resultados do estudo de incerteza utilizando métodos analíticos ou estatísticos. Esta mostra o horizonte de previsão na primeira coluna, os modelos em análise na segunda, os erros médios de previsão dos modelos propostos na terceira, a amplitude da banda de incerteza na quarta, e por último o intervalo de confiança do erro de previsão à 95%.

Os resultados mostram que todos os modelos propostos tiveram suas médias de erro de predição como valores negativos em todos horizontes de previsão, o que significa que os modelos de predição subestimaram os valores observados, porém, com valores médios de erro muito pequenos.

O modelo XGB e suas variações com seleção de de variáveis de entrada, XGB-LASSO e XGB-PMI apresentaram menores médias de erro em termos absolutos, menores amplitudes de erro de incerteza ou intervalos de confiança de menor amplitude, em comparação com outros modelos, em todos os horizontes de previsão em análise, assegurando ainda mais confiança nas suas previsões. Ademais, comparando estes modelos entre si, observa-se que os modelos com seleção de variáveis, XGB-LASSO e XGB-PMI apresentam relativamente melhores resultados de estimativas de erro de previsão em relação ao método XGB sem seleção em todos horizontes de previsão, utilizando um número reduzido de variáveis de entradas. Isto reforça a afirmação de que, a utilização de um número elevado de variáveis de entrada nos modelos não garante maior precisão das previsões, pelo contrário, aumenta a incerteza sobre as mesmas. Portanto, a utilização dos métodos de seleção propostos neste trabalho ou outros é indispensável para identificação de variáveis de maior poder preditivo sobre a variável resposta.

Porém, apesar de que os modelos obtiveram estimativas de erros de previsão baixos, mesmo em horizontes de previsão relativamente distantes, essas estimativas aumentam com o aumento do horizonte de previsão, indicando ainda maiores desafios de fazer previsões para horizontes mais distantes.

Tabela 28 – Estimativas de incerteza para modelos de aprendizado de máquina.

DA	Modelo	Média do Erro de Previsão	Amplitude do Erro de Incerteza	Intervalo de Confiança de erro à 95%
1	ELM	-0.008	$\pm 0.036$	0.867 to 1.20
	EN	-0.008	$\pm 0.036$	0.867 to 1.20
	MARS	-0.004	$\pm 0.032$	0.872 to 1.17
	SVR	-0.007	$\pm 0.035$	0.868 to 1.19
	<b>XGB</b>	<b>-0.003</b>	<b><math>\pm 0.031</math></b>	<b>0.875 to 1.16</b>
	<b>XGB-LASSO</b>	<b>-0.003</b>	<b><math>\pm 0.031</math></b>	<b>0.874 to 1.16</b>
	<b>XGB-PMI</b>	<b>-0.001</b>	<b><math>\pm 0.031</math></b>	<b>0.872 to 1.15</b>
3	ELM	-0.010	$\pm 0.035$	0.873 to 1.20
	EN	-0.010	$\pm 0.035$	0.872 to 1.20
	MARS	-0.006	$\pm 0.031$	0.883 to 1.16
	SVR	-0.007	$\pm 0.036$	0.862 to 1.20
	<b>XGB</b>	<b>-0.005</b>	<b><math>\pm 0.030</math></b>	<b>0.885 to 1.16</b>
	<b>XGB-LASSO</b>	<b>-0.003</b>	<b><math>\pm 0.030</math></b>	<b>0.879 to 1.15</b>
	<b>XGB-PMI</b>	<b>-0.004</b>	<b><math>\pm 0.031</math></b>	<b>0.878 to 1.16</b>
5	ELM	-0.010	$\pm 0.044$	0.839 to 1.25
	EN	-0.009	$\pm 0.044$	0.837 to 1.25
	MARS	-0.005	$\pm 0.042$	0.839 to 1.22
	SVR	-0.011	$\pm 0.045$	0.837 to 1.25
	<b>XGB</b>	<b>-0.006</b>	<b><math>\pm 0.042</math></b>	<b>0.838 to 1.23</b>
	<b>XGB-LASSO</b>	<b>-0.006</b>	<b><math>\pm 0.043</math></b>	<b>0.836 to 1.23</b>
	<b>XGB-PMI</b>	<b>-0.004</b>	<b><math>\pm 0.043</math></b>	<b>0.831 to 1.22</b>
7	ELM	-0.008	$\pm 0.058$	0.786 to 1.32
	EN	-0.008	$\pm 0.058$	0.784 to 1.32
	MARS	-0.011	$\pm 0.060$	0.782 to 1.34
	SVR	-0.006	$\pm 0.060$	0.775 to 1.33
	<b>XGB</b>	<b>-0.008</b>	<b><math>\pm 0.056</math></b>	<b>0.789 to 1.31</b>
	<b>XGB-LASSO</b>	<b>-0.007</b>	<b><math>\pm 0.056</math></b>	<b>0.788 to 1.31</b>
	<b>XGB-PMI</b>	<b>-0.006</b>	<b><math>\pm 0.058</math></b>	<b>0.779 to 1.32</b>

Fonte: Elaborada pelo autor (2023)

Os resultados da análise de incerteza para o  $Q_t$  usando os modelos propostos calculados com base no método de Monte Carlo descrito ainda na seção 4.6 são resumidos através de medidas estatísticas descritivas robustas das distribuições empíricas geradas em cada horizonte de previsão, Mediana, o Desvio médio absoluto e a porcentagem de incerteza, e são mostrados na Tabela 29.

Nos experimentos conduzidos neste estudo, as variações dos parâmetros de entrada foram modeladas usando uma distribuição uniforme com os limites inferior e superior iguais ao menor e maior valores para cada parâmetro, no conjunto de teste, conforme apresentado na Tabela 15.

A análise de incerteza mostra que todos os modelos apresentaram baixos níveis de incertezas, com a porcentagem inferior a 36% em todos horizontes de previsão, exceto o modelo MARS para 1, 5 e 7 passos a frente que apresentou valores altos de incerteza. No entanto, pode-se observar que o modelo SVR apresentou a menor porcentagem de incerteza quando comparado com as outras abordagens.

Exceto para o modelo XGB-PMI 7 passos a frente, as incertezas obtidas em modelos

baseados em XGB (XGB-LASSO e XGB-PMI) são baixas e parecidas na comparação entre si. Além disso, os modelos baseados em XGB permanecem com níveis de incertezas baixos, utilizando um número reduzido de variáveis de entrada.

Tabela 29 – Análise de incerteza para o conjunto de teste usando Simulação de Monte Carlo (MCS).

DA	Modelo	No. Variáveis	Mediana	MAD	Incerteza %
1	ELM	28	3.855	0.809	20.990
	EN	28	3.857	0.801	20.774
	MARS	28	3.388	1.264	37.303
	<b>SVR</b>	<b>28</b>	<b>2.889</b>	<b>0.415</b>	<b>14.373</b>
	<b>XGB</b>	<b>28</b>	<b>3.186</b>	<b>0.833</b>	<b>26.150</b>
	<b>XGB-LASSO</b>	<b>14</b>	<b>3.117</b>	<b>0.781</b>	<b>25.062</b>
	<b>XGB-PMI</b>	<b>7</b>	<b>3.382</b>	<b>0.862</b>	<b>25.478</b>
3	ELM	28	4.097	0.790	19.278
	EN	28	4.106	0.787	19.177
	MARS	28	3.610	1.230	34.073
	<b>SVR</b>	<b>28</b>	<b>3.265</b>	<b>0.414</b>	<b>12.666</b>
	<b>XGB</b>	<b>28</b>	<b>3.206</b>	<b>0.788</b>	<b>24.577</b>
	<b>XGB-LASSO</b>	<b>12</b>	<b>3.098</b>	<b>0.759</b>	<b>24.504</b>
	<b>XGB-PMI</b>	<b>7</b>	<b>3.209</b>	<b>0.785</b>	<b>24.472</b>
5	ELM	28	4.302	0.883	20.531
	EN	28	4.352	0.887	20.369
	MARS	28	2.856	1.413	49.482
	<b>SVR</b>	<b>28</b>	<b>3.178</b>	<b>0.440</b>	<b>13.832</b>
	<b>XGB</b>	<b>28</b>	<b>2.987</b>	<b>0.827</b>	<b>27.688</b>
	<b>XGB-LASSO</b>	<b>18</b>	<b>2.993</b>	<b>0.788</b>	<b>26.335</b>
	<b>XGB-PMI</b>	<b>8</b>	<b>3.237</b>	<b>0.971</b>	<b>30.000</b>
7	ELM	28	4.502	0.953	21.170
	EN	28	4.473	0.964	21.545
	MARS	28	2.714	1.571	57.861
	<b>SVR</b>	<b>28</b>	<b>2.913</b>	<b>0.650</b>	<b>22.313</b>
	<b>XGB</b>	<b>28</b>	<b>2.628</b>	<b>0.587</b>	<b>22.335</b>
	<b>XGB-LASSO</b>	<b>22</b>	<b>2.582</b>	<b>0.600</b>	<b>23.225</b>
	<b>XGB-PMI</b>	<b>8</b>	<b>2.306</b>	<b>0.812</b>	<b>35.194</b>

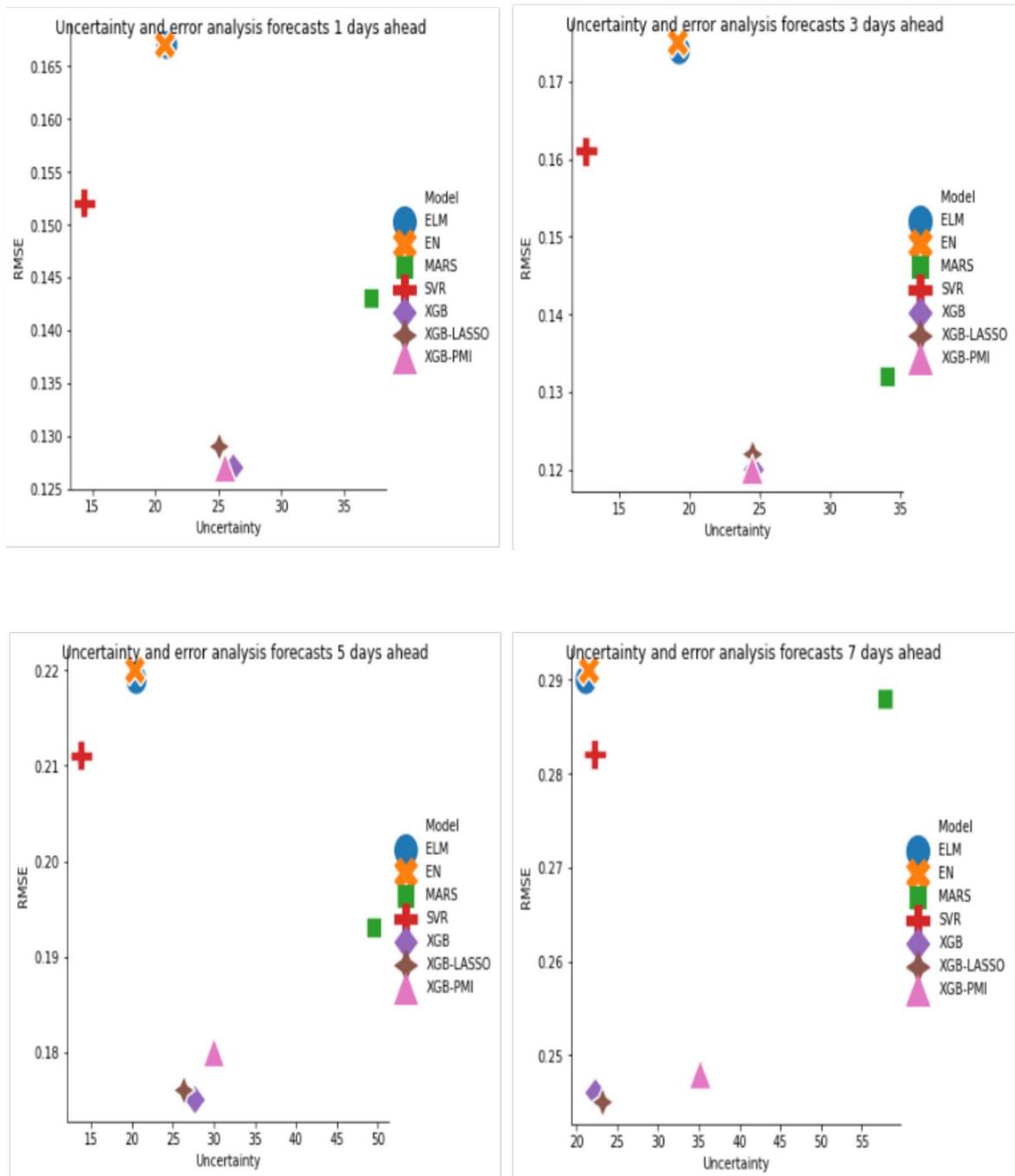
Fonte: Elaborada pelo autor (2023)

A Figura 36 mostra o gráfico de comparação que mapeia os modelos em relação ao RMSE e à Incerteza para cada horizonte de previsão. Esta figura mostra qualitativamente que os resultados do MARS levaram a uma maior incerteza e RMSE, enquanto SVR produziu uma pequena incerteza, porém com um RMSE alto, em todos horizontes de previsão.

Os modelos XGB, XGB-LASSO e XGB-PMI apresentaram valores muito baixos de RMSE e níveis de incerteza menores, portanto superaram os demais modelos quanto a precisão e confiabilidade em todos horizontes de previsão. No entanto, comparado ao

XGB, os modelo XGB-LASSO e XGB-PMI obtiveram erros e desempenho de incerteza semelhantes com menor complexidade considerando o número de variáveis para todos horizontes de previsão em análise, exceto 7 passos a frente, onde XGB-PMI atingiu nível alto de incerteza. Mas, por outro lado, XGB-PMI é menos complexo que XGB-LASSO, considerando o número de variáveis de entrada, assegurando assim a sua superioridade quanto à confiança.

Figura 36 - Análise de incertezas e erros.



Fonte: Elaborada pelo autor (2023)

## 6 CONCLUSÃO

No contexto da gestão e planejamento sustentável ou otimizado de recursos hídricos é necessário prever vazões acuradas e confiáveis. A previsão precisa da vazão continua sendo um desafio científico e tem recebido atenção considerável devido à natureza não-linear, não-estacionária e estocástica destas séries.

Com avanços em tecnologia, o futuro da pesquisa hidrológica provavelmente envolverá a maximização da informação, extração de observações complexas e dados coletados em todos os nossos sistemas ambientais, para melhorar a capacidade de previsão de variáveis ambientais complexas. Muitas vezes, a previsão dessas variáveis requer extensos conjuntos de dados e computação significativa. Este estudo visa superar essas restrições, desenvolvendo e examinando cinco modelos de aprendizado de máquinas nomeadamente EN, ELM, SVR, MARS e XGB. Além disso, quatro algoritmos evolutivos inspirados na natureza, GA, DE, PSO e GWO são empregados como estratégias para selecionar parâmetros internos dos cinco modelos. Através de um estudo de caso na barragem de Cahora-Bassa, na bacia hidrográfica do rio Zambeze, em Moçambique, o desempenho dos cinco modelos foi comparado com base em previsões de vários passos (multipassos) à frente da vazão afluente à barragem. Os dados foram fornecidos pelo Departamento de Recursos Hídricos e Meio Ambiente da Hidroelétrica de Cahora-Bassa (HCB), e são correspondentes ao período entre 2003 e 2018. O método k-fold ( $k = 5$ ) é usado como critério de particionamento de dados para testar e treinar a separação de conjuntos de dados.

Os experimentos foram realizados para avaliar suas capacidades de previsões pela aplicação de medidas de desempenho e testes de hipóteses estatísticas (Anova e Tukey). Os resultados obtidos mostraram que,

1. Os algoritmos evolutivos inspirados na natureza, aplicados para auxiliar na seleção de parâmetros dos modelos, podem auxiliar na implementação dos modelos de aprendizado de máquina.
2. O algoritmo GWO se mostrou superior aos algoritmos PSO, DE e GA, na determinação dos hiperparâmetros ótimos dos modelos de aprendizado de máquinas, para previsão com base em valores obtidos em cada passo do horizonte de tempo considerado.
3. O modelo XGB superou os outros (SVR, MARS, ELM e EN) na aplicação a todos algoritmos de busca evolutiva. Foi o melhor nas previsões feitas com os diferentes passos à frente, conforme as medidas de desempenho consideradas e o resultado dos testes estatísticos, sendo XGB com GWO integrado o modelo superior.

4. Houve uma boa aderência ou aproximação dos dados previstos pelos modelos com os observados, isto é, os modelos foram capazes de reproduzir as características da série de dados observados com excelente qualidade, porém, valores extremos são previstos com alguma incerteza.
5. O desempenho se deteriora à medida que o horizonte de previsão aumenta, isto é, horizontes menos distantes obtêm melhores resultados do que os relativamente mais distantes.

Embora os modelos de aprendizado de máquinas tenham mostrado um grande potencial para a simulação da vazão, o diagnóstico detalhado e a análise da interpretação das tendências incorporadas nesses modelos é importante para criar confiança em tais modelos. O uso de modelos como *caixa-preta*, ou seja, sem interpretação de como eles simulam a vazão, pode ser extremamente perigoso. Além disso, em relação ao XGB híbrido proposto, embora seja considerado robusto para o ajuste, gasta um tempo considerável para realizar o treinamento do modelo quando grandes bancos de dados são utilizados. Para reverter essa situação, uma abordagem utilizando dois métodos de seleção de variáveis de entrada: o LASSO e PMI foi incorporada ao modelo XGB híbrido para reduzir o número de variáveis de entrada, entre as sugeridas pelos métodos de autoregressão, o que resulta nos modelos XGB-LASSO e XGB-PMI. Os resultados indicaram que,

1. Nem todas as variáveis sugeridas pelas combinações de entrada recomendadas pela análise de correlação (ACF, PACF e CCFs) devem ser diretamente aplicadas aos métodos de aprendizado de máquinas. A utilização de métodos eficientes de seleção de variáveis é recomendada, por que permitem não só reduzir o número de variáveis, mas também capturar relações ocultas que não foram detectadas utilizando a análise de correlação, e fornecer informações abrangentes sobre a influência das variáveis no fluxo simulado, o que torna os modelos de aprendizado de máquina hidrológicamente robustos.
2. Maior capacidade de seleção dos modelos LASSO e PMI, sendo o PMI superior ao LASSO. O PMI pode determinar as entradas que melhor caracterizam a relação de entrada/saída subjacente, e fornece uma melhor representação do processo modelado quando comparado a alternativas lineares.
3. O número médio de variáveis aumenta conforme aumenta o horizonte de previsão. Esse aumento está relacionado à necessidade de incluir mais informações para horizontes maiores, indicando maior dificuldade em representar a vazão.
4. Alguns efeitos das variáveis hidrometeorológicas (anteriores de vazão, precipitação, evaporação e umidade relativa) anteriormente considerados como não-significativos pela análise de correlação, foram adicionados como preditores da vazão pelos

métodos de seleção de entradas. A vazão anterior, seguida da umidade relativa são as principais variáveis selecionadas; no entanto, outras variáveis são adicionadas em horizontes relativamente distantes (longo prazo).

As variações do modelo XGB desenvolvidas (XGB-PMI e XGB-LASSO) são opções eficientes para a previsão da vazão, pois fornecem ao XGB dados de entrada-saída mais informativos para melhorar seu desempenho, diminuem o esforço computacional, e promovem um melhor entendimento do processo gerador dos dados, tornando os modelos de aprendizado de máquina hidrológicamente robustos. Essa eficiência dos modelos foi comprovada através dos métodos de análise de incerteza (analíticos e de Montecarlo), nos quais em geral observou-se que os modelos apresentaram níveis de incerteza baixos, mesmo em horizontes distantes, mostrando, por sua vez, maior confiabilidade para sua aplicação. Ademais, seus resultados foram validados através da comparação com estudos recentes na literatura sobre previsão da vazão de rios localizados em regiões com diferentes condições hidroclimáticas e geomorfológicas.

Portanto, o modelo híbrido XGB proposto pode ser considerado uma alternativa superior aos modelos atualmente utilizados para a previsão diária da vazão, crucial para a atividade das usinas hidroelétricas. Esta previsão pode auxiliar na alocação da capacidade de armazenamento da barragem e na otimização dos procedimentos operacionais, na gestão da geração de energia elétrica, na manutenção das vazões ecológicas do reservatório e na obtenção contínua de registros de vazão em bacias não-calibradas, nas quais a vazão medida não esteja disponível. Além disso, a capacidade de compressão ou seleção do modelo desenvolvido neste trabalho sugere a sua aplicação em problemas em que grandes volumes dados são utilizados.

Este estudo é o primeiro desenvolvido sobre modelagem e previsão utilizando modelos de inteligência computacional ou de aprendizado de máquinas sobre a bacia hidrográfica do rio Zambeze; portanto, deu os primeiros passos na literatura, contribuindo no preenchimento da lacuna que existia nos estudos dessa natureza naquela região. Além disso, enquadra-se nos novos avanços sobre modelagem e previsão aplicando modelos de inteligência computacional em hidrologia.

## 6.1 PERSPECTIVAS DE TRABALHOS FUTUROS

Apresenta-se, a seguir, uma relação dos desenvolvimentos previstos para a continuidade da pesquisa. Entre os principais desenvolvimentos destacam-se:

- Aplicação de algoritmos multi-objetivos na otimização de parâmetros dos modelos de aprendizado de máquinas;
- Aplicação de métodos de quantificação da importância de variáveis de entrada como o método *SHapley Additive exPlanations*(SHAP);
- Aplicação de métodos de decomposição robustos, como a transformação *Wavelet* e outros;
- Aplicação de modelos de aprendizagem profunda (*Deep Learning*);
- O aprimoramento do modelo para valores extremos;
- Aplicação do método recursivo de previsão;
- Acoplamento de modelos analíticos (empíricos) de previsão hidrológica;
- Melhoramento contínuo do arcabouço computacional.

## 6.2 TRABALHOS PUBLICADOS

**Martinho, A. D.**, Saporetti, C. M., & Goliatt, L. (2022). Approaches for the short-term prediction of natural daily streamflows using hybrid machine learning enhanced with grey wolf optimization. *Hydrological Sciences Journal*, 1-18.

**Martinho, A. D.**, Ribeiro, C., Gorodetskaya, Y., Fonseca, T. L., & Goliatt, L. (2020, November). Extreme learning machine with evolutionary parameter tuning applied to forecast the daily natural flow at cahora bassa dam, mozambique. In *International Conference on Bioinspired Methods and Their Applications* (pp. 255-267). Springer, Cham.

**Martinho, A. D.**, Fonseca, T. L., & Goliatt, L. (2020, December). Automated Extreme Learning Machine to Forecast the Monthly Flows: A Case Study at Zambezi River. In *International Conference on Intelligent Systems Design and Applications* (pp. 1314-1324). Springer, Cham.

**Martinho, A. D.**, Goliatt, L., & Hippert, H. (2020). Gaussian process models applied for monthly forecast coal price mineral: a case study of mozambique. In *Proceedings of the XLI Ibero-Latin-American Congress on Computational Methods in Engineering, ABMEC*.

Souza, D. P., **Martinho, A. D.**, Rocha, C. C., da S Christo, E., & Goliatt, L. (2022). Hybrid particle swarm optimization and group method of data handling for short-term prediction of natural daily streamflows. *Modeling Earth Systems and Environment*, 8(4), 5743-5759.

Souza, D.P.M., **Martinho, A.D.**, Rocha, C.C. et al. Group method of data handling to forecast the daily water flow at the Cahora Bassa Dam. *Acta Geophys.* (2022). <https://doi.org/10.1007/s11600-022-00834-3>

Rosa, I. G., Carvalho, R. M., **Martinho, A. D.**, Goliatt, L., Neto, R. F., & Borges, C. C. Extensions and Improvements of the Extreme Learning Machine (ELM) Applied to Face Recognition.

Um artigo intitulado: Short-term multistep streamflow modeling using data-intelligence evolutionary machine learning models, em revisão na revista *Scientific Reports (SR)*.

Um Trabalho de Conclusão de Curso (TCC) intitulado: Modelos de aprendizado profundo com transferência de aprendizagem para modelagem hidrológica, desenvolvido pelo aluno de Graduação em Ciências da Computação Henrique Colonese Echternacht, com artigo submetido na revista *Stochastic Environmental Research and Risk Assessment (SERRA)*, no âmbito do projeto da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

## REFERÊNCIAS

ABUDU, Shalamu; CUI, Chun-liang; KING, James Phillip; ABUDUKADEER, Kaiser. Comparison of performance of statistical models in forecasting monthly streamflow of kizil river, china. **Water Science and Engineering**, Elsevier, v. 3, n. 3, p. 269–281, 2010.

ADNAN, Rana Muhammad; KISI, Ozgur; MOSTAFA, Reham R; AHMED, Ali Najah; EL-SHAFIE, Ahmed. The potential of a novel support vector machine trained with modified mayfly optimization algorithm for streamflow prediction. **Hydrological Sciences Journal**, Taylor & Francis, v. 67, n. 2, p. 161–174, 2022.

ADNAN, Rana Muhammad; LIANG, Zhongmin; HEDDAM, Salim; ZOUNEMAT-KERMANI, Mohammad; KISI, Ozgur; LI, Binquan. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. **Journal of Hydrology**, v. 586, p. 124371, 2020. ISSN 0022-1694.

ADNAN, Rana Muhammad; LIANG, Zhongmin; PARMAR, Kulwinder Singh; SONI, Kirti; KISI, Ozgur. Modeling monthly streamflow in mountainous basin by mars, gmdh-nn and denfis using hydroclimatic data. **Neural Computing and Applications**, Springer, v. 33, n. 7, p. 2853–2871, 2021.

ADNAN, Rana Muhammad; LIANG, Zhongmin; TRAJKOVIC, Slavisa; ZOUNEMAT-KERMANI, Mohammad; LI, Binquan; KISI, Ozgur. Daily streamflow prediction using optimally pruned extreme learning machine. **Journal of Hydrology**, Elsevier, v. 577, p. 123981, 2019.

ADNAN, Rana Muhammad; LIANG, Zhongmin; TRAJKOVIC, Slavisa; ZOUNEMAT-KERMANI, Mohammad; LI, Binquan; KISI, Ozgur. Daily streamflow prediction using optimally pruned extreme learning machine. **Journal of Hydrology**, v. 577, p. 123981, 2019. ISSN 0022-1694.

ADNAN, Rana Muhammad; MOSTAFA, Reham R; ELBELTAGI, Ahmed; YASEEN, Zaher Mundher; SHAHID, Shamsuddin; KISI, Ozgur. Development of new machine learning model for streamflow prediction: Case studies in pakistan. **Stochastic Environmental Research and Risk Assessment**, Springer, p. 1–35, 2022.

ADNAN, Rana Muhammad; MOSTAFA, Reham R; KISI, Ozgur; YASEEN, Zaher Mundher; SHAHID, Shamsuddin; ZOUNEMAT-KERMANI, Mohammad. Improving streamflow prediction using a new hybrid elm model combined with hybrid particle swarm optimization and grey wolf optimization. **Knowledge-Based Systems**, Elsevier, v. 230, p. 107379, 2021.

AFAN, Haitham Abdulmohsin; ALLAWI, Mohammed Falah; EL-SHAFIE, Amr; YASEEN, Zaher Mundher; AHMED, Ali Najah; MALEK, Marlinda Abdul; KOTING, Suhana Binti; SALIH, Sinan Q; MOHTAR, Wan Hanna Melini Wan; LAI, Sai Hin et al. Input attributes optimization using the feasibility of genetic nature inspired algorithm: application of river flow forecasting. **Scientific Reports**, Nature Publishing Group UK London, v. 10, n. 1, p. 4684, 2020.

AGHELPOUR, Pouya; BAHRAMI-PICHAGHCHI, Hadigheh; VARSHAVIAN, Vahid; GRAF, Renata; SEFIDKOUHI, Mohammad Ali Gholami; KHOSRAVESH, Mojtaba.

Evaluating the impact of large-scale climatic indices as inputs for forecasting monthly river flow in mazandaran province, iran. **Pure and Applied Geophysics**, Springer, p. 1–23, 2022.

AHMED, Ali Najah; LAM, To Van; HUNG, Nguyen Duy; THIEU, Nguyen Van; KISI, Ozgur; EL-SHAFIE, Ahmed. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. **Applied Soft Computing**, Elsevier, v. 105, p. 107282, 2021.

AKHTER, Muhammad Naveed; MEKHILEF, Saad; MOKHLIS, Hazlie; SHAH, Noraisyah Mohamed. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. **IET Renewable Power Generation**, IET, v. 13, n. 7, p. 1009–1023, 2019.

AL-JAWARNEH, Abdullah S; ISMAIL, Mohd Tahir; AWAJAN, Ahmad M. Elastic net regression and empirical mode decomposition for enhancing the accuracy of the model selection. **International Journal of Mathematical, Engineering and Management Sciences**, International Journal of Mathematical, Engineering and Management Sciences, v. 6, n. 2, p. 564, 2021.

AL-SUDANI, Zainab Abdulelah; SALIH, Sinan Q; YASEEN, Zaher Mundher et al. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. **Journal of Hydrology**, Elsevier, v. 573, p. 1–12, 2019.

AL-SUDANI, Zainab Abdulelah; SALIH, Sinan Q; YASEEN, Zaher Mundher et al. Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. **Journal of hydrology**, Elsevier, v. 573, p. 1–12, 2019.

AL-TASHI, Qasem; RAIS, Helmi Md; ABDULKADIR, Said Jadid; MIRJALILI, Seyedali; ALHUSSIAN, Hitham. A review of grey wolf optimizer-based feature selection methods for classification. **Evolutionary machine learning techniques**, Springer, p. 273–286, 2020.

ALBADRA, Musatafa Abbas Abbood; TIUNA, Sabrina. Extreme learning machine: a review. **International Journal of Applied Engineering Research**, v. 12, n. 14, p. 4610–4623, 2017.

ALIZADEH, Zahra; SHOURIAN, Mojtaba; YASEEN, Zaher Mundher. Simulating monthly streamflow using a hybrid feature selection approach integrated with an intelligence model. **Hydrological Sciences Journal**, Taylor & Francis, v. 65, n. 8, p. 1374–1384, 2020.

ALKHAMMASH, Eman H; KAMEL, Abdelmonaim Fakhry; AL-FATTAH, Saud M; ELSHEWEY, Ahmed M. Optimized multivariate adaptive regression splines for predicting crude oil demand in saudi arabia. **Discrete Dynamics in Nature and Society**, Hindawi, v. 2022, 2022.

ALMEIDA, Lizane; SERRA, Juan Carlos Valdés. Modelos hidrológicos, tipos e aplicações mais utilizadas. **Revista da FAE**, v. 20, n. 1, p. 129–137, 2017.

AN, Nguyen Hoang; ANH, Duong Tuan. Comparison of strategies for multi-step-ahead prediction of time series using neural network. In: IEEE. **2015 International Conference on Advanced Computing and Applications (ACOMP)**. [S.l.], 2015. p. 142–149.

ANJO, Antonio Batel. Objectivos de desenvolvimento sustentável e os recursos hídricos - uma reflexão. **Seminário Internacional sobre a Meteorologia e Dessalinização da Água**”, 18 de Novembro de 2015, Ministério da Ciência e Tecnologia, Ensino Superior e Técnico Profissional e o Instituto de Investigação em Águas, 11 2015.

ARAUJO, RL d; BARBOSA, HJC; BERNARDINO, HS. Evolução diferencial para problemas de otimização com restrições lineares. **Universidade Federal de Juiz de Fora (UFJF)**, p. 46, 2016.

ARNOLD, Jeffrey G; MORIASI, Daniel N; GASSMAN, Philip W; ABBASPOUR, Karim C; WHITE, Michael J; SRINIVASAN, Raghavan; SANTHI, Chinnasamy; HARMEL, RD; GRIENSVEN, Ann Van; LIEW, Michael W Van et al. Swat: Model use, calibration, and validation. **Transactions of the ASABE**, American Society of Agricultural and Biological Engineers, v. 55, n. 4, p. 1491–1508, 2012.

BARROS, MTL de et al. A programação dinâmica aplicada à engenharia de recursos hídricos. **TÉCNICAS quantitativas para gerenciamento de recursos hídricos. Porto Alegre: Ed. da Universidade/UFRGS: ABRH**, p. 239–277, 1997.

BARROS, Rita. **Integrated Water Resource Management in Mozambique: The case of the Limpopo River Basin**. 2009. Dissertação (Mestrado) — ETH Zürich, Departement für Umweltwissenschaften, 2009.

BAYER, Fábio Mariano. Wavelets e modelos tradicionais de previsão: Um estudo comparativo. **Rev. Bras. Biom**, v. 28, n. 2, p. 40–61, 2010.

BENIAND, G; WANG, J. **Swarm Intelligence in Cellular Robotic Systems**. [S.l.]: Springer, Heidelberg, 1993.

BERGMEIR, Christoph; BENÍTEZ, José M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, Elsevier, v. 191, p. 192–213, 2012.

BERTONI, JC; TUCCI, CEM; TUCCI, CEM. Hidrologia: ciência e aplicação. **Porto Alegre: UFRGS**, p. 177–241, 2007.

BEZERRA, Manoel Ivanildo Silvestre. Apostila de análise de séries temporais. **UNESP: Curso de estatística. São Paulo, SP**, 2006.

BISHOP, Christopher M; NASRABADI, Nasser M. **Pattern recognition and machine learning**. [S.l.]: Springer, 2006. v. 4.

BOX, George EP; JENKINS, Gwilym M; REINSEL, Gregory C; LJUNG, Greta M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.

BOX, Gwilym M. Jenkins George E. P. **Time Series Analysis: Forecasting and Control**. [S.l.]: Holden-Day, 1976. (Holden-Day series in time series analysis and digital processing). ISBN 9780816211043.

BRAVO, JM. **Subsídios à Operação de Reservatórios Baseada na Previsão de Variáveis Hidrológicas**. 226 f. 2010. Tese (Doutorado) — Tese (Doutorado)—Instituto de Pesquisas Hidráulicas, Universidade Federal do . . . , 2010.

BRAVO, Juan Martín; COLLISCHONN, Walter; PILAR, Jorge Víctor; SILVA, BC; TUCCI, C. Avaliação dos benefícios de previsões de vazão na operação de reservatórios. **Revista Brasileira de Recursos Hídricos**, v. 13, p. 197–213, 2009.

BRENTAN, Bruno M; JR, Edevar Luvizotto; HERRERA, Manuel; IZQUIERDO, Joaquín; PÉREZ-GARCÍA, Rafael. Hybrid regression model for near real-time urban water demand forecasting. **Journal of Computational and Applied Mathematics**, Elsevier, v. 309, p. 532–541, 2017.

BRITO, Luis de et al. Cidadania e governação em moçambique. 2008.

BROWNLEE, Jason. **Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python**. [S.l.]: Machine Learning Mastery, 2018.

BRUN, Gerti W; TUCCI, Carlos EM. Previsão em tempo real do volume afluente ao reservatório de Ernestina. **Revista Brasileira de Recursos Hídricos**, v. 6, n. 2, p. 73–79, 2001.

BUENO, Eduardo de Oliveira; MELLO, Carlos Rogério de; ALVES, Geovane Junqueira. Evaporação do reservatório da usina hidrelétrica de Camargos: caracterização da pegada hídrica. **RBRH**, SciELO Brasil, v. 21, n. 3, p. 570–575, 2016.

CARVALHO, Thiago Morato de. Técnicas de medição de vazão por meios convencionais e não convencionais. **Revista Brasileira de Geografia Física**, v. 1, n. 1, p. 73–85, 2008.

CARVALHO, Wanderson Laerte de Oliveira. **Estudo de parâmetros ótimos em algoritmos genéticos elitistas**. 2017. Dissertação (Mestrado) — Brasil, 2017.

CASAGRANDE, Leandro; TOMASELLA, Javier; ALVALÁ, Regina Célia dos Santos; BOTTINO, Marcus Jorge; CARAM, Rochane de Oliveira. Early flood warning in the Itajaí-açu river basin using numerical weather forecasting and hydrological modeling. **Natural hazards**, Springer, v. 88, n. 2, p. 741–757, 2017.

CHANG, Chih-Chung; LIN, Chih-Jen. LIBSVM: A Library for Support Vector Machines. **ACM Trans. Intell. Syst. Technol.**, Association for Computing Machinery, New York, NY, USA, v. 2, n. 3, 2011. ISSN 2157-6904.

CHATFIELD, Chris; XING, Haipeng. **The analysis of time series: an introduction with R**. [S.l.]: CRC press, 2019.

CHEN, Lu; SUN, Na; ZHOU, Chao; ZHOU, Jianzhong; ZHOU, Yanlai; ZHANG, Junhong; ZHOU, Qing. Flood forecasting based on an improved extreme learning machine model combined with the backtracking search optimization algorithm. **Water**, Multidisciplinary Digital Publishing Institute, v. 10, n. 10, p. 1362, 2018.

CHEN, Rong; LIANG, Chang-Yong; HONG, Wei-Chiang; GU, Dong-Xiao. Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. **Applied Soft Computing**, Elsevier, v. 26, p. 435–443, 2015.

- CHEN, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**, p. 785–794, 2016.
- CHEN, Tianqi; HE, Tong. Higgs boson discovery with boosted trees. In: **NIPS 2014 workshop on high-energy physics and machine learning**. [S.l.: s.n.], 2015. p. 69–80.
- CHENG, M; FANG, F; KINOUCI, T; NAVON, IM; PAIN, CC. Long lead-time daily and monthly streamflow forecasting using machine learning methods. **Journal of Hydrology**, Elsevier, v. 590, p. 125376, 2020.
- CHENG, Min-Yuan; CAO, Minh-Tu. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. **Applied Soft Computing**, v. 22, p. 178 – 188, 2014. ISSN 1568-4946.
- CHU, Haibo; WEI, Jiahua; WU, Wenyan. Streamflow prediction using lasso-fcm-dbn approach based on hydro-meteorological condition classification. **Journal of Hydrology**, Elsevier, v. 580, p. 124253, 2020.
- CHU, Haibo; WEI, Jiahua; WU, Wenyan; JIANG, Yuan; CHU, Qi; MENG, Xiujing. A classification-based deep belief networks model framework for daily streamflow forecasting. **Journal of Hydrology**, Elsevier, v. 595, p. 125967, 2021.
- CLAESEN, Marc; MOOR, Bart De. Hyperparameter search in machine learning. **CoRR**, abs/1502.02127, 2015.
- COLLISCHONN, Walter; DORNELLES, Fernando. Hidrologia para engenharia e ciências ambientais. **Porto Alegre: Associação Brasileira de Recursos Hídricos**, v. 1, p. 336, 2013.
- COSTA, Geslie Fernandes. Previsão multi-passos em séries temporais: Estratégias clássicas e de aprendizagem automática. 2015.
- COSTA, Samira Domingos. Estratégias de previsão multipassos à frente para vazão afluente em bacias hidrográficas de diferentes dinâmicas. Universidade Federal de Minas Gerais, 2014.
- CURTARELLI, Marcelo Pedroso; ALCÂNTARA, Enner Herenio; ARAÚJO, Carlos Alberto Sampaio de; STECH, José Luiz; LORENZZETTI, João Antônio. Avaliação da dinâmica temporal da evaporação no reservatório de itumbiara, go, utilizando dados obtidos por sensoriamento remoto. **Revista Ambiente & Água**, SciELO Brasil, v. 8, p. 272–289, 2013.
- DEHGHANI, Majid; SEIFI, Akram; RIAHI-MADVAR, Hossien. Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining anfis and grey wolf optimization. **Journal of Hydrology**, Elsevier, v. 576, p. 698–725, 2019.
- DEKA, Paresh Chandra et al. Support vector machine applications in the field of hydrology: a review. **Applied soft computing**, Elsevier, v. 19, p. 372–386, 2014.
- DELFORGE, Damien; VIRON, Olivier de; VANCLOOSTER, Marnik; CAMP, Michel Van; WATLET, Arnaud. Detecting hydrological connectivity using causal inference from time series: synthetic and real karstic case studies. **Hydrology and Earth System Sciences**, Copernicus GmbH, v. 26, n. 8, p. 2181–2199, 2022.

DIFI, Salah; ELMEDDAHI, Yamina; HEBAL, Aziz; SINGH, Vijay P; HEDDAM, Salim; KIM, Sungwon; KISI, Ozgur. Monthly streamflow prediction using hybrid extreme learning machine optimized by bat algorithm: a case study of cheliff watershed, algeria. **Hydrological Sciences Journal**, Taylor & Francis, n. just-accepted, 2022.

DING, Daizong; ZHANG, Mi; PAN, Xudong; YANG, Min; HE, Xiangnan. Modeling extreme events in time series prediction. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**. [S.l.: s.n.], 2019. p. 1114–1122.

DING, Shifei; ZHAO, Han; ZHANG, Yanan; XU, Xinzheng; NIE, Ru. Extreme learning machine: algorithm, theory and applications. **Artificial Intelligence Review**, Springer, v. 44, n. 1, p. 103–115, 2015.

DMITRIEVA, Ksenia. **Forecasting of a hydropower plant energy production**. 2015. Dissertação (Mestrado) — Norway, 2015.

DUAN, Junbo; SOUSSEN, Charles; BRIE, David; IDIER, Jérôme; WAN, Mingxi; WANG, Yu-Ping. Generalized lasso with under-determined regularization matrices. **Signal processing**, Elsevier, v. 127, p. 239–246, 2016.

DUARTE, Victor Braga Rodrigues; SILVA, Francisca de Cássia Silva da; SOUZA, Igor Viana; SILVA, Marcos Vinicius Cardoso; SOUSA, Hygor Gomes de Almeida; GIONGO, Marcos; VIOLA, Marcelo Ribeiro. Previsão de vazão na bacia hidrográfica do rio manuel alves da natividade utilizando o modelo de séries temporais sarima. **Journal of Biotechnology and Biodiversity**, v. 7, n. 4, p. 457–468, 2019.

ESSAM, Yusuf; HUANG, Yuk Feng; NG, Jing Lin; BIRIMA, Ahmed H; AHMED, Ali Najah; EL-SHAFIE, Ahmed. Predicting streamflow in peninsular malaysia using support vector machine and deep learning algorithms. **Scientific Reports**, Nature Publishing Group, v. 12, n. 1, p. 1–26, 2022.

ESTÉVEZ, Pablo A; TESMER, Michel; PEREZ, Claudio A; ZURADA, Jacek M. Normalized mutual information feature selection. **IEEE Transactions on neural networks**, IEEE, v. 20, n. 2, p. 189–201, 2009.

EVERINGHAM, YL; SEXTON, J et al. An introduction to multivariate adaptive regression splines for the cane industry. In: **Proceedings of the 2011 Conference of the Australian Society of Sugar Cane Technologists**. [S.l.: s.n.], 2011.

EYDURAN, Ecevit; AKIN, Meleksen; EYDURAN, Sadiye Peral. Application of multivariate adaptive regression splines in agricultural sciences through r software. **Nobel Bilimsel Eserler Sertifika**, n. 20779, 2019.

FAHIMI, Farzad; YASEEN, Zaher Mundher; EL-SHAFIE, Ahmed. Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review. **Theoretical and applied climatology**, Springer, v. 128, n. 3, p. 875–903, 2017.

FALSARELLA, Orandi; SILVA, Maria Luiza Ramos da; MARIOSA, Duarcides Ferreira. O processo de decisão na gestão de recursos hídricos: a contribuição da internet das coisas (iot) e big data. **Journal on Innovation and Sustainability RISUS**, v. 13, n. 2, p. 45–58, 2022.

- FARIS, Hossam; ALJARAH, Ibrahim; AL-BETAR, Mohammed Azmi; MIRJALILI, Seyedali. Grey wolf optimizer: a review of recent variants and applications. **Neural computing and applications**, Springer, v. 30, n. 2, p. 413–435, 2018.
- FERREIRA, GILZA SANTOS SIMÃO. Intervalo de confiança para previsão de carga a curto-prazo. 2013.
- FERREIRA, Lucas Borges; DUARTE, Anunciene Barbosa; CUNHA, Fernando França da; FERNANDES, Elpídio Inácio. Multivariate adaptive regression splines (mars) applied to daily reference evapotranspiration modeling with limited weather data. **Acta Scientiarum. Agronomy**, SciELO Brasil, v. 41, 2019.
- FERREIRA, Renan Gon; SILVA, Demetrius David da; ELESBON, Abrahão Alexandre Alden; FERNANDES-FILHO, Elpídio Inácio; VELOSO, Gustavo Vieira; FRAGA, Micael de Souza; FERREIRA, Lucas Borges. Machine learning models for streamflow regionalization in a tropical watershed. **Journal of Environmental Management**, v. 280, p. 111713, 2021. ISSN 0301-4797.
- FERREIRA, Tamiris Capellaro. **Impactos e desafios da construção civil brasileira para os Objetivos de Desenvolvimento Sustentável**. 2018. Tese (Doutorado) — Universidade de São Paulo, 2018.
- FIGUEREDO, Clodoaldo José. Previsão de séries temporais utilizando a metodologia box & jenkins e redes neurais para inicialização de planejamento e controle da produção. **Curitiba: UFPR**, 2008.
- FILHO, Mauricio Camargo; FREITAS, Andreza Rocha de; SILVA, Sebastião Celso Ferreira da; SOUZA, Edilberto Leopolski de; TRATZ, Eliza do Belém. Hidrometria aplicada: medição de vazão da seção 2 do rio cascavel, guarapuava (pr) applied hydrometry: outflow measurement of section 2 of cascavel river. **Ambiência**, v. 5, n. 2, p. 333–342, 2009.
- FLUET-CHOUINARD, Etienne; AEBERHARD, William; SZEKELY, Eniko; ZAPPA, Massimiliano; BOGNER, Konrad; SENEVIRATNE, Sonia I; GUDMUNDSSON, Lukas. **Machine learning-derived predictions of river flow across Switzerland**. [S.l.], 2022.
- FRASER, Andrew M; SWINNEY, Harry L. Independent coordinates for strange attractors from mutual information. **Physical review A**, APS, v. 33, n. 2, p. 1134, 1986.
- FREITAS, Antonio Airton Carneiro de et al. Previsão de series temporais via seleção de variáveis, reconstrução dinamica, arma-garch e redes neurais artificiais. [sn], 2007.
- FRIEDMAN, Jerome H. Multivariate adaptive regression splines. **Ann. Statist.**, The Institute of Mathematical Statistics, v. 19, n. 1, p. 1–67, 03 1991.
- FRIEDMAN, Jerome H. **The elements of statistical learning: Data mining, inference, and prediction**. [S.l.]: springer open, 2017.
- GABRIEL, Paulo Henrique Ribeiro; DELBEM, Alexandre Cláudio Botazzo. **Fundamentos de algoritmos evolutivos**. [S.l.]: ICMC-USP, 2008.
- GODEIRO, Lucas Lúcio et al. Ensaio sobre modelos de previsão econômica. Universidade Federal da Paraíba, 2018.

GOLIATT, Leonardo; SULAIMAN, Sadeq Oleiwi; KHEDHER, Khaled Mohamed; FAROOQUE, Aitazaz Ahsan; YASEEN, Zaher Mundher. Estimation of natural streams longitudinal dispersion coefficient using hybrid evolutionary machine learning model. **Engineering Applications of Computational Fluid Mechanics**, Taylor & Francis, v. 15, n. 1, p. 1298–1320, 2021.

GOMES, Frederico Pimentel. **A comparação entre médias de tratamentos na análise da variância**. [S.l.]: SciELO Brasil, 1954. 1–12 p.

GÓMEZ, Carlos Arenal; HUERTAS, Eduardo Caro. Modelo de temperatura para la mejora de la predicción de la demanda eléctrica: Aplicación al sistema peninsular español. 2016.

GONÇALVES, André Ricardo. Máquina de vetores suporte. **Acesso em**, v. 21, 2015.

GONZALES, Gerard Bryan; SAEGER, Sarah De. Elastic net regularized regression for time-series analysis of plasma metabolome stability under sub-optimal freezing condition. **Scientific Reports**, Nature Publishing Group, v. 8, n. 1, p. 1–10, 2018.

GORODETSKAYA, Yulia; FONSECA, Leonardo Goliatt da; RIBEIRO, Celso Bandeira de Melo. Previsão de vazão de curto prazo utilizando redes neurais artificiais. estudo de caso: Rio paraíba do sul. 2018.

GU, Jirong; ZHU, Mingcang; JIANG, Liuguangyan. Housing price forecasting based on genetic algorithm and support vector machine. **Expert Systems with Applications**, v. 38, n. 4, p. 3383–3386, 2011. ISSN 0957-4174. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0957417410009310>.

GUETTER, Kolodinskye. **ATUALIZAÇÃO DE MODELOS CHUVA-VAZÃO-PROPAGAÇÃO COM ESTIMADORES DE ESTADO**. 2005. Tese (Doutorado) — Universidade Federal do Paraná, 2005.

GUILHON, Luiz Guilherme Ferreira. **Modelo heurístico de previsão de vazões naturais médias semanais aplicado à usina de Foz do Areia**. 2002. Tese (Doutorado) — Dissertação de Mestrado. COPPE UFRJ, 2002.

GUILHON, Luiz Guilherme Ferreira; ROCHA, Vinicius Forain; MOREIRA, Jonatas Costa. Comparação de métodos de previsão de vazões naturais afluentes a aproveitamentos hidroelétricos. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 3, p. 13–20, 2007.

GUILHON, Luiz Guilherme Ferreira; ROCHA, Vinicius Forain; MOREIRA, Jonatas Costa. Comparação de métodos de previsão de vazões naturais afluentes a aproveitamentos hidroelétricos. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 3, p. 13–20, 2007.

GUNN, Steve R et al. Support vector machines for classification and regression. **ISIS technical report**, v. 14, n. 1, p. 5–16, 1998.

GUO, Peng; CHENG, Wenming; WANG, Yi. Hybrid evolutionary algorithm with extreme machine learning fitness function evaluation for two-stage capacitated facility location problems. **Expert Systems with Applications**, v. 71, p. 57 – 68, 2017. ISSN 0957-4174.

GUO, Pi; ZENG, Fangfang; HU, Xiaomin; ZHANG, Dingmei; ZHU, Shuming; DENG, Yu; HAO, Yuantao. Improved variable selection algorithm using a lasso-type penalty, with an application to assessing hepatitis b infection relevant factors in community residents. **PLoS One**, Public Library of Science San Francisco, CA USA, v. 10, n. 7, p. e0134151, 2015.

GUPTA, Hoshin V; KLING, Harald; YILMAZ, Koray K; MARTINEZ, Guillermo F. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. **Journal of hydrology**, Elsevier, v. 377, n. 1-2, p. 80–91, 2009.

HADI, Sinan Jasim; ABBA, Sani Isah; SAMMEN, Saad Sh; SALIH, Sinan Q; AL-ANSARI, Nadhir; YASEEN, Zaher Mundher. Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation. **IEEE Access**, IEEE, v. 7, p. 141533–141548, 2019.

HARRISON, Robert L. Introduction to monte carlo simulation. In: AMERICAN INSTITUTE OF PHYSICS. **AIP conference proceedings**. [S.l.], 2010. v. 1204, n. 1, p. 17–21.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning - Data Mining, Inference, and Prediction**. 2. ed. Verlag, New York: Springer, 2009. 745 p. ISBN 978-0-387-84857-0.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. New York: Springer Science+Business Media, 2009.

HAVEROTH, Roberto; PITZ, Joabe; ELI, Katiani; NEVES, Leonardo; FERNANDES, Elizabete. Modelagem para a estimativa da umidade relativa do ar para a cidade de ituporanga/sc. **Enciclopédia Biosfera**, v. 8, n. 15, 2012.

HAZNEDAR, Bulent; KILINC, Huseyin Cagan. A hybrid anfis-ga approach for estimation of hydrological time series. **Water Resources Management**, Springer, v. 36, n. 12, p. 4819–4842, 2022.

HEINEN, Eduarth Dapper. Redes neurais recorrentes e xgboost aplicados à previsão de radiação solar no horizonte de curto prazo. Universidade Federal de São Carlos, 2018.

HERRERA, Manuel; TORGO, Luís; IZQUIERDO, Joaquín; PÉREZ-GARCÍA, Rafael. Predictive models for forecasting hourly urban water demand. **Journal of hydrology**, Elsevier, v. 387, n. 1-2, p. 141–150, 2010.

HOERL, Arthur E; KENNARD, Robert W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, Taylor & Francis, v. 12, n. 1, p. 55–67, 1970.

HONG, Jiyeong; LEE, Seoro; BAE, Joo Hyun; LEE, Jimin; PARK, Woon Ji; LEE, Dongjun; KIM, Jonggun; LIM, Kyoung Jae. Development and evaluation of the combined machine learning models for the prediction of dam inflow. **Water**, Multidisciplinary Digital Publishing Institute, v. 12, n. 10, p. 2927, 2020.

HU, Zhongyi; BAO, Yukun; XIONG, Tao; CHIONG, Raymond. Hybrid filter–wrapper feature selection for short-term load forecasting. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 40, p. 17–27, 2015.

HUANG, Guang-Bin; CHEN, Lei; SIEW, Chee Kheong et al. Universal approximation using incremental constructive feedforward networks with random hidden nodes. **IEEE Trans. Neural Networks**, v. 17, n. 4, p. 879–892, 2006.

HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE. **Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on**. [S.l.], 2004. v. 2, p. 985–990.

HUGHES, DA; MANTEL, S; FARINOSI, F. Assessing development and climate variability impacts on water resources in the zambezi river basin: Initial model calibration, uncertainty issues and performance. **Journal of Hydrology: Regional Studies**, Elsevier, v. 32, p. 100765, 2020.

HULSMAN, Petra; SAVENIJE, Hubert HG; HRACHOWITZ, Markus. Satellite-based drought analysis in the zambezi river basin: Was the 2019 drought the most extreme in several decades as locally perceived? **Journal of Hydrology: Regional Studies**, Elsevier, v. 34, p. 100789, 2021.

HUSSAIN, Dostdar; KHAN, Aftab. Machine learning techniques for monthly river flow forecasting of hunza river, pakistan. **Earth Science Informatics**, p. 1–11, 02 2020.

HUSSAIN, Dostdar; KHAN, Aftab Ahmed. Machine learning techniques for monthly river flow forecasting of hunza river, pakistan. **Earth Science Informatics**, v. 13, n. 3, p. 939–949, Sep 2020. ISSN 1865-0481. Disponível em: <https://doi.org/10.1007/s12145-020-00450-z>.

IBRAHIM, Karim Sherif Mostafa Hassan; HUANG, Yuk Feng; AHMED, Ali Najah; KOO, Chai Hoon; EL-SHAFIE, Ahmed. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. **Alexandria Engineering Journal**, Elsevier, v. 61, n. 1, p. 279–303, 2022.

IKRAM, Rana Muhammad Adnan; GOLIATT, Leonardo; KISI, Ozgur; TRAJKOVIC, Slavisa; SHAHID, Shamsuddin. Covariance matrix adaptation evolution strategy for improving machine learning approaches in streamflow prediction. **Mathematics**, MDPI, v. 10, n. 16, p. 2971, 2022.

ISLAM, SFN; SHOLAHUDDIN, A; ABDULLAH, AS. Extreme gradient boosting (xgboost) method in making forecasting application and analysis of usd exchange rates against rupiah. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2021. v. 1722, n. 1, p. 012016.

IZBICKI, Rafael; SANTOS, Tiago Mendonça dos. **Aprendizado de máquina: uma abordagem estatística**. [S.l.]: Rafael Izbicki, 2020.

JAMEI, Mehdi; ALI, Mumtaz; MALIK, Anurag; PRASAD, Ramendra; ABDULLA, Shahab; YASEEN, Zaher Mundher. Forecasting daily flood water level using hybrid advanced machine learning based time-varying filtered empirical mode decomposition approach. **Water Resources Management**, Springer, v. 36, n. 12, p. 4637–4676, 2022.

JESSEN, Gustavo; SILVA, Henrique. Gestão hidrológica da albufeira de cahora bassa em períodos críticos (cheias e secas). In: **5º Congresso Luso-Moçambicano de**

**Engenharia & 2º Congresso de Engenharia de Moçambique, Tete.** [S.l.: s.n.], 2008.

JIANG, Yan; BAO, Xin; HAO, Shaonan; ZHAO, Hongtao; LI, Xuyong; WU, Xianing. Monthly streamflow forecasting using elm-ipso based on phase space reconstruction. **Water Resources Management**, Springer, v. 34, n. 11, p. 3515–3531, 2020.

JUNIOR, Gilberto Pereira de Alcântara. Avaliação do lasso e métodos alternativos em modelos de regressão logística. Universidade Federal de São Carlos, 2021.

KARGAR, Katayoun; AL. et. Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. **Engineering Applications of Computational Fluid Mechanics**, Taylor & Francis, v. 14, n. 1, p. 311–322, 2020.

KARTHIKEYAN, Muthukumarasamy; VYAS, Renu. Machine learning methods in chemoinformatics for drug discovery. In: **Practical chemoinformatics**. [S.l.]: Springer, 2014. p. 133–194.

KHOSRAVI, Khabat; GOLKARIAN, Ali; TIEFENBACHER, John P. Using optimized deep learning to predict daily streamflow: A comparison to common machine learning algorithms. **Water Resources Management**, Springer, v. 36, n. 2, p. 699–716, 2022.

KILINC, Huseyin Cagan. Daily streamflow forecasting based on the hybrid particle swarm optimization and long short-term memory model in the orontes basin. **Water**, MDPI, v. 14, n. 3, p. 490, 2022.

KILINC, Huseyin Cagan; YURTSEVER, Adem. Short-term streamflow forecasting using hybrid deep learning model based on grey wolf algorithm for hydrological time series. **Sustainability**, MDPI, v. 14, n. 6, p. 3352, 2022.

KISI, Ozgur; HEDDAM, Salim; KESHTEGAR, Behrooz; PIRI, Jamshid; ADNAN, Rana Muhammad. Predicting daily streamflow in a cold climate using a novel data mining technique: Radial m5 model tree. **Water**, Multidisciplinary Digital Publishing Institute, v. 14, n. 9, p. 1449, 2022.

KLEMEŠ, Vit. Operational testing of hydrological simulation models. **Hydrological sciences journal**, Taylor & Francis, v. 31, n. 1, p. 13–24, 1986.

KLING, Harald; STANZEL, Philipp; PREISHUBER, Martin. Impact modelling of water resources development and climate scenarios on zambezi river discharge. **Journal of Hydrology: Regional Studies**, Elsevier, v. 1, p. 17–43, 2014.

KOLLER, Daphne; SAHAMI, Mehran. **Toward optimal feature selection**. [S.l.], 1996.

KRASKOV, Alexander; STÖGBAUER, Harald; GRASSBERGER, Peter. Erratum: estimating mutual information [phys. rev. e 69, 066138 (2004)]. **Physical Review E**, APS, v. 83, n. 1, p. 019903, 2011.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. [S.l.]: Springer, 2013. v. 26.

KULLBACK, Solomon. **Information theory and statistics**. [S.l.]: Courier Corporation, 1997.

KWON, Sunghoon; LEE, Sangin; NA, Okyoung. Tuning parameter selection for the adaptive lasso in the autoregressive model. **Journal of the Korean Statistical Society**, Elsevier, v. 46, n. 2, p. 285–297, 2017.

LAARNE, Petri; ZAIDAN, Martha A; NIEMINEN, Tuomo. Ennemi: non-linear correlation detection with mutual information. **SoftwareX**, Elsevier, v. 14, p. 100686, 2021.

LATIF, Sarmad Dashti; AHMED, Ali Najah; SATHIAMURTHY, Edlic; HUANG, Yuk Feng; EL-SHAFIE, Ahmed. Evaluation of deep learning algorithm for inflow forecasting: a case study of durian tunggal reservoir, peninsular malaysia. **Natural Hazards**, Springer, v. 109, n. 1, p. 351–369, 2021.

LATIFOĞLU, Levent. A novel approach for prediction of daily streamflow discharge data using correlation based feature selection and random forest method. **International Advanced Researches and Engineering Journal**, v. 6, n. 1, p. 1–7, 2022.

LI, Jun; WANG, Zhaoli; LAI, Chengguang; ZHANG, Zhenxing. Tree-ring-width based streamflow reconstruction based on the random forest algorithm for the source region of the Yangtze River, China. **CATENA**, v. 183, p. 104216, 2019. ISSN 0341-8162.

LI, Xuyuan; MAIER, Holger R; ZECCHIN, Aaron C. Improved pmi-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. **Environmental Modelling & Software**, Elsevier, v. 65, p. 15–29, 2015.

LI, Yujie; WEI, Jing; WANG, Dong; LI, Bo; HUANG, Huaping; XU, Bin; XU, Yueping. A medium and long-term runoff forecast method based on massive meteorological data and machine learning algorithms. **Water**, v. 13, n. 9, 2021. ISSN 2073-4441. Disponível em: <https://www.mdpi.com/2073-4441/13/9/1308>.

LIANG, Jun; HOU, Liang; LUAN, Zhenhua; HUANG, Weiping. Feature selection with conditional mutual information considering feature interaction. **Symmetry**, v. 11, n. 7, 2019. ISSN 2073-8994. Disponível em: <https://www.mdpi.com/2073-8994/11/7/858>.

LIMA, Aranildo R; CANNON, Alex J; HSIEH, William W. Forecasting daily streamflow using online sequential extreme learning machines. **Journal of hydrology**, Elsevier, v. 537, p. 431–443, 2016.

LIMA, Aranildo R; HSIEH, William W; CANNON, Alex J. Variable complexity online sequential extreme learning machine, with applications to streamflow prediction. **Journal of Hydrology**, Elsevier, v. 555, p. 983–994, 2017.

LIMA, Rômulo Galdino da Rocha et al. Desenvolvimento de um modelo de previsão hidrológico usando máquina vetorial de suporte regressivo: uma aproximação computacional para modelagem da bacia do rio arkansas. Universidade Federal Rural do Semi-Árido, 2018.

LIN, Yongen; WANG, Dagang; WANG, Guiling; QIU, Jianxiu; LONG, Kaihao; DU, Yi; XIE, Hehai; WEI, Zhongwang; SHANGGUAN, Wei; DAI, Yongjiu. A hybrid deep learning algorithm and its application to streamflow prediction. **Journal of Hydrology**, Elsevier, v. 601, p. 126636, 2021.

LIU, Wei; DOU, Zhenhai; WANG, Weiguo; LIU, Yueyu; ZOU, Hao; ZHANG, Bo; HOU, Shoujun. Short-term load forecasting based on elastic net improved gmdh and difference degree weighting optimization. **Applied Sciences**, MDPI, v. 8, n. 9, p. 1603, 2018.

LIU, Yanfeng; ZHOU, Yong; WANG, Dengjia; WANG, Yingying; LI, Yong; ZHU, Ying. Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in china. **Energy Conversion and Management**, v. 154, p. 168–179, 2017. ISSN 0196-8904. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0196890417309779>.

LIU, Zhiyong; ZHOU, Ping; CHEN, Xiuzhi; GUAN, Yinghui. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. **Journal of Geophysical Research: Atmospheres**, Wiley Online Library, v. 120, n. 19, p. 10–116, 2015.

MACHIWAL, Deepesh; JHA, Madan Kumar. **Hydrologic time series analysis: theory and practice**. [S.l.]: Springer Science & Business Media, 2012.

MALIK, Anurag; TIKHAMARINE, Yazid; SOUAG-GAMANE, Doudja; KISI, Ozgur; PHAM, Quoc Bao. Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 34, n. 11, p. 1755–1773, 2020.

MANDAL, Ankush; DAS, Swagatam; ABRAHAM, Ajith. A differential evolution based memetic algorithm for workload optimization in power generation plants. In: IEEE. **2011 11th International Conference on Hybrid Intelligent Systems (HIS)**. [S.l.], 2011. p. 271–276.

MARACAJÁ, José Rosenilton de Araújo et al. Previsão regionalizada de vazão sazonal utilizando redes neurais artificiais. Universidade Federal de Campina Grande, 2005.

MARTENS, David; BAESEENS, Bart; FAWCETT, Tom. Editorial survey: swarm intelligence for data mining. **Machine Learning**, Springer, v. 82, n. 1, p. 1–42, 2011.

MARTINHO, Alfeu D; SAPORETTI, Camila M; GOLIATT, Leonardo. Hybrid machine learning approaches enhanced with grey wolf optimization to the short-term prediction of natural daily streamflows. **Hydrological Sciences Journal**, Taylor & Francis, n. just-accepted, 2022.

MASINI, Ricardo P; MEDEIROS, Marcelo C; MENDES, Eduardo F. Machine learning advances for time series forecasting. **Journal of Economic Surveys**, Wiley Online Library, 2021.

MASUM, Shamsul; LIU, Ying; CHIVERTON, John. Multi-step time series forecasting of electric load using machine learning models. In: SPRINGER. **International conference on artificial intelligence and soft computing**. [S.l.], 2018. p. 148–159.

MAY, Robert; DANDY, Graeme; MAIER, Holger. Review of input variable selection methods for artificial neural networks. **Artificial neural networks-methodological advances and biomedical applications**, InTech Rijeka, Croatia, v. 10, n. 1, p. 19–45, 2011.

MAY, Robert J; MAIER, Holger R; DANDY, Graeme C; FERNANDO, TMK Gayani. Non-linear variable selection for artificial neural networks using partial mutual information. **Environmental Modelling & Software**, Elsevier, v. 23, n. 10-11, p. 1312–1326, 2008.

MCCARTNEY, Matthew; BEILFUSS, Richard D.; REBELO, Lisa-Maria. Zambezi river basin. In: \_\_\_\_\_. **The Wetland Book: II: Distribution, Description, and Conservation**. Dordrecht: Springer Netherlands, 2018. p. 1217–1232. ISBN 978-94-007-4001-3.

MEHDIZADEH, Saeid; FATHIAN, Farshad; ADAMOWSKI, Jan F. Hybrid artificial intelligence-time series models for monthly streamflow modeling. **Applied Soft Computing**, Elsevier, v. 80, p. 873–887, 2019.

MEISENBACHER, Stefan; PINTER, Tim Martin Janik; HAGENMEYER, Veit; MIKUT, Ralf. Concepts for automated machine learning in smart grid applications. In: **Proceedings-31. Workshop Computational Intelligence: Berlin**. [S.l.: s.n.], 2021. v. 25, p. 26.

MESHRAM, Sarita Gajbhiye; GHORBANI, Mohmmmad Ali; SHAMSHIRBAND, Shahaboddin; KARIMI, Vahid; MESHRAM, Chandrashekhar. River flow prediction using hybrid psogsa algorithm based on feed-forward neural network. **Soft Computing**, Springer, v. 23, p. 10429–10438, 2019.

MIRANDA, Nathallia Mercedes; CATALDI, Marcio; SILVA, Felipe das Neves Roque da. Simulação do regime hidrológico da cabeceira do rio são francisco a partir da utilização dos modelos smap e regcm. **Anuário do Instituto de Geociencias**, v. 40, n. 3, p. 328–339, 2019.

MIRJALILI, Seyedali. How effective is the grey wolf optimizer in training multi-layer perceptrons. **Applied Intelligence**, Springer, v. 43, n. 1, p. 150–161, 2015.

MIRJALILI, Seyedali; MIRJALILI, Seyed Mohammad; LEWIS, Andrew. Grey wolf optimizer. **Advances in Engineering Software**, v. 69, p. 46–61, 2014. ISSN 0965-9978.

MUAIEVELA, Nordino Martinho. **GESTÃO DE RECURSOS HÍDRICOS DE BACIAS DE RIOS TRANSFRONTEIRIÇOS: ANÁLISE DO GRAU DE ATENDIMENTO À DEMANDA DE ÁGUA NA BACIA DO RIO INCOMÁTI EM MOÇAMBIQUE**. 2011. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2011.

MUNGÓI, Cláudio Artur. A produção de energia: usos políticos e econômicos do território e desenvolvimento rural no vale do zambeze, moçambique-a barragem de cahora bassa em questão. **GEOUSP Espaço e Tempo (Online)**, v. 15, n. 2, p. 77–84, 2011.

MUSHTAQ, Haris; AKHTAR, Taimoor; HASHMI, Muhammad Zia-ur Rahman; MASOOD, Amjad. Hydrologic interpretation of machine learning models for 10-daily streamflow simulation in climate sensitive upper indus catchments. **Hydrology and Earth System Sciences Discussions**, Copernicus GmbH, p. 1–32, 2022.

NETO, Ricardo Krauskopf; GUETTER, Alexandre K; MINE, Miriam Rita Moro. Modelo hidrológico com atualização de estado parte i: Modelagem hidrológica. **Revista Brasileira de Recursos Hídricos**, v. 12, p. 91–103, 2007.

NGUYEN, Duc Hai; LE, Xuan Hien; HEO, Jae-Yeong; BAE, Deg-Hyo. Development of an extreme gradient boosting model integrated with evolutionary algorithms for hourly water level prediction. **IEEE Access**, IEEE, v. 9, p. 125853–125867, 2021.

NGUYEN, Hieu; NGUYEN, Ngoc-Mai; CAO, Minh-Tu; HOANG, Nhat-Duc; TRAN, Xuan-Linh. Prediction of long-term deflections of reinforced-concrete members using a novel swarm optimized extreme gradient boosting machine. **Engineering with Computers**, Springer, v. 38, n. 2, p. 1255–1267, 2022.

NGUYEN, Huu Duy. Daily streamflow forecasting by machine learning in tra khuc river in vietnam. **Science of the Earth**, 2022.

NI, Lingling; WANG, Dong; SINGH, Vijay P.; WU, Jianfeng; WANG, Yuankun; TAO, Yuwei; ZHANG, Jianyun. Streamflow and rainfall forecasting by two long short-term memory-based models. **Journal of Hydrology**, v. 583, p. 124296, 2020. ISSN 0022-1694.

NI, Lingling; WANG, Dong; WU, Jianfeng; WANG, Yuankun; TAO, Yuwei; ZHANG, Jianyun; LIU, Jiufu. Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. **Journal of Hydrology**, Elsevier, v. 586, p. 124901, 2020.

NI, Lingling; WANG, Dong; WU, Jianfeng; WANG, Yuankun; TAO, Yuwei; ZHANG, Jianyun; LIU, Jiufu. Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. **Journal of Hydrology**, Elsevier, v. 586, p. 124901, 2020.

NILESH, Rathod; SUNIL, Wankhade. Improving extreme learning machine through optimization a review. In: IEEE. **2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)**. [S.l.], 2021. v. 1, p. 906–912.

NIU, Peifeng; NIU, Songpeng; LIU, Nan; CHANG, Lingfang. The defect of the grey wolf optimization algorithm and its verification method. **Knowledge-Based Systems**, v. 171, p. 37–43, 2019. ISSN 0950-7051.

NIU, Wen-jing; FENG, Zhong-kai. Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. **Sustainable Cities and Society**, Elsevier, v. 64, p. 102562, 2021.

NOURANI, Vahid; BAGHANAM, Aida Hosseini; ADAMOWSKI, Jan; KISI, Ozgur. Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. **Journal of Hydrology**, Elsevier, v. 514, p. 358–377, 2014.

OCHOA-TOCACHI, Boris F; BUYTAERT, Wouter; BIÈVRE, Bert De. Regionalization of land-use impacts on streamflow using a network of paired catchments. **Water Resources Research**, Wiley Online Library, v. 52, n. 9, p. 6710–6729, 2016.

OLIVEIRA, Andréia Fróes Galuci. Testes estatísticos para comparação de médias. **Revista Eletrônica Nutritime**, v. 5, n. 6, p. 777–788, 2008.

OLIVEIRA, Ana Paula Nunes de; HENKES, Jairo Afonso. Condomínios sustentáveis: desafios da escassez dos recursos naturais. **Revista Gestão & Sustentabilidade Ambiental**, v. 4, n. 2, p. 602–625, 2015.

OSABA, Eneko; VILLAR-RODRIGUEZ, Esther; SER, Javier Del; NEBRO, Antonio J; MOLINA, Daniel; LATORRE, Antonio; SUGANTHAN, Ponnuthurai N; COELLO, Carlos A Coello; HERRERA, Francisco. A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems. **Swarm and Evolutionary Computation**, Elsevier, v. 64, p. 100888, 2021.

OSMAN, Ahmedbahaaaldin Ibrahim Ahmed; AHMED, Ali Najah; CHOW, Ming Fai; HUANG, Yuk Feng; EL-SHAFIE, Ahmed. Extreme gradient boosting (xgboost) model to predict the groundwater levels in selangor malaysia. **Ain Shams Engineering Journal**, Elsevier, v. 12, n. 2, p. 1545–1556, 2021.

PAIXÃO, Rafael Christian Fonseca da; PENIDO, Rúben El-Katib; CURY, Alexandre Abrahão; MENDES, Júlia Castro. Comparison of machine learning techniques to predict the compressive strength of concrete and considerations on model generalization. **Revista IBRACON de Estruturas e Materiais**, SciELO Brasil, v. 15, 2022.

PALERMO, Stefania Anna; MAIOLO, Mario; BRUSCO, Anna Chiara; TURCO, Michele; PIROUZ, Behrouz; GRECO, Emilio; SPEZZANO, Giandomenico; PIRO, Patrizia. Smart technologies for water resource management: An overview. **Sensors**, MDPI, v. 22, n. 16, p. 6225, 2022.

PARISOIJ, Peiman; MOHEBZADEH, Hamid; LEE, Taesam. Employing machine learning algorithms for streamflow prediction: A case study of four river basins with different climatic zones in the united states. **Water Resources Management**, v. 34, n. 13, p. 4113–4131, Oct 2020. ISSN 1573-1650.

PARMEZAN, Antonio Rafael Sabino; BATISTA, Gustavo Enrique de Almeida Prado Alves. Descrição de modelos estatísticos e de aprendizado máquina para predição de séries temporais. 2016.

PARMEZAN, Antonio Rafael Sabino; SOUZA, Vinicius MA; BATISTA, Gustavo EAPA. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. **Information sciences**, Elsevier, v. 484, p. 302–337, 2019.

PAZ, Adriano Rolim da. Hidrologia aplicada. 2004.

PEDREGOSA, Fabian; VAROQUAUX, Gaël; GRAMFORT, Alexandre; MICHEL, Vincent; THIRION, Bertrand; GRISEL, Olivier; BLONDEL, Mathieu; PRETTENHOFER, Peter; WEISS, Ron; DUBOURG, Vincent et al. Scikit-learn: Machine learning in python. **The Journal of Machine Learning Research**, JMLR. org, v. 12, p. 2825–2830, 2011.

PEREIRA, Daniel Pena; LIMA, Julião Soares de Souza; XAVIER, Alexandre Cândido; PASSOS, Renato Ribeiro; FIEDLER, Nilton César. Aplicação do diagrama de taylor para avaliação de interpoladores espaciais em atributos de solo em cultivo com eucalipto. **Revista Árvore**, SciELO Brasil, v. 38, p. 899–905, 2014.

PEREIRA, Matheus Machado. Protótipo de uma plataforma de coleta de dados hidrológicos. 2022.

PEREIRA, RS; NETO, AS; TUCCI, CEM. Princípios da hidrometria. **Texto de apoio da disciplina de Hidrologia I. Programa de Pós-Graduação em Recursos**

**Hídricos e Saneamento Ambiental. Instituto de Pesquisas Hidráulicas. Universidade Federal do Rio Grande do Sul, 2003.**

PHAM, Binh Thai; NGUYEN-THOI, Trung; LY, Hai-Bang; NGUYEN, Manh Duc; AL-ANSARI, Nadhir; TRAN, Van-Quan; LE, Tien-Think. Extreme learning machine based prediction of soil shear strength: a sensitivity analysis using monte carlo simulations and feature backward elimination. **Sustainability**, MDPI, v. 12, n. 6, p. 2339, 2020.

PHAM, Quoc Bao; ABBA, Sani Isah; USMAN, Abdullahi Garba; LINH, Nguyen Thi Thuy; GUPTA, Vivek; MALIK, Anurag; COSTACHE, Romulus; VO, Ngoc Duong; TRI, Doan Quang. Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. **Water Resources Management**, Springer, v. 33, n. 15, p. 5067–5087, 2019.

PHAN, Thi-Thu-Hong; NGUYEN, Xuan Hoai. Combining statistical machine learning models with ARIMA for water level forecasting: The case of the Red river. **Advances in Water Resources**, v. 142, p. 103656, 2020. ISSN 0309-1708.

PINI, Michele; SCALVINI, Andrea; LIAQAT, Muhammad Usman; RANZI, Roberto; SERINA, Ivan; MEHMOOD, Tahir. Evaluation of Machine Learning Techniques for Inflow Prediction in Lake Como, Italy. **Procedia Computer Science**, v. 176, p. 918–927, 2020. ISSN 1877-0509. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 24th International Conference KES2020.

PINTO, Elis. Geopolítica da água. **Revista de Geopolítica**, v. 8, n. 1, p. 19–32, 2017.

POPOCA, Esperanza Nicolás et al. **Un método de regresión bayesiana para selección genómica**. 2013. Dissertação (Mestrado) — México, 2013.

PORTELA, Maria Manuela; QUINTELA, AC; SANTOS, João Filipe; VAZ, C; MARTINS, C. Tendências em séries temporais de variáveis hidrológicas. Associação Portuguesa dos Recursos Hídricos, 2011.

POVAK, Nicholas A; HESSBURG, Paul F; REYNOLDS, Keith M; SULLIVAN, Timothy J; MCDONNELL, Todd C; SALTER, R Brion. Machine learning and hurdle models for improving regional predictions of stream water acid neutralizing capacity. **Water Resources Research**, Wiley Online Library, v. 49, n. 6, p. 3531–3546, 2013.

PUMA-VILLANUEVA, Wilfredo J; SANTOS, Eurípedes P dos; ZUBEN, Fernando J Von. Data partition and variable selection for time series prediction using wrappers. In: IEEE. **The 2006 IEEE International Joint Conference on Neural Network Proceedings**. [S.l.], 2006. p. 4740–4747.

QUILTY, John; ADAMOWSKI, Jan; KHALIL, Bahaa; RATHINASAMY, Maheswaran. Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. **Water Resources Research**, v. 52, n. 3, p. 2299, 2016.

RADHIKA, Y; SHASHI, M. Atmospheric temperature prediction using support vector machines. **International journal of computer theory and engineering**, IACSIT Press, v. 1, n. 1, p. 55, 2009.

- RAJ, Nawin; GHARINEIAT, Zahra. Evaluation of multivariate adaptive regression splines and artificial neural network for prediction of mean sea level trend around northern australian coastlines. **Mathematics**, MDPI, v. 9, n. 21, p. 2696, 2021.
- RAMSON, SR Jino; RAJU, K Lova; VISHNU, S; ANAGNOSTOPOULOS, Theodoros. Nature inspired optimization techniques for image processing—a short review. **Nature Inspired Optimization Techniques for Image Processing Applications**, Springer, p. 113–145, 2019.
- RAMSON, SR Jino; RAJU, K Lova; VISHNU, S; ANAGNOSTOPOULOS, Theodoros. Nature inspired optimization techniques for image processing—a short review. **Nature Inspired Optimization Techniques for Image Processing Applications**, Springer, p. 113–145, 2019.
- RASOULI, Kabir; HSIEH, William W; CANNON, Alex J. Daily streamflow forecasting by machine learning methods with weather and climate inputs. **Journal of Hydrology**, Elsevier, v. 414, p. 284–293, 2012.
- RATHOD, Nilesh; WANKHADE, Sunil. Review of optimization in improving extreme learning machine. **EAI Endorsed Transactions on Industrial Networks and Intelligent Systems**, v. 8, n. 28, p. e2–e2, 2021.
- REN, Kun; FANG, Wei; QU, Jihong; ZHANG, Xia; SHI, Xiaoyu. Comparison of eight filter-based feature selection methods for monthly streamflow forecasting—three case studies on camels data sets. **Journal of Hydrology**, Elsevier, v. 586, p. 124897, 2020.
- REZAIE-BALF, Mohammad; NAGANNA, Sujay Raghavendra; KISI, Ozgur; EL-SHAFIE, Ahmed. Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of aswan high dam. **Hydrological Sciences Journal**, Taylor & Francis, v. 64, n. 13, p. 1629–1646, 2019.
- RIAHI-MADVAR, Hossien; DEHGHANI, Majid; MEMARZADEH, Rasoul; GHARABAGHI, Bahram. Short to long-term forecasting of river flows by heuristic optimization algorithms hybridized with anfis. **Water Resources Management**, Springer, v. 35, p. 1149–1166, 2021.
- RIBEIRO, Victor Henrique Alves; REYNOSO-MEZA, Gilberto; SIQUEIRA, Hugo Valadares. Multi-objective ensembles of echo state networks and extreme learning machines for streamflow series forecasting. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 95, p. 103910, 2020.
- RIZVON, Shamili Syed; JAYAKUMAR, Karthikeyan. Strength prediction models for recycled aggregate concrete using random forests, ann and lasso. **Journal of Building Pathology and Rehabilitation**, Springer, v. 7, n. 1, p. 1–10, 2022.
- ROSENDO, Matheus. Um algoritmo de otimização por nuvem de partículas para resolução de problemas combinatórios. 2010.
- ROSS, Brian C. Mutual information between discrete and continuous data sets. **PloS one**, Public Library of Science San Francisco, USA, v. 9, n. 2, p. e87357, 2014.
- ROUGIER, Nicolas. **Matplotlib tutorial**. 2012. Tese (Doutorado) — INRIA, 2012.

ROZA, Felipe Schmoeller da et al. Aprendizagem de máquina para apoio à tomada de decisão em vendas do varejo utilizando registros de vendas. Florianópolis, SC., 2016.

SÁ, Lucas Amorim de. Utilização de imagens de satélite para determinação das curvas cota x área x volume do açude itans, localizado no município de Caicó, rio grande do norte, brasil. 2018.

SAHOO, Abinash; SAMANTARAY, Sandeep; GHOSE, Dillip K. Multilayer perceptron and support vector machine trained with grey wolf optimiser for predicting floods in barak river, india. **Journal of Earth System Science**, Springer, v. 131, n. 2, p. 1–23, 2022.

SAHOUR, Hossein; GHOLAMI, Vahid; TORKAMAN, Javad; VAZIFEDAN, Mehdi; SAEEDI, Sirwe. Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. **Environmental Earth Sciences**, Springer, v. 80, n. 22, p. 1–14, 2021.

SALES, João Marcos de Jesus. Monitoramento e modelagem hidrológica em uma bacia hidrográfica no semiárido do brasil. Pós-Graduação em Recursos Hídricos, 2019.

SAMANTARAY, Sandeep; SAHOO, Abinash. A comparative study on prediction of monthly streamflow using hybrid anfis-pso approaches. **KSCE Journal of Civil Engineering**, Springer, v. 25, n. 10, p. 4032–4043, 2021.

SANTOS, Carlos Eduardo da Silva. Seleção de parâmetros de máquinas de vetores de suporte usando otimização multiobjetivo baseada em meta-heurísticas. 2019.

SANTOS, Eulanda Miranda dos et al. Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência. Universidade Federal de Campina Grande, 2002.

SAPITANG, Michelle; RIDWAN, Wanie M; KUSHIAR, Khairul Faizal; AHMED, Ali Najah; EL-SHAFIE, Ahmed. Machine learning application in reservoir water level forecasting for sustainable hydropower generation strategy. **Sustainability**, Multidisciplinary Digital Publishing Institute, v. 12, n. 15, p. 6121, 2020.

SARAIVA, Samuel Vitor; CARVALHO, Frede de Oliveira; SANTOS, Celso Augusto Guimarães; BARRETO, Lucas Costa; FREIRE, Paula Karenina de Macedo Machado. Daily streamflow forecasting in sobradinho reservoir using machine learning models coupled with wavelet transform and bootstrapping. **Applied Soft Computing**, Elsevier, v. 102, p. 107081, 2021.

SATTAR, Ahmed MA; GHARABAGHI, Bahram. Gene expression models for prediction of longitudinal dispersion coefficient in streams. **Journal of Hydrology**, Elsevier, v. 524, p. 587–596, 2015.

SCIENTIFIC, Campbell. **CR800/CR850 measurement and control system**. [S.l.], 2010.

SHIELDS, Joe. A review of monte carlo methods and their application in medical physics for simulating radiation transport. 2022.

SHINDE, Pramila P; SHAH, Seema. A review of machine learning and deep learning applications. In: IEEE. **2018 Fourth international conference on computing communication control and automation (ICCUBEA)**. [S.l.], 2018. p. 1–6.

SILVA, Allan F; LEMONGE, AC; LIMA, Beatriz S. Algoritmo de otimização com enxame de partículas auxiliado por metamodelos. **XI Simpósio de Mecânica Computacional, II Encontro Mineiro de Modelagem Computacional, SIMMEC/EMMCOMP**, 2014.

SILVA, Luciene Pimentel da; EWEN, John. Modelagem hidrológica de grandes bacias hidrográficas: A necessidade de novas metodologias. 2000.

SIQUEIRA, Hugo; MACEDO, Mariana; TADANO, Yara de Souza; ALVES, Thiago Antonini; JR, Sergio L Stevan; JR, Domingos S Oliveira; MARINHO, Manoel HN; NETO, Paulo SG de Mattos; OLIVEIRA, João FL de; LUNA, Ivette et al. Selection of temporal lags for predicting riverflow series from hydroelectric plants using variable selection methods. **Energies**, MDPI, v. 13, n. 16, p. 4236, 2020.

SMOLA, Alex J; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. **Statistics and computing**, Springer, v. 14, n. 3, p. 199–222, 2004.

SNIEDER, Everett; SHAKIR, R; KHAN, UT. A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. **Journal of Hydrology**, Elsevier, v. 583, p. 124299, 2020.

SOLOMATINE, Dimitri P; XUE, Yunpeng. M5 model trees and neural networks: application to flood forecasting in the upper reach of the huai river in china. **Journal of Hydrologic Engineering**, American Society of Civil Engineers, v. 9, n. 6, p. 491–501, 2004.

SONG, Jiancai; ZHANG, Liyi; XUE, Guixiang; MA, YunPeng; GAO, Shan; JIANG, QingLing. Predicting hourly heating load in a district heating system based on a hybrid cnn-lstm model. **Energy and Buildings**, Elsevier, v. 243, p. 110998, 2021.

SOUZA, Danilo PM; MARTINHO, Alfeu D; ROCHA, Caio C; CHRISTO, Eliane da S; GOLIATT, Leonardo. Group method of data handling to forecast the daily water flow at the cahora bassa dam. **Acta Geophysica**, Springer, p. 1–13, 2022.

SOUZA, Rodrigo Clemente Thom De. Previsão de séries temporais utilizando rede neural treinada por filtro de kalman e evolução diferencial. Curitiba: Pontifícia Universidade Católica do Paraná. Tese de doutorado, 2008.

SUN, Alexander Y.; WANG, Dingbao; XU, Xianli. Monthly streamflow forecasting using gaussian process regression. **Journal of Hydrology**, v. 511, p. 72 – 81, 2014. ISSN 0022-1694.

SUN, Shiliang; CAO, Zehui; ZHU, Han; ZHAO, Jing. A survey of optimization methods from a machine learning perspective. **IEEE transactions on cybernetics**, IEEE, v. 50, n. 8, p. 3668–3681, 2019.

TAYLAN, Pakize; WEBER, Gerhard-Wilhelm. Multivariate adaptive regression spline and continuous optimization for modern applications in science, economy and technology. **METU Institute of Applied Mathematics, Preprints**, n. 2007-09, 2007.

TEODORO, Anderson; SILVA, Jhonatan Barbosa da; MACHADO, Daniel Okamoto; NOBUYOSHI, Carlos; GEALH, Batista Pereira. O curso de capacitação em hidrometria para gestão de recursos hídricos (hidrotec-ufms) e resultados de algumas atividades. 2011.

TEUTSCHBEIN, Claudia; GRABS, Thomas; LAUDON, Hjalmar; KARLSEN, Reinert H; BISHOP, Kevin. Simulating streamflow in ungauged basins under a changing climate: The importance of landscape characteristics. **Journal of Hydrology**, Elsevier, v. 561, p. 160–178, 2018.

THA, Daniel; SEAGER, Daniel. Análise estendida de custo-benefício para adoção de ação ecológica em moçambique: Compartilhando o rio zambeze. **Planejamento e Políticas Públicas**, n. 40, 2013.

TIBSHIRANI, R.J. Regression shrinkage and selection via the lasso. **JOURNAL OF THE ROYAL STATISTICAL SOCIETY: SERIES B (STATISTICAL METHODOLOGY)**, 2011.

TIKHAMARINE, Yazid; SOUAG-GAMANE, Doudja; AHMED, Ali Najah; KISI, Ozgur; EL-SHAFIE, Ahmed. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey wolf optimization (gwo) algorithm. **Journal of Hydrology**, Elsevier, v. 582, p. 124435, 2020.

TIKHAMARINE, Yazid; SOUAG-GAMANE, Doudja; KISI, Ozgur. A new intelligent method for monthly streamflow prediction: hybrid wavelet support vector regression based on grey wolf optimizer (wsvr-gwo). **Arabian Journal of Geosciences**, Springer, v. 12, n. 17, p. 1–20, 2019.

TONGAL, Hakan; BOOIJ, Martijn J. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. **Journal of hydrology**, Elsevier, v. 564, p. 266–282, 2018.

TUCCI, CEM; HIDROLÓGICOS, Modelos. Porto alegre, ed. **Universidade/UFRGS/ABRH**, 1998.

TUCCI, Carlos EM et al. Hidrologia: Ciência e aplicação, 3ª edição. **Porto Alegre, Editora da UFRGS/ABRH**, 2004.

TUKEY, John W. Comparing individual means in the analysis of variance. **Biometrics**, JSTOR, p. 99–114, 1949.

TUMBARE, Michael J; AUTHORITY, Zambezi River. The zambezi river: Its threats and opportunities. In: **Zambezi River: It's Threats and Opportunities, 7th River Symposium**. [S.l.: s.n.], 2004. v. 1, n. 3.

ULIANA, Eduardo Morgan; PEREIRA, Donizete dos Reis; SILVA, Demetrius David da; ALMEIDA, Frederico Terra de; SOUZA, Adilson Pacheco de. Modelagem chuva-vazão em uma bacia tropical utilizando o modelo iph ii. **Geo Uerj**, n. 33, p. 30557, 2018.

UNNIKRISHNAN, Poornima; JOTHIPRAKASH, V. Daily rainfall forecasting for one year in a single run using singular spectrum analysis. **Journal of Hydrology**, Elsevier, v. 561, p. 609–621, 2018.

VANEGAS, Jairo; VÁSQUEZ, Fabián. Multivariate adaptative regression splines (mars), una alternativa para el análisis de series de tiempo. **Gaceta sanitaria**, SciELO Public Health, v. 31, p. 235–237, 2017.

VAPNIK, Vladimir; GOLOWICH, Steven E; SMOLA, Alex et al. Support vector method for function approximation, regression estimation, and signal processing. **Advances in neural information processing systems**, Morgan Kaufmann Publishers, p. 281–287, 1997.

VAROQUAUX, Gaël; RAAMANA, Pradeep Reddy; ENGEMANN, Denis A.; HOYOS-IDROBO, Andrés; SCHWARTZ, Yannick; THIRION, Bertrand. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. **NeuroImage**, v. 145, p. 166–179, 2017. ISSN 1053-8119. Individual Subject Prediction.

VATANCHI, Sajjad M; ETEMADFARD, Hossein; MAGHREBI, Mahmoud Faghfour; SHAD, Rouzbeh. A comparative study on forecasting of long-term daily streamflow using ann, anfis, bilstm, and cnn-gru-lstm. 2022.

VENKATESAN, Ellakkia; MAHINDRAKAR, Amit B. Forecasting floods using extreme gradient boosting-a new approach. **International Journal of Civil Engineering and Technology**, v. 10, n. 2, p. 1336–1346, 2019.

VIANA, Jussara Freire de Souza; MONTENEGRO, Suzana Maria Gico Lima; SILVA, Bernardo Barbosa da; SILVA, Richarde Marques da; SOUSA, Wanderson dos Santos. Modelagem hidrológica da bacia hidrográfica do rio pirapama-pe utilizando o modelo swat. **Journal of Environmental Analysis and Progress**, p. 155–172, 2018.

VILLANUEVA, Wilfredo Jaime Puma. Comitê de máquinas em predição de séries temporais. In: **Mestrado em Engenharia Elétrica e Computação**. [S.l.]: Faculdade de Engenharia Elétrica e de Computação, 2006.

VILLELA, Swami Marcondes; MATTOS, Arthur. **Hidrologia aplicada**. [S.l.]: Editora McGraw-Hill do Brasil, 1975.

WANG, Jianjun; LI, Li; NIU, Dongxiao; TAN, Zhongfu. An annual load forecasting model based on support vector regression with differential evolution algorithm. **Applied Energy**, Elsevier, v. 94, p. 65–70, 2012.

WANG, Jian; LU, Siyuan; WANG, Shui-Hua; ZHANG, Yu-Dong. A review on extreme learning machine. **Multimedia Tools and Applications**, Springer, p. 1–50, 2021.

WANG, Maofa; REZAIE-BALF, Mohammad; NAGANNA, Sujay Raghavendra; YASEEN, Zaher Mundher. Sourcing chirps precipitation data for streamflow forecasting using intrinsic time-scale decomposition based machine learning models. **Hydrological Sciences Journal**, Taylor & Francis, v. 66, n. 9, p. 1437–1456, 2021.

WANG, Weizeng; SHI, Yuliang; LYU, Gaofan; DENG, Wanghua. Electricity consumption prediction using xgboost based on discrete wavelet transform. **DEStech Transactions on Computer Science and Engineering**, 2017.

WANG, Wen-Chuan; CHAU, Kwok-Wing; CHENG, Chun-Tian; QIU, Lin. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. **Journal of Hydrology**, v. 374, n. 3, p. 294 – 306, 2009. ISSN 0022-1694.

WEGAYEHU, Eyob Betru; MULUNEH, Fiseha Behulu. Multivariate streamflow simulation using hybrid deep learning models. **Computational Intelligence and Neuroscience**, Hindawi, v. 2021, 2021.

- WHITE, Elaheh. **Predicting Unimpaired Flow in Ungauged Basins: "Random Forests" Applied to California Streams**. [S.l.]: University of California, Davis, 2017.
- WHITLEY, Darrell. A genetic algorithm tutorial. **Statistics and computing**, Springer, v. 4, n. 2, p. 65–85, 1994.
- WU, Lifeng; ZHOU, Hanmi; MA, Xin; FAN, Junliang; ZHANG, Fucang. Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: Application in contrasting climates of china. **Journal of Hydrology**, Elsevier, v. 577, p. 123960, 2019.
- XUE, Lianqing; YANG, Fan; YANG, Changbing; WEI, Guanghui; LI, Wenqian; HE, Xinlin. Hydrological simulation and uncertainty analysis using the improved topmodel in the arid manas river basin, china. **Scientific reports**, Nature Publishing Group, v. 8, n. 1, p. 1–12, 2018.
- YANG, Kai; JIN, Huaiping; CHEN, Xiangguang; DAI, Jiayu; WANG, Li; ZHANG, Dongxiang. Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time gaussian process regression models. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 155, p. 170–182, 2016.
- YASEEN, Zaher Mundher; EBTEHAJ, Isa; BONAKDARI, Hossein; DEO, Ravinesh C; MEHR, Ali Danandeh; MOHTAR, Wan Hanna Melini Wan; DIOP, Lamine; EL-SHAFIE, Ahmed; SINGH, Vijay P. Novel approach for streamflow forecasting using a hybrid anfis-ffa model. **Journal of Hydrology**, Elsevier, v. 554, p. 263–276, 2017.
- YASEEN, Zaher Mundher; FARIS, Hossam; AL-ANSARI, Nadhir. Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application. **Complexity**, Hindawi, v. 2020, 2020.
- YU, Xiang; WANG, Yuhao; WU, Lifeng; CHEN, Genhua; WANG, Lei; QIN, Hui. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. **Journal of Hydrology**, Elsevier, v. 582, p. 124293, 2020.
- ZAFAR, Ayesha; SHAH, Samia; KHALID, Rabiya; HUSSAIN, Sardar Mehboob; RAHIM, Hassan; JAVAID, Nadeem. A meta-heuristic home energy management system. In: **IEEE. 2017 31st international conference on advanced information networking and applications workshops (WAINA)**. [S.l.], 2017. p. 244–250.
- ZAINI, Nuratiah; MALEK, MA; YUSOFF, M; MARDI, NH; NORHISHAM, S. Daily river flow forecasting with hybrid support vector machine–particle swarm optimization. In: **IOP PUBLISHING. IOP Conference Series: Earth and Environmental Science**. [S.l.], 2018. v. 140, n. 1, p. 012035.
- ZHANG, G Peter. Time series forecasting using a hybrid arima and neural network model. **Neurocomputing**, Elsevier, v. 50, p. 159–175, 2003.
- ZHANG, Hongbo; SINGH, Vijay P; WANG, Bin; YU, Yinghao. Ceref: A hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. **Journal of hydrology**, Elsevier, v. 540, p. 246–256, 2016.

ZHANG, WG; GOH, Anthony Teck Chee. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. **Computers and Geotechnics**, Elsevier, v. 48, p. 82–95, 2013.

ZHANG, Yong; SONG, Xian fang; GONG, Dun wei. A return-cost-based binary firefly algorithm for feature selection. **Information Sciences**, v. 418-419, p. 561 – 574, 2017. ISSN 0020-0255.

ZHAO, Ziyue; YANG, Wen; ZHAI, Yixiao; LIANG, Yingjian; ZHAO, Yuming. Identify dna-binding proteins through the extreme gradient boosting algorithm. **Frontiers in Genetics**, Frontiers Media SA, v. 12, p. 821996, 2022.

ZHOU, Yong; LIU, Yanfeng; WANG, Dengjia; LIU, Xiaojun. Comparison of machine-learning models for predicting short-term building heating load using operational parameters. **Energy and Buildings**, Elsevier, v. 253, p. 111505, 2021.

ZHU, Shuang; LUO, Xiangang; XU, Zhanya; YE, Lei. Seasonal streamflow forecasts using mixture-kernel gpr and advanced methods of input variable selection. **Hydrology Research**, IWA Publishing, v. 50, n. 1, p. 200–214, 2019.

ZOU, Hui; HASTIE, Trevor. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005.