

PGCOMP - Programa de Pós-Graduação em Ciência da Computação
Universidade Federal da Bahia (UFBA)
Av. Milton Santos, s/n - Ondina
Salvador, BA, Brasil, 40170-110

<https://pgcomp.ufba.br>
pgcomp@ufba.br

A água é essencial para a vida humana e prever sua crescente demanda é um grande desafio, o objetivo deste trabalho é propor um novo modelo híbrido, SVR-ANN, para previsão de demanda de água, onde uma adaptação da metodologia proposta por Zhang (2003) é utilizada para decompor as séries temporais de 10 reservatórios que abastecem a Região Metropolitana de Salvador (RMS). Os dados utilizados são do consumo histórico de janeiro/2017 a fevereiro/2022, em litros por hora, obtidos junto à empresa de abastecimento local, Empresa Baiana de Águas e Saneamento (EMBASA), e dados meteorológicos obtidos junto ao Instituto Nacional de Meteorologia do Brasil (INMET). Os resultados demonstraram a viabilidade do uso do modelo proposto, comparado a outros modelos tradicionais como o *Multilayer Perceptron* (MLP), *Support Vector Regression* (SVR), *Short Long Term Memory* (LSTM) e *Autoregressive and Integrated Moving Average* (ARIMA).

Palavras-chave: Séries Temporais, Aprendizagem de Máquina, Modelos Híbridos, demanda de água, dados meteorológicos.

MSC | 160 | 2023

Modelos de aprendizagem de máquina para previsão da demanda de água da
Região Metropolitana de Salvador, Bahia

Edmilson dos Santos de
Jesus

Modelos de aprendizagem de máquina para previsão da demanda de água da Região Metropolitana de Salvador, Bahia

Edmilson dos Santos de Jesus

Dissertação de Mestrado

Universidade Federal da Bahia

Programa de Pós-Graduação em
Ciência da Computação

Julho | 2023

UFBA





Universidade Federal da Bahia
Instituto de Computação

Programa de Pós-Graduação em Ciência da Computação

**MODELOS DE APRENDIZAGEM DE
MÁQUINA PARA PREVISÃO DA DEMANDA
DE ÁGUA DA REGIÃO METROPOLITANA
DE SALVADOR, BAHIA**

Edmilson dos Santos de Jesus

DISSERTAÇÃO DE MESTRADO

Salvador
31 de julho de 2023

EDMILSON DOS SANTOS DE JESUS

**MODELOS DE APRENDIZAGEM DE MÁQUINA PARA
PREVISÃO DA DEMANDA DE ÁGUA DA REGIÃO
METROPOLITANA DE SALVADOR, BAHIA**

Esta Dissertação de Mestrado foi apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientadora: Profa. Dra. Gecynalda Soares da Silva Gomes

Salvador
31 de julho de 2023

Sistema de Bibliotecas - UFBA

C871 Jesus, Edmilson dos Santos de.

Modelos de aprendizagem de máquina para previsão da demanda de água da Região Metropolitana de Salvador, Bahia / Edmilson dos Santos de Jesus – Salvador, 2023.

56p.: il.

Orientadora: Prof. Dr. Profa. Dra. Gecynalda Soares da Silva Gomes.

Dissertação (Mestrado) – Universidade Federal da Bahia, Instituto de Computação, 2023.

1. Análise de Séries Temporais. 2. Meteorologia. 3. Água. I. Gomes, Gecynalda Soares da Silva. II. Universidade Federal da Bahia. Instituto de Computação. III Título.

CDU – 219.246



“MODELOS DE APRENDIZAGEM DE MÁQUINA PARA PREVISÃO DA DEMANDA DE ÁGUA DA REGIÃO METROPOLITANA DE SALVADOR, BAHIA”

EDMILSON DOS SANTOS DE JESUS

Dissertação apresentada ao Colegiado do Programa de Pós-Graduação em Ciência da Computação na Universidade Federal da Bahia, como requisito parcial para obtenção do Título de Mestre em Ciência da Computação.

Banca Examinadora



Documento assinado digitalmente
GECYNALDA SOARES DA SILVA GOMES
Data: 01/08/2023 11:25:45-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Gecynalda Soares da Silva Gomes (Orientadora-UFBA)



Documento assinado digitalmente
PAULO HENRIQUE FERREIRA DA SILVA
Data: 31/07/2023 18:24:07-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Paulo Henrique Ferreira da Silva (UFBA)



Documento assinado digitalmente
ANDERSON LUIZ ARA SOUZA
Data: 01/08/2023 09:45:40-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Anderson Luiz Ara Souza (UFPR)

Ao meu pai José Pimentel de Jesus que com toda a sua humildade e sabedoria, dedicou sua vida para a família e nos ensinou os valores morais e o caminho de Deus.

AGRADECIMENTOS

Todo projeto, pela definição da palavra tem início, meio e fim. Quando iniciei o projeto do Mestrado tinha em mente essa definição e no horizonte a expectativa de muito aprendizado e da conclusão. Durante a escrita desses agradecimentos me dei conta de que essa jornada não termina aqui graças à contribuição de muitas pessoas, pois um universo de possibilidades se abre à minha frente, com inúmeras oportunidades, para atender uma sociedade que anseia, cada vez mais e mais, por contribuições científicas e avanços tecnológicos. Por isso, para que seja possível seguir em frente, preciso agradecer às pessoas extraordinárias que fizeram parte dessa jornada e que contribuíram para a minha formação.

Agradeço inicialmente a Deus, pelo dom da vida e pela perfeição da criação, em todos os lugares que olho vejo e sinto a presença do criador, sou grato por ele ter me dado o privilégio desse dom e dessa sensibilidade.

Agradeço aos meus pais Maria da Conceição e José Pimentel, que com muito amor, humildade e dedicação criou a mim e minhas duas irmãs Eliani e Erica. Mesmo sem terem concluído o ensino médio foram meus maiores professores, pois sempre me ensinaram o caminho do bem, da humildade e da honestidade. São a razão do meu esforço e da minha jornada.

Agradeço às minhas irmãs Eliani e Erica, por participarem de minha jornada de crescimento desde a infância; elas talvez não saibam mas continuo aprendendo com elas até hoje, sempre que memoro os momentos juntos que vivemos desde o jardim de infância até o colegial.

Agradeço à minha esposa Maria Geniceia, pelo amor incondicional, compreensão e dedicação, ela sempre respeitou os momentos em que eu precisei me afastar para dedicar aos estudos, controlando e ocupando as crianças nos momentos decisivos.

Agradeço aos meus filhos Maria Clara e Aiko Rafael por existirem em minha vida, são os meus tesouros e na companhia deles pude me distrair, relaxar e aliviar o *stress* adquirido ao longo da jornada.

Agradeço a EMBASA (Empresa Baiana de Águas e Saneamento S.A.) e aos colegas de trabalho, gestores, analistas, técnicos e prestadores de serviço, pelo apoio que me foi dado ao longo dos estudos e pelas orientações e informações, os quais foram fundamentais para realização dessa pesquisa. Aqui deixo um agradecimento especial a Denise Britto, Jenner Pitombo e Rita Couto que acreditaram em meu potencial e me incentivaram antes mesmo do início dessa jornada.

Agradeço à turma do OOLAB (antigo Laboratório de Orientação a Objetos da UNEB), em especial a Eduardo Rocha, Carlos Eduardo Martins, Roque Jorge, Felipe Titonel, Erico Ventura e Weber Souza, pelo companheirismo e conhecimentos compartilhados, nos estudos e na vida.

Agradeço aos colegas do Mestrado e Doutorado, Anderson Boa Morte, Beatriz Santana, Cláudio Júnior, Elen Reis, Eliabe Silva, Elisangela Carneiro, Gustavo Antunes, Jonatas Silva, Marcelo Pereira, Uandisson Miranda e outros que tanto contribuíram nas aulas, artigos e trabalhos sempre com discussões de alto nível, de muito conhecimento e troca compartilhada de informações.

Agradeço à professora Gecynalda Soares da Silva Gomes, pelo acolhimento, orientações e direcionamento durante toda a jornada do mestrado, sempre focada e precisa em suas contribuições e ensinamentos, uma pessoa muito especial e vocacionada para sua função sem dúvidas, minha eterna gratidão.

Um agradecimento especial ao eterno professor e mestre Jorge Farias, pela sua extraordinária contribuição em minha vida, ele sempre demonstrou confiança em meu potencial e sempre me incentivou a ir mais longe, desde os primeiros dias da graduação na Universidade Estadual da Bahia.

Por fim, agradeço a todos que, de alguma forma, contribuíram para essa jornada, mesmo que eu possa ter esquecido o nome de alguém que não tenha citado aqui ou tenha perdido o contato; cada pessoa que passa em nossas vidas é única e com certeza deixa sua contribuição.

*A mente que se abre a uma nova ideia
jamais voltará ao seu tamanho original.*

—ALBERT EINSTEIN

RESUMO

A água é essencial para a vida humana e prever sua crescente demanda é um grande desafio, o objetivo deste trabalho é propor um novo modelo híbrido, SVR-ANN, para previsão de demanda de água, onde uma adaptação da metodologia proposta por Zhang (2003) é utilizada para decompor as séries temporais de 10 reservatórios que abastecem a Região Metropolitana de Salvador (RMS). Os dados utilizados são do consumo histórico de janeiro/2017 a fevereiro/2022, em litros por hora, obtidos junto à empresa de abastecimento local, Empresa Baiana de Águas e Saneamento (EMBASA), e dados meteorológicos obtidos junto ao Instituto Nacional de Meteorologia do Brasil (INMET). Os resultados demonstraram a viabilidade do uso do modelo proposto, comparado a outros modelos tradicionais como o *Multilayer Perceptron* (MLP), *Support Vector Regression* (SVR), *Short Long Term Memory* (LSTM) e *Autoregressive and Integrated Moving Average* (ARIMA).

Palavras-chave: Séries Temporais, Aprendizagem de Máquina, Modelos Híbridos, demanda de água, dados meteorológicos.

ABSTRACT

Water is essential for human life and predicting its growing demand is a great challenge, the objective of this work is to propose a new hybrid model, SVR-ANN, for forecasting water demand, where an adaptation of the methodology proposed by Zhang (2003) is used to decompose the time series of 10 reservoirs that supply the Metropolitan Region of Salvador (RMS). The data used are from the historical consumption from January/2017 to February/2022, in liters per hour, received from the local supply company, Empresa Baiana de Águas e Saneamento (EMBASA), and meteorological data obtained from the Instituto Nacional de Meteorologia do Brasil (INMET). The results tolerated the feasibility of using the proposed model, compared to other traditional models such as *Multilayer Perceptron* (MLP), *Support Vector Regression* (SVR), *Short Long Term Memory* (LSTM) and *Autoregressive and Integrated Moving Average* (ARIMA).

Keywords: Time Series, Machine Learning, Hybrid Models, water demand, meteorological data.

SUMÁRIO

Capítulo 1—Introdução	1
1.1 Problema	2
1.2 Motivação	4
1.3 Objetivos	5
1.4 Metodologia	5
1.5 Resultados Alcançados	6
1.6 Estrutura do Texto	7
Capítulo 2—Fundamentação Teórica	9
2.1 Introdução e conceitos básicos	9
Capítulo 3—Trabalhos Relacionados	21
Capítulo 4—Previsão de demanda de água para a Região Metropolitana de Salvador	29
4.1 Proposta	29
4.2 Metodologia	29
4.3 Mapeamento e coleta dos dados	30
4.4 Pré-processamento	31
4.5 Calibragem e treinamento	37
4.6 Predição e coleta dos resultados	39
4.6.1 MLP	39
4.6.2 SVR	40
4.6.3 ARIMA	40
4.6.4 Modelo híbrido SVR-ANN	40
4.7 Análise dos resultados e conclusão	41
Capítulo 5—Resultados alcançados	43
Capítulo 6—Conclusão	49
Referências Bibliográficas	51

Apêndice A—Tabela de trabalhos relacionados

55

LISTA DE FIGURAS

1.1	Sistema Integrado de Abastecimento de Água da Região Metropolitana de Salvador.	3
1.2	Fluxo proposto na metodologia de trabalho	6
2.1	Arquitetura de uma ANN	10
2.2	Arquitetura geral de uma MLP.	11
2.3	Arquitetura de uma rede neural artificial dinâmica.	12
2.4	Configuração de margem de perda para uma SVM baseada em regressão.	14
4.1	Distribuição dos dados meteorológicos inválidos ou ausentes, antes do pré-processamento, por período	31
4.2	Gráfico <i>boxplot</i> para os valores de vazão em litros/hora dos reservatórios.	32
4.3	Histórico de vazão do reservatório R23A, com apenas 2 meses de registros de vazão.	34
4.4	Histórico de vazão dos reservatórios R1, R7, R10, R14, R17, R18, R20, R21, R23B e R25 com dados inválidos, por período.	35
4.5	Histórico de vazão do reservatório R7, com dados inválidos, por período.	36
4.6	Histórico de vazão do reservatório R7, após o pré-processamento, por período.	37
4.7	Histórico de vazão dos reservatórios R1, R7, R10, R14, R17, R18, R20, R21, R23B e R25 após o pré-processamento, por período.	38
4.8	Arquitetura do modelo SVR-ANN.	41
5.1	Previsão horária para o reservatório R23B usando o modelo LSTM com lag de atraso de 3 valores de consumo.	43
5.2	Gráfico <i>boxplot</i> das estatísticas MAPE por modelo implementado.	47

LISTA DE TABELAS

1.1	Lista dos dez principais reservatórios da RMS.	2
4.1	Mínimos, Máximos e Quartis por reservatório.	33
4.2	Média, Mediana, SD, Variância, CV e CA por reservatório.	33
4.3	Combinação das variáveis de entrada usadas para treinamento e teste. . .	39
5.1	Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para a previsão horária.	44
5.2	Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para previsão diária.	45
5.3	Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para previsão semanal.	46
5.4	Resultados das menores estatísticas MAPE por modelo na previsão horária de cada reservatório (RSVR).	46
5.5	Tabela de diferenças entre modelos, com estatísticas de teste e p -valor respectivamente calculados para cada par.	47
A.1	Trabalhos relacionados por métrica de avaliação e técnica utilizada. . . .	56

LISTA DE SIGLAS

1D-CNN	Rede Neural Convolutacional Unidimensional	
AARE	Erro Relativo Absoluto Médio	
ANN	Artificial Neural Network	
AR	Autorregressivo	13
ARIMA	Modelo Integrado Autorregressivo e de Média Móvel	
CV	Coefficiente de Variação	32
CA	Coefficiente de Assimetria	32
DAN2	Arquitetura dinâmica para ANN	
EMBASA	Empresa Baiana de Águas e Saneamento S.A.	1
EV	Variância Estimada	22
IA	Inteligência Artificial	1
INMET	Instituto Nacional de Meteorologia do Brasil	30
IUWM	Modelo de Gestão Integrada de Recursos Hídricos	
LLSSIM	MLP baseada em mínimos quadrados lineares e no algoritmo Simplex	21
LR	Regressão Linear	
LSSVM	SVM baseada em Mínimos Quadrados	
LSTM	Rede Neural com Memória de Curto e Longo Prazo	
MA	Média móvel	13
MAE	Erro Médio Absoluto	
MAPE	Erro Médio Percentual Absoluto	
MASP	Região Metropolitana de São Paulo	21
Max ARE	Erro Relativo Absoluto Máximo	
ME	Erro Médio	
MLP	Rede Neural Perceptron Multicamadas	
MLP-BP	MLP com algoritmo de retropropagação	25
MLR	Regressão Linear Multipla	
MRE	Erro Relativo Médio	
MSE	Erro Médio Quadrático	

NSCE	Coeficiente de Eficiência <i>Nash-Sutcliffe</i>	25
OLS	Regressão Linear dos Mínimos Quadrados Ordinários	
PIPPE	Plataforma de Informações das Plantas de Processos da EMBASA . .	4
PSO	Otimização por Enxame de Partícula	
PU	Unidade de Processamento	
<i>R</i>	Coeficiente de Correlação	18
<i>R</i> ²	Coeficiente de Determinação	18
RMS	Região Metropolitana de Salvador	1
RMSE	Raiz Quadrada do Erro Médio	
RSVR	Reservatório	1
SCADA	Sistema de Supervisão e Aquisição de Dados	
SVM	Máquina de Vetores de Suporte	
SVR	Máquina de Vetores de Suporte para Regressão	
SVR-ANN	Máquina de Vetores de Suporte para Regressão com Redes Neurais Artificiais	2

INTRODUÇÃO

O aquecimento global e o crescimento da população mundial (DINIZ, 2019) tem como consequência o crescimento da demanda por água potável. A população mundial aumenta em cerca de 80 milhões de pessoas ao ano, ou em mais de 200 mil pessoas por dia, o que confirma a necessidade crescente de água disponível para consumo humano (RIPPLE et al., 2019). Nesse aspecto, o gerenciamento adequado dos serviços públicos de abastecimento de água é de fundamental importância para garantir a disponibilidade contínua de água potável para a população. A decisão sobre o gerenciamento de serviços públicos de abastecimento de água, depende de níveis futuros de demanda, onde o horizonte e a periodicidade de previsão são os principais motores para a escolha da técnica e do modelo para previsão, segundo Donkor et al. (2014). Por isso, a necessidade de, cada vez mais, aprimorar a precisão na previsão da demanda de água, sendo que a utilização de modelos de Aprendizagem de Máquina, como Redes Neurais Artificiais (ANNs, do inglês *Artificial Neural Networks*) e suas variações, podem contribuir de forma decisiva neste processo com a predição de cenários futuros de demanda.

Ao longo do tempo, alguns autores desenvolveram estudos utilizando técnicas de Inteligência Artificial (IA) e Aprendizagem de Máquina na construção de métodos e modelos que possam determinar com precisão a previsão da demanda de água, seja a curto, médio ou longo prazo, como em Bakker et al. (2014), Santos e Filho (2014), Sharvelle et al. (2017) e Rees, Clark e Nawaz (2020). Além disso, a previsão da demanda de água pode ser utilizada para melhorar a eficiência da rede de abastecimento, no sentido de reduzir custos operacionais e eliminar problemas de falta d'água e vazamentos ao longo da rede.

A Região Metropolitana de Salvador (RMS) é composta por 13 municípios (IPEA, 2015), dos quais 06 são atendidos exclusivamente pelo Sistema Integrado de Abastecimento da RMS (Candeias, Lauro de Freitas, Madre de Deus, Salvador, São Francisco do Conde, Simões Filho). O sistema integrado de abastecimento é operado pela Empresa Baiana de Águas e Saneamento S.A. (EMBASA), sendo composto por 16 reservatórios, 05 captações e 04 estações de tratamento de água. Esse complexo sistema responde pelo fornecimento de água potável para uma população de mais de 3 milhões de habitantes, que corresponde a 85% da população da RMS, de acordo com estimativas do Instituto Brasileiro de Geografia e Estatística (IBGE, 2021).

Embora o sistema integrado de abastecimento que atende a RMS seja composto por 16 reservatórios, apenas 11 recebem água diretamente das estações de tratamento, listadas na Tabela 1.1 (exceto o R23A), onde **RSVR** significa Reservatório, **Estação** é a Estação de Tratamento de Água, **Capacidade** é a capacidade de armazenamento e **Cobertura** refere-se à área atendida com água do respectivo Reservatório (RSVR). Os outros 05 reservatórios são abastecidos indiretamente pelos primeiros. Este estudo também teve como um dos objetivos prever a vazão de água para os 10 reservatórios que recebem água

Tabela 1.1 Lista dos dez principais reservatórios da RMS.

RSVR	Estação(s)	Capacidade	Cobertura(s)
R1	ETA Vieira de Melo + ETA Teodoro Sampaio	11.500 m ³	Zonas
R7	ETA Principal + ETA Vieira de Melo + ETA Teodoro Sampaio	36.000 m ³	Zonas e R3, R19, R15 e R5
R25	ETA Principal	8.700 m ³	Zonas
R14	ETA Principal	17.400 m ³	Zonas
R20	ETA Principal	8.700 m ³	Zonas
R17	ETA Principal	7.616 m ³	Zonas
R10	ETA Principal	8.700 m ³	Zonas
R18	ETA Principal	8.700 m ³	Zonas e R12
R21	ETA Principal	8.700 m ³	Zonas
R23B	ETA Principal	17.400 m ³	Zonas e R23A

Fonte: Próprio autor.

diretamente das estações de tratamento, pois é destes reservatórios que a água potável é distribuída para zonas, bairros e reservatórios intermediários, garantindo o abastecimento contínuo de água para toda a população. O décimo primeiro reservatório, R23A, não foi estudado porque não forneceu dados suficientes para previsão.

O principal objetivo desta pesquisa é verificar a viabilidade de um novo modelo híbrido utilizando Máquina de Vetores de Suporte para Regressão com Redes Neurais Artificiais (SVR-ANN) para previsão da demanda de água dos reservatórios da RMS, em comparação a outros modelos de aprendizagem de máquina tradicionais, com intuito de efetuar predições mais precisas, avaliadas através de medidas de erro, para curto e médio prazo, que contribuam para melhoria na eficiência do sistema de abastecimento.

Após a implementação dos modelos de Aprendizagem de Máquina, os dados de consumo e meteorológicos disponíveis foram utilizados para treinamento, teste e predição. Os dados ausentes ou inválidos foram substituídos utilizando a técnica de imputação por média móvel. Essa junção dos dados de consumo e meteorológicos são importantes porque melhoram a precisão da previsão (BAKKER et al., 2014), principalmente se o modelo for utilizado como uma ferramenta de controle do sistema de abastecimento de água e também para detecção de anomalias.

1.1 PROBLEMA

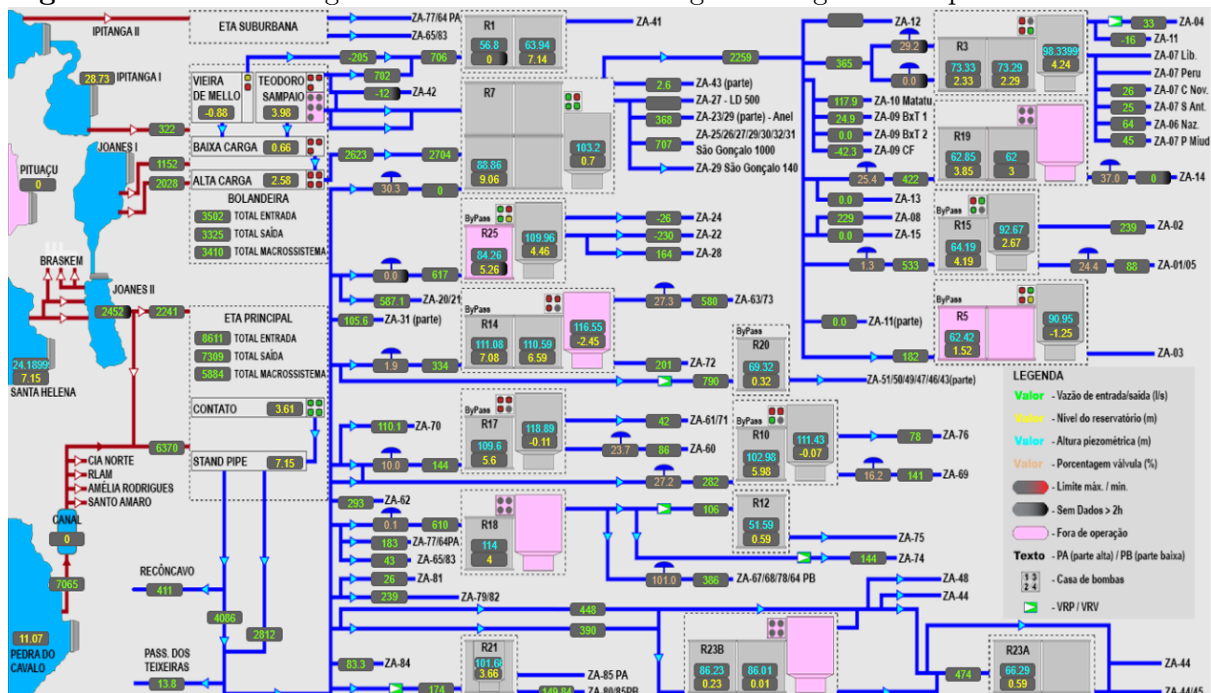
O transporte de água é um dos principais subprocessos de um sistema integrado de abastecimento, pois é através de adutoras e redes de tubulações que a água é transportada, desde o ponto de captação nas barragens até as unidades de tratamento de água e depois na distribuição para os bairros e localidades abastecidas.

Segundo Farias et al. (2018), o transporte de água representa custos de energia e depende da disponibilidade de água, estando sujeito (i) ao crescimento populacional - se

a população cresce, mais água precisa ser transportada; e (ii) à escassez de água - quanto menor a disponibilidade de água maior o custo de transporte para trazê-la de locais mais distantes. E, conforme Rees, Clark e Nawaz (2020), investimentos em infraestrutura de abastecimento de água são caros e levam tempo para implementação e para retorno do investimento, por serem obras de grande magnitude, que demoram para serem planejadas e implementadas, podendo ainda sofrer mudanças na data de entrega. Sendo assim, investimento nesse tipo de infraestrutura requer previsões assertivas a longo prazo.

A RMS possui um sistema complexo de abastecimento de água, composto por 16 reservatórios interconectados, 05 pontos de captação em barragens e 04 estações de tratamento de água, que juntos garantem o abastecimento contínuo de água potável para toda a população, conforme mostra a Figura 1.1 com o panorama geral do Sistema Integrado de Abastecimento. Contudo, a previsão assertiva da demanda de água para as regiões abastecidas é um dos pontos críticos desse sistema, pois os níveis dos reservatórios são controlados levando-se em consideração essa demanda variável que tende sempre a baixar o nível dos mesmos à medida que ocorre o consumo. Assim, as condições climáticas e outros fatores, como descrito por Yassuda et al. (1976 apud SILVA et al., 2008), tendem a influenciar essa demanda, levando o sistema ao desequilíbrio à medida em que acelera a vazão de saída dos reservatórios, tendendo a esvaziá-los mais rapidamente.

Figura 1.1 Sistema Integrado de Abastecimento de Água da Região Metropolitana de Salvador.



Fonte: Imagem georreferenciada do sistema PIPPE (2022).

A previsão adequada da demanda de água para os reservatórios de água potável, em horas, dias e semanas, é de fundamental importância para tempestividade na execução das manobras de bombeamento que garantam o abastecimento contínuo de água para

os bairros da cidade de Salvador e RMS, mantendo sempre os reservatórios com níveis adequados e evitando o esvaziamento por completo dos mesmos. O sistema atualmente utilizado para esse monitoramento são os dados da Plataforma de Informações das Plantas de Processos da EMBASA (PIPPE), no entanto, o *dashboard* de previsão da demanda de água está desatualizado em relação à estrutura dos reservatórios e precisa de melhorias para que volte a ser utilizado com eficiência.

1.2 MOTIVAÇÃO

A água é essencial para os seres vivos, estando presente nas células de quase todas as formas de vida existentes na Terra. Os seres humanos, por exemplo, possuem uma forte dependência de água para sua sobrevivência sob diversos aspectos. Segundo Farias et al. (2018, p. 01, tradução nossa) “a água é um dos recursos naturais mais importantes para sustentar a vida e garantir a qualidade de vida das pessoas”. Além disso, a água tem múltiplas utilidades para os seres humanos, mais de 70% do corpo humano é composto por água, de acordo com Ribeiro (2008).

Segundo Rees, Clark e Nawaz (2020), as projeções de crescimento populacional podem ser convertidas em previsões domésticas adequadas para predição de consumo de água. Entretanto, outras questões também influenciam no consumo de água, como temperatura, umidade, precipitação, preço do serviço e nível socioeconômico da população (YASSUDA et al., 1976 apud SILVA et al., 2008). Dessa forma, a previsão adequada da demanda de água é fundamental para garantir a continuidade no abastecimento de água, reduzir custos operacionais e de energia, otimizar o sistema de abastecimento e evitar interrupções do serviço com consequente falta d’água.

As redes de distribuição que interligam os reservatórios aos bairros e zonas de abastecimento possuem uma determinada capacidade de pressão e vazão, portanto a previsão adequada da demanda de água pode contribuir para a utilização adequada da rede, reduzindo o risco de estouros nas tubulações ou mesmo falta d’água em regiões subdimensionadas.

De acordo com Ludermir (2021) estamos vivendo uma nova revolução industrial, com o desenvolvimento de tecnologias avançadas de computação, como a IA, aliada à crescente disponibilidade de dados e o crescimento do poder computacional disponível. O que permite a expansão do uso computacional para tarefas mais complexas como a execução dos modelos de predição.

Neste cenário, torna-se cada vez mais viável a aplicação de técnicas de IA para resolver problemas complexos e não lineares como, por exemplo, a previsão da demanda de água, através da implementação de algoritmos de Aprendizagem de Máquina com o processamento de grandes volumes de dados. Além disso, a aplicação de técnicas de Aprendizagem de Máquina sobre um problema real e cotidiano traz consigo uma motivação à parte, no sentido de contribuir para melhoria da sociedade, da saúde das pessoas e da qualidade de vida da população atendida, a qual poderá ser diretamente beneficiada com os resultados deste trabalho.

1.3 OBJETIVOS

O objetivo geral deste trabalho é verificar a viabilidade de utilização de um novo modelo híbrido de predição SVR-ANN na previsão da demanda de água para a RMS, em comparação a outros modelos de aprendizagem de máquina mais tradicionais, como os baseados em ANNs, Máquina de Vetores de Suporte para Regressão (SVR, do inglês *Support Vector Machine*) e Modelo Autorregressivo Integrado de Média Móvel (ARIMA, do inglês *Autoregressive Integrated Moving Average*), para previsão de curto e médio prazo (hora, dias e semanas), considerando os dados históricos de consumo e meteorológicos disponíveis.

Além da análise sobre a viabilidade de utilização de um novo modelo híbrido SVR-ANN na previsão da demanda de água, também se constituem objetivos específicos deste trabalho:

1. Analisar os dados históricos de vazão dos reservatórios que abastecem a RMS para entender o comportamento do consumo e suas variações;
2. Implementar modelos de Aprendizagem de Máquina, aplicáveis a séries temporais, para previsão de demanda de água, principalmente ANNs, ANNs híbridas e suas variações, utilizando os dados de consumo e meteorológicos disponíveis;
3. Efetuar a análise comparativa sobre o desempenho dos modelos implementados, estabelecendo critérios objetivos de medição e classificação dos resultados, de maneira que possibilite a identificação do modelo melhor ajustado para predição da demanda de água para cada reservatório;
4. Implementar e atualizar *dashboard* de previsão da demanda de água, através da exibição de gráficos e informações, que permitam aos operadores monitorar as previsões e otimizar o processo de tomada de decisão.

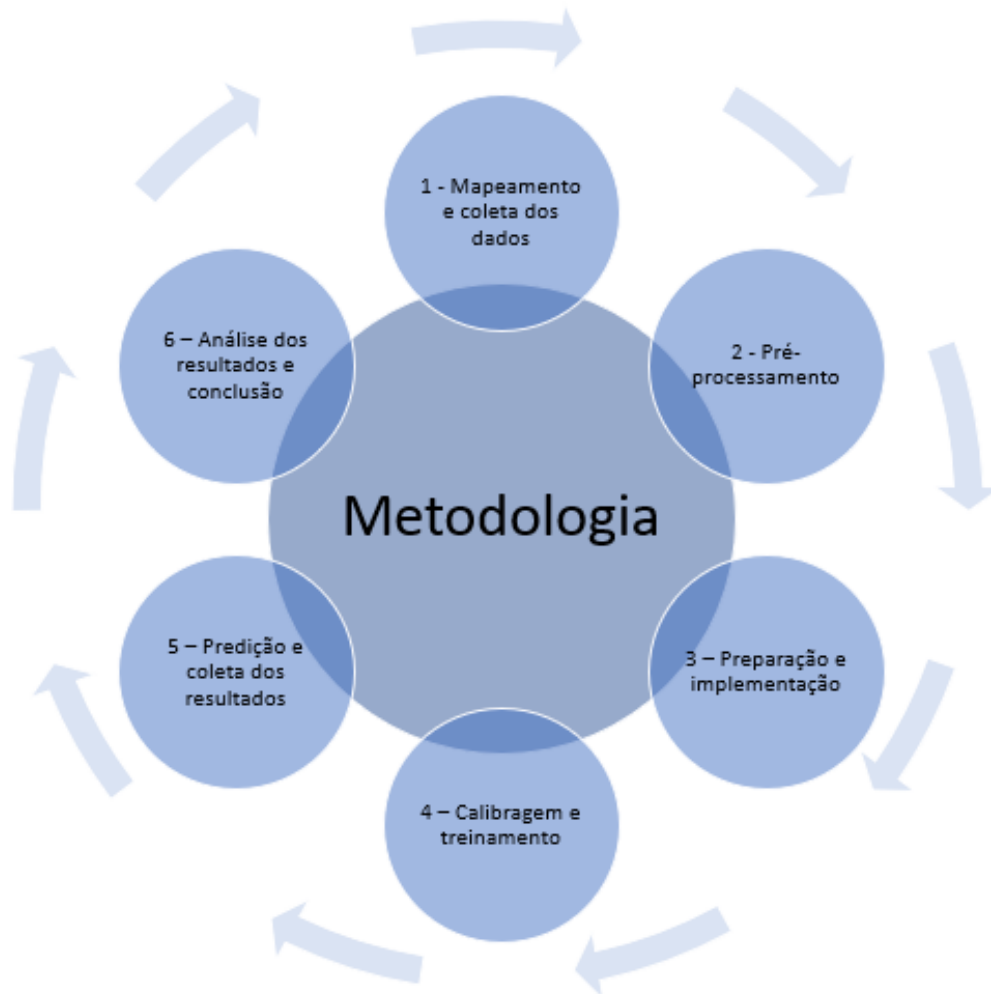
1.4 METODOLOGIA

A Figura 1.2 demonstra o fluxo das ações previstas na metodologia e aplicadas para o desenvolvimento das atividades de pesquisa, implementadas para alcance dos objetivos propostos nesse trabalho.

A metodologia de pesquisa adotada para o trabalho foi dividida em etapas que corresponderam ao mapeamento e coleta dos dados, pré-processamento, preparação e implementação, calibragem e treinamento, predição e coleta dos resultados, e, por fim, análise dos resultados e conclusão.

Durante a fase de coleta, foram levantados e carregados dados históricos meteorológicos e de consumo dos reservatórios de água potável que abastecem diariamente os bairros e localidades da região metropolitana, para o período de janeiro/2017 a fevereiro/2022. A partir destes dados foi possível verificar o comportamento do consumo de água ao longo do tempo para as regiões abastecidas, bem como a sazonalidade e distribuição da demanda. Durante essa fase de pré-processamento foram utilizadas ferramentas como o SAP®Data Services Designer (SAP S.A., Walldorf, Alemanha) para cópia e consolidação dos dados

Figura 1.2 Fluxo proposto na metodologia de trabalho



Fonte: Próprio autor.

extraídos do ambiente de produção da PIPPE e à linguagem R para o pré-processamento com tratamento dos dados ausentes e inválidos.

Na fase de preparação e implementação, os modelos de Aprendizagem de Máquina foram codificados utilizando à linguagem Python e R com técnicas de otimização para busca da melhor eficiência em cada modelo. Em seguida, na fase de predição e coleta de resultados, foram realizadas as previsões, com coleta dos resultados e cálculo das métricas de erro. Por fim, foi realizada a análise dos resultados com escolha do modelo melhor ajustado para cada reservatório e implementação do *dashboard* de previsão.

1.5 RESULTADOS ALCANÇADOS

Ao final da fase de implementação e execução dos modelos propostos, o quadro de resultados das predições viabilizou a análise e identificação do modelo de predição mais eficiente

para cada reservatório. A partir dessa escolha, os modelos escolhidos foram utilizados para implementação do *dashboard* de monitoramento das previsões de demanda de água para visualização pelas equipes de operação da EMBASA através da PIPPE.

O *dashboard* criado, contendo as informações de predição atualizadas, poderá ser utilizado pelos técnicos da operação, no sentido de otimizar o processo de tomada de decisão em relação às manobras para equilíbrio do nível dos reservatórios, com objetivo de que se evite o esvaziamento dos mesmos e a consequente interrupção do abastecimento de água em determinadas regiões. Isso também evitará que esforços desnecessários sejam efetuados para bombeamento de água para regiões que não necessitariam, em função da previsão de redução do consumo, e direcionamento dessa energia para acionamento de bombas visando elevação do nível de reservatórios para outros locais que possuem previsão de aumento na demanda de água a curto prazo.

Além da análise comparativa dos resultados, com identificação do modelo que apresente a maior precisão com menor taxa de erros na previsão da demanda de água e da elaboração de um *dashboard* com as predições, também é uma contribuição deste trabalho, bem como a enumeração das possíveis dificuldades encontradas durante a fase de implementação e treinamento, e o registro dos pontos críticos e possíveis melhorias para implementação de trabalhos futuros.

1.6 ESTRUTURA DO TEXTO

No Capítulo 2 é apresentada a fundamentação teórica com descrição de conceitos básicos inerentes ao tema, bem como alguns modelos de Aprendizagem de Máquina comumente utilizados; o Capítulo 3 apresenta uma revisão dos trabalhos relacionados, como os que tratam das Máquinas de Vetores de Suporte (SVM, do inglês *Support Vector Machines*) baseadas em regressão, ANNs e suas variações, para previsão de demanda de água de curto, médio e longo prazo. Já no Capítulo 4 é apresentada a proposta de pesquisa, contemplando o novo modelo proposto, os demais modelos implementados como base comparativa e a metodologia utilizada. Os resultados alcançados são apresentados no Capítulo 5. Por fim, as conclusões são descritas no Capítulo 6.

FUNDAMENTAÇÃO TEÓRICA

2.1 INTRODUÇÃO E CONCEITOS BÁSICOS

Segundo Sharvelle et al. (2017), as ANNs, Redes Neurais Difusas (FNNs, do inglês *Fuzzy Neural Networks*) e Sistemas de Inferência Fuzzy (FISs, do inglês *Fuzzy Inference Systems*) têm sido utilizados para melhorar as previsões de curto prazo, pois na previsão da demanda de água as interações são fortemente não lineares entre fatores naturais e sociais que influenciam o consumo e, por isso, elas inibem a ampla aplicabilidade das metodologias tradicionais de análise de séries temporais ou baseadas em regressão. Para Wu (1992), a estrutura e funcionamento das ANNs são baseadas no conhecimento que temos sobre o funcionamento do cérebro humano e problemas não estruturados compõem um terreno fértil para sua aplicação.

Segundo Ludermir (2021, p. 89) as ANNs “são modelos matemáticos que se inspiram nas estruturas neurais biológicas e que têm a capacidade computacional adquirida por meio de aprendizado” e o aprendizado de máquina pode acontecer de três formas principais:

- Aprendizado Supervisionado, onde o algoritmo recebe, para treinamento, um conjunto de dados de entrada e os respectivos resultados desejados;
- Aprendizado Não Supervisionado, onde o conjunto de dados de entrada é apresentado sem rótulos e o algoritmo analisa os dados e tenta determinar se os mesmos podem ser colocados em agrupamentos ou *clusters* pela similaridade de seus atributos;
- Aprendizado por Reforço, onde além do conjunto de dados de entrada, o algoritmo recebe um sinal de reforço ao invés da resposta, sendo bastante utilizado, por exemplo, em jogos e robótica.

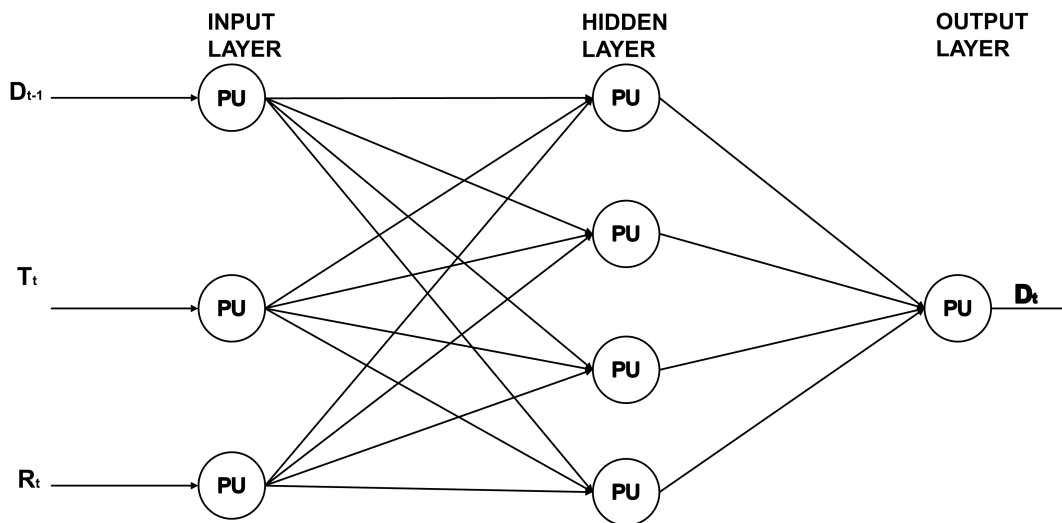
A principal vantagem na utilização de modelos de ANNs é a sua capacidade de autoaprendizagem (BOUGADIS; ADAMOWSKI; DIDUCH, 2005), onde a partir das entradas e saídas desejadas, o modelo tentará reproduzir as saídas por meio de uma série de iterações até atingir uma taxa de erro aceitável. As estruturas básicas das ANNs são as Unidades de Processamento (PUs, do inglês *Process Units*).

As estruturas básicas das ANNs são as PUs; essas estruturas simulam o funcionamento de um neurônio no cérebro humano, onde cada PU recebe um valor como entrada e produz uma saída baseada na soma de seus pesos de entrada e um determinado valor limite. A Figura 2.1 ilustra uma ANN, utilizada por Bougadis, Adamowski e Diduch (2005), onde as camadas internas são formadas por um conjunto de PUs altamente interligadas, para

as seguintes variáveis de entrada e saída: demanda de água anterior (D_{t-1}), temperatura (T_t), precipitação (R_t) e previsão da demanda de água (D_t). Nesse conceito, segundo Wu (1992), as ANNs seriam implementadas e operadas em dois estágios:

1. Treinamento, onde as PUs são calibradas a partir dos dados de entrada, e suas saídas serviriam como entrada para as próximas PUs;
2. Operação, onde as ANNs seriam utilizadas para solução de problemas reais.

Figura 2.1 Arquitetura de uma ANN



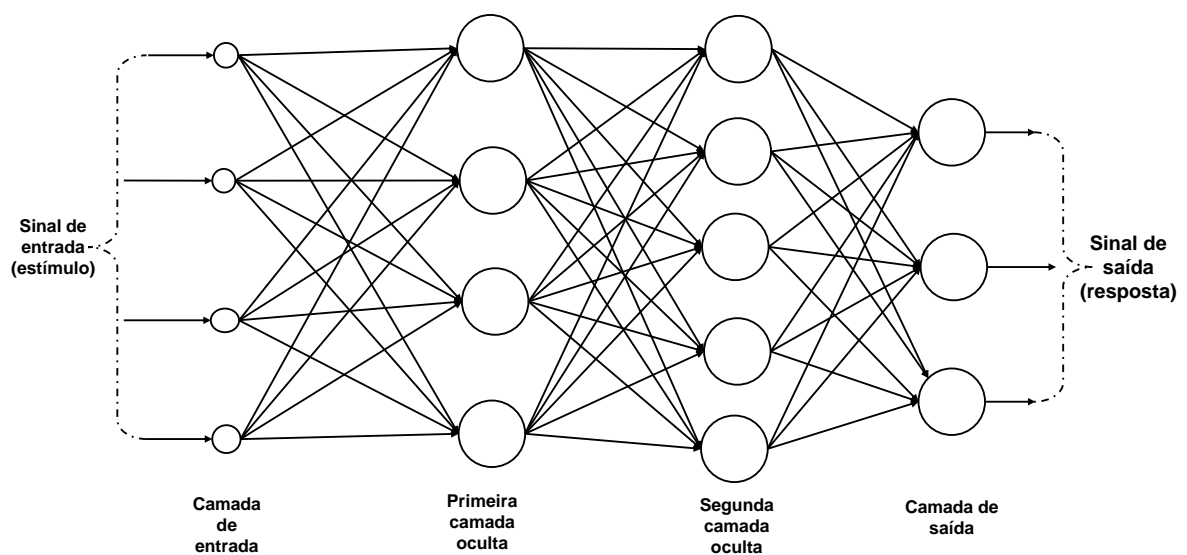
Fonte: Bougadis, Adamowski e Diduch (2005).

Os modelos de ANNs tradicionais geralmente são eficazes no reconhecimento de padrões, agrupamentos e previsão. Contudo, sua implementação e calibração não é simples, pois depende muito da situação-problema, onde o número de camadas, de PUs ocultas e as variáveis de entrada, precisam ser experimentalmente determinados (GHIASSI; SAIDANE, 2005). Embora existam bibliotecas disponíveis para simplificar sua implementação, as ANNs demandam um tempo para parametrização e treinamento até a operação. Além disso, a depender do volume de dados e da complexidade do problema podem resultar em alto custo computacional. Para Wu (1992) as ANNs não supervisionadas dependem de um maior grau de inspiração neural, em relação as supervisionadas, para que menos dados sejam necessários durante o treinamento. Elas também têm a capacidade de se adaptar às mudanças ambientais após o treinamento, mas não possuem regras formais relacionadas à escolha da topologia e da matriz de peso inicial.

Existe uma classe de ANN chamada Rede Neural Perceptron Multicamadas (MLP, do inglês *Multilayer perceptron*), que consiste de uma camada de entrada, uma ou mais

camadas computacionais ocultas e uma camada de saída. Cada camada é composta por PUs, onde o sinal de entrada se propaga para frente através da rede, camada por camada. A rede é totalmente conectada, onde os neurônios de cada camada se conectam a todos os neurônios da camada anterior. Segundo HAYKIN (2001), essa classe de ANNs tem sido aplicada com sucesso na solução de diversos problemas complexos com o algoritmo de retropropagação de erro (*error backpropagation*) para treinamento de forma supervisionada.

Figura 2.2 Arquitetura geral de uma MLP.



Fonte: Adaptado de HAYKIN (2001).

cada camada, propagando e ajustando esse conhecimento para as camadas subsequentes. Repetindo este processo até que o grau de desempenho previamente configurado seja alcançado.

O modelo DAN2 de Ghiassi e Saidane (2005) também foi classificada por eles como um modelo de rede neural puramente *feedforward*, justamente pela capacidade de aprender e acumular conhecimento repetidamente em cada camada até o alcance do resultado desejado. A arquitetura DAN2 apresentada é diferente da tradicional, onde a estrutura da rede é simplificada, a quantidade de PUs ocultas é fixo (quatro), as PUs das camadas internas se relacionam de muitos-para-um. Para demonstrar a eficiência da DAN2 Ghiassi e Saidane (2005) utilizaram como série de entrada, por exemplo, a série de manchas solares de 1920-1987, que é um conjunto de dados não linear, não gaussiano e tradicionalmente usado para medir a eficácia de modelos estatísticos não lineares.

Um dos métodos para predição de séries temporais mais conhecidos é o ARIMA, popularizada por Box e Jenkins (1970), que desenvolveram uma metodologia prática e iterativa para construção de modelos ARIMA, proporcionando sua expansão e utilização por um longo tempo. De forma resumida, o ARIMA utiliza uma função linear composta por várias observações passadas e erros aleatórios para efetuar a previsão do valor futuro de uma variável. Com isso, em geral, uma série temporal pode ser modelada como uma combinação de valores passados e erros, que podem ser representados como ARIMA (p,d,q) , pela Equação (2.1):

$$Y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (2.1)$$

onde Y_t e ε_t representam, respectivamente, o valor real e o erro aleatório no período t . Os parâmetros ϕ_p e θ_q representam, respectivamente, a parte autorregressiva e a parte de média móvel do modelo, e p e q se referem ao número de ordem dos polinômios Autorregressivo (AR) e Média móvel (MA), respectivamente. Assim, se p for igual a 0 o modelo se reduz a uma MA de ordem q , da mesma forma, se q for igual a 0, o modelo se resume a um AR de ordem p . A variável d , por sua vez, indica o número de diferenciações necessárias para tornar a série estacionária.

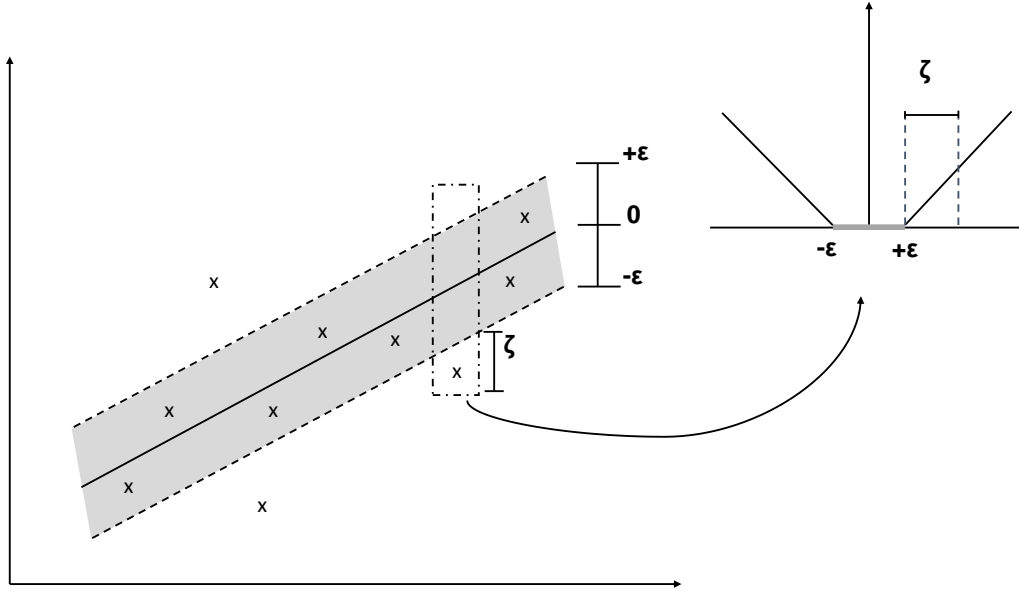
Os modelos ARIMA e suas variações são bastante flexíveis e, por isso, bastante empregados na previsão de séries temporais. Entretanto, para Zhang (2003), sua principal limitação é a presunção de linearidade, ou seja, ele não consegue capturar os padrões não lineares e, por isso, possui limitações quanto à aplicabilidade a problemas mais complexos.

As SVMs são modelos de aprendizagem supervisionada baseados nos conceitos de Cortes e Vapnik (1995), onde um hiperplano é encontrado através de um processo de otimização que garante a maior distância de separação entre os grupos de dados. Existe uma variação de SVM baseada em regressão (DRUCKER et al., 1996), também conhecida como SVR, que vem sendo utilizada para previsão da demanda de água, como verificado em Herrera et al. (2010), Adamowski e Karapataki (2010) e Candelieri (2017).

Enquanto as SVMs procuram dividir os dados em diferentes classes e classificar os dados futuros de forma correta, as SVRs buscam uma função $f(x)$ que aproxime bem os dados da amostra (VERDÉRIO, 2015). De forma bastante simplificada, considerando o conjunto de treinamento como sendo $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\} \subset X \times \mathbb{R}$, onde X denota o espaço do padrão de entrada, a ideia básica por trás da SVR consiste em

encontrar uma função $f(x)$ em que tenha no máximo ε desvios dos alvos reais y_i para todos os dados de treinamento. Ou seja, como mostra a Figura 2.4, a SVR não se preocupa com erros desde que sejam menores que ε e também não aceita qualquer desvio maior do que ε (SMOLA; SCHÖLKOPF, 2004).

Figura 2.4 Configuração de margem de perda para uma SVM baseada em regressão.



Fonte: Smola e Schölkopf (2002 apud SMOLA; SCHÖLKOPF, 2004)

Para exemplificar, considerando que f seja uma função linear, dada, por exemplo, pela Equação (2.2):

$$f(x) = \langle w, x \rangle + b, \text{ com } w \in X, b \in \mathbb{R} \quad (2.2)$$

onde $\langle w, x \rangle$ denota o espaço escalar em X . A planicidade em (2.2) significa que se busca um w menor possível. Uma forma de se garantir isso seria minimizar a função norma, ou seja, $\|w\|^2 = \langle w, w \rangle$, que pode ser escrito como um problema convexo de otimização, conforme Equação (2.4):

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 \\ & \text{sujeito a } \begin{cases} y_i - \langle w, x_i \rangle - b \\ \langle w, x_i \rangle + b - y_i \end{cases} \end{aligned} \quad (2.3)$$

A suposição dada pela Equação (2.4) seria de que tal função f que aproxima todos os pares (x_i, y_i) , com precisão ε , realmente existe, ou seja, que o problema de otimização convexa é viável. Todavia, nem sempre isso ocorre ou podemos permitir a introdução de algum limite de erros. Assim como sugerido por (CORTES; VAPNIK, 1995), pode-se introduzir variáveis de folga ξ_i, ξ_i^* para lidar de outra forma com a inviabilidade do

problema de otimização (2.4). Utilizando a formulação sugerida por (CORTES; VAPNIK, 1995).

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi + \xi_i^*) \\ & \text{sujeito a } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (2.4)$$

A constante $C > 0$ determina a troca entre o nivelamento de f e a quantidade de desvios maiores do que ϵ que são tolerados. Também chamada de função de perda ϵ -insensitiva $|\xi|_\epsilon$ descrita pela Equação (2.5):

$$|\xi|_\epsilon := \begin{cases} 0 & \text{se } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{caso contrário.} \end{cases} \quad (2.5)$$

Uma série de modelos de previsão com SVRs para otimização da programação de bombeamento e detecção de anomalias é apresentada por Candelieri (2017). Usando uma abordagem em dois estágios, com base em agrupamento de séries temporais e regressão SVM, para analisar dados de séries temporais de demanda de água em um nível agregado (demanda de água urbana com dados do Sistema de Supervisão e Aquisição de Dados (SCADA, do inglês *Supervisory Control and Data Acquisition*)) e em um nível individual (cliente único com dados de Leitura Automática do Medidor). Segundo Candelieri (2017) as duas aplicações visam a previsão de demanda urbana (utilizando o agrupamento em *clusters*) e a detecção de anomalias (utilizando a regressão SVM). Enquanto o agrupamento em *clusters* das séries temporais com os dados do SCADA foram usados para otimização das operações, em particular a otimização do cronograma de bombeamento, os dados dos consumidores individuais foram utilizados em modelos de regressão SVM para identificar possíveis anomalias *online* devido a falhas de medidores, fraudes, ataques ciberfísicos e mudanças significativas nos hábitos de consumo.

Em diversos estudos publicados sobre o tema, a previsão de demanda de água frequentemente é dividida em previsão de curto prazo e previsão de longo prazo. Geralmente, a previsão de demanda de curto prazo diz respeito à previsão horária, diária ou até semanal e sua predição está mais associada à otimização dos sistemas de abastecimento, como por exemplo, para redução de custos com eletricidade ou para otimização do abastecimento. Segundo Alvisi, Franchini e Marinelli (2007, p. 01, tradução nossa) “Para a operação eficiente da rede de distribuição de água, alguma forma de previsão de demanda geralmente é necessária para programar os arranjos de bombeamento nas próximas 24 horas, para aproveitar a estrutura de tarifas de energia elétrica”.

A previsão de demanda de água de longo prazo geralmente se refere a períodos mais longos como semanas, meses e até anos. Neste caso, sendo mais associada a questões de planejamento dos sistemas de abastecimento para iniciativas de médio a longo prazo.

Em Donkor et al. (2014) é apresentado um estudo comparativo, de forma organizada e estruturada, de um conjunto de publicações sobre os métodos e modelos utilizados para previsão da demanda de água, entre os anos de 2000 a 2010. Elaborado pensando em profissionais e pesquisadores que buscam orientação e direcionamento, na adoção de

métodos e modelos de previsões, o guia elenca as principais técnicas utilizadas naquele período, de acordo com o conjunto de dados de entrada utilizado e o objetivo de predição desejado.

Alguns autores como Zhang (2003), Babu e Reddy (2014) e Oliveira e Ludermir (2016) desenvolveram estudos sobre a utilização de modelos híbridos para previsão, onde a modelagem é realizada de forma a combinar diferentes métodos que possam considerar a característica linear e não linear das séries temporais. A motivação para utilização de modelos híbridos, segundo Zhang (2003), vem do fato de que nem sempre é possível determinar se uma série temporal em estudo é gerada a partir de um processo subjacente não linear ou linear, ou ainda se um método específico é mais eficaz do que outro para previsão fora da amostra.

Foram encontrados poucos trabalhos de pesquisa que tratam sobre a utilização de modelos híbridos de aprendizagem especificamente voltados para previsão da demanda de água e nenhum deles com referência à aplicação para a RMS. Em Zhang (2003) é apresentada uma abordagem combinada para a previsão de séries temporais, utilizando o modelo linear ARIMA e o modelo não linear ANN para capturar diferentes formas de relacionamento nos dados da série temporal; o modelo híbrido aproveita a força única do ARIMA e da ANN na modelagem linear e não linear. Para problemas complexos que possuem estruturas de correlação linear e não linear, o método de combinação pode ser uma maneira eficaz de melhorar o desempenho da previsão, superando os resultados de cada modelo quando utilizados de forma isolada.

Babu e Reddy (2014) desenvolveram um modelo híbrido ARIMA-ANN para a previsão de dados de séries temporais. Eles levaram em consideração a natureza da volatilidade da série temporal utilizando um filtro de média móvel para, em seguida, aplicar o modelo ARIMA-ANN de forma adequada. Para teste e treinamento foram utilizados os conjuntos de dados de manchas solares, dados de preços de eletricidade e dados do mercado de ações. Os resultados obtidos para esses conjuntos de dados mostram que, tanto para previsões de um passo à frente quanto de vários passos, o modelo híbrido proposto teve maior precisão de previsão em relação aos demais modelos individuais ARIMA e ANN e a alguns modelos híbridos ARIMA-ANN existentes.

No trabalho apresentado por Oliveira e Ludermir (2016) foi proposto um sistema híbrido de decomposição evolutiva onde a série temporal é decomposta em relação aos seus padrões de baixa e alta volatilidade com um filtro de suavização exponencial. Em seguida, o termo de alta volatilidade é decomposto em padrões lineares e não lineares, sendo aplicados os modelos ARs e de MAs, além da SVR, respectivamente. Além disso, a otimização dos parâmetros é realizada através do emprego de um algoritmo de Otimização por Enxame de Partícula (PSO, do inglês *Pratic Swarm Optimization*), a fim de produzir previsões mais precisas. O modelo proposto alcançou resultados promissores aumentando, porém, a complexidade computacional em comparação com outros modelos; foram utilizados 13 conjuntos de dados da literatura e seis métodos de previsão, entre eles o método de Zhang (2003), Babu e Reddy (2014) e ARIMA. Os resultados mostraram que nenhum método foi melhor para todos os conjuntos de dados e, em alguns casos, um modelo linear obteve melhores resultados (menores erros). O método proposto obteve os melhores resultados gerais em termos de acurácia das previsões.

Zhang (2003) e Wang et al. (2013) apresentaram abordagem híbridas baseadas no ARIMA-ANN. Enquanto o primeiro utilizou um modelo aditivo (Equação (2.6)), onde o resultado da previsão é dado pela soma da previsão efetuada no ARIMA mais o resultado da ANN, o segundo utilizou um modelo multiplicativo (Equação (2.7)) para chegar ao resultado da predição, ou seja, a previsão é o resultado do produto entre a previsão dada pelo modelo ARIMA e a previsão extraída da ANN.

$$y_t = L_t + N_t, \quad (2.6)$$

$$y_t = L_t \times N_t. \quad (2.7)$$

Nos dois modelos (aditivo e multiplicativo) L_t representa a porção linear da série temporal, N_t representa a porção não linear. Para que seja efetuada a previsão usando os modelos aditivo ou multiplicativo, o passo inicial é efetuar a previsão da porção linear da série temporal, utilizando para isso o modelo ARIMA e usando como entrada a série y_t . A série temporal resultante é chamada de \hat{L}_t ; em seguida, deve-se comparar o valor real y_t com o valor obtido para se extrair a série de componentes não lineares ε_t .

Para o modelo aditivo, temos:

$$\varepsilon_t = y_t - \hat{L}_t, \quad (2.8)$$

para o modelo multiplicativo, temos:

$$\varepsilon_t = y_t / \hat{L}_t. \quad (2.9)$$

A segunda etapa é efetuar a modelagem da componente não linear, utilizando a série ε_t como entrada do modelo ANN. O resultado da predição efetuada pela ANN é chamada de \hat{N}_t e deve ser utilizada para se obter o resultado combinado da predição.

Para o modelo aditivo, temos:

$$y_t = \hat{L}_t + \hat{N}_t, \quad (2.10)$$

para o modelo multiplicativo, temos:

$$y_t = \hat{L}_t \times \hat{N}_t. \quad (2.11)$$

Neste trabalho de pesquisa, um novo modelo híbrido é proposto, chamado SVR-ANN, para predição da demanda de água, utilizando a mesma lógica da metodologia proposta por Zhang (2003) e Wang et al. (2013), com o objetivo de verificar a viabilidade de utilização desse novo modelo, seja pela abordagem aditiva.

De acordo com Donkor et al. (2014), para medir a precisão de um determinado modelo os dados de amostragem são comparados com os dados da previsão e o modelo que demonstrar o menor valor de medida de erro é considerado o mais preciso. Alguns métodos estatísticos são recorrentes na análise de desempenho dos modelos de predição, como

por exemplo, o Erro Médio (ME, do inglês *Mean Error*), Erro Médio Absoluto (MAE, do inglês *Mean Absolute Error*), Erro Médio Percentual Absoluto (MAPE, do inglês *Mean Absolute Percentage Error*), Erro Médio Quadrático (MSE, do inglês *Mean Squared Error*), Raiz Quadrada do Erro Médio (RMSE, do inglês *Root Mean Squared Error*), Coeficiente de Correlação (R) e Coeficiente de Determinação (R^2). Cada função mede características diferentes, podendo gerar também classificações distintas, dependendo de qual medida de erro é usada. É comum, por exemplo, a comparação do resultado de funções distintas aplicadas ao mesmo modelo, como pode ser visto em Bakker et al. (2014), Santos e Filho (2014), Farias et al. (2018) e Farias et al. (2018).

O ME calcula a média de todos os erros do conjunto de dados, denotado pela Equação (2.12)

$$\text{ME} = \frac{1}{n} \sum_{t=1}^n Y_t - Y'_t, \quad (2.12)$$

onde:

Y - representa a série temporal da demanda de água observada,

Y' - representa a série temporal da demanda de água predita,

Y_t - a demanda de água observada no momento t ,

Y'_t - demanda de água predita no momento t , e

n - comprimento da série temporal, em número de observações.

O MAE utiliza o valor absoluto dos erros no cálculo evitando assim as distorções que podem ocorrer no ME e que são provocadas pelos valores negativos. O MAE possui a Equação (2.13)

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |Y_t - Y'_t|, \quad (2.13)$$

onde:

Y - representa a série temporal da demanda de água observada,

Y' - representa a série temporal da demanda de água predita,

Y_t - a demanda de água observada no momento t ,

Y'_t - demanda de água predita no momento t , e

n - comprimento da série temporal, em número de observações.

O MAPE expressa o percentual médio dos erros, em termos absolutos; e quanto menor esse valor, melhor será a previsão, sendo representado pela Equação (2.14)

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - Y'_t}{Y_t} \right|, \quad (2.14)$$

em que:

Y - representa a série temporal da demanda de água observada,

Y' - representa a série temporal da demanda de água predita,

Y_t - a demanda de água observada no momento t ,

Y'_t - demanda de água predita no momento t , e

n - comprimento da série temporal, em número de observações.

O MSE eleva os erros individuais ao quadrado, para depois somá-los, e então divide o resultado pelo número de amostras. Sendo expressado pela Equação (2.15)

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (Y_t - Y'_t)^2, \quad (2.15)$$

em que:

Y - representa a série temporal da demanda de água observada,

Y' - representa a série temporal da demanda de água predita,

Y_t - a demanda de água observada no momento t ,

Y'_t - demanda de água predita no momento t , e

n - comprimento da série temporal, em número de observações.

O RMSE é a raiz quadrada do MSE e é representado pela Equação (2.16)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - Y'_t)^2}, \quad (2.16)$$

em que:

Y - representa a série temporal da demanda de água observada,

Y' - representa a série temporal da demanda de água predita,

Y_t - a demanda de água observada no momento t ,

Y'_t - demanda de água predita no momento t , e

n - comprimento da série temporal, em número de observações.

O Coeficiente de Correlação R , também conhecido como correlação de *Pearson*, que mede a relação ou semelhança entre duas variáveis é dado pela Equação (2.17)

$$R = \frac{S_{xy}}{S_x S_y}, \quad (2.17)$$

onde:

S_{xy} - representa a covariância entre as variáveis x e y ,

S_x e S_y - representam o desvio padrão das variáveis x e y , respectivamente.

O Coeficiente de Determinação R^2 , que varia de 0 a 1, mede o quanto da fração da variável total é explicada modelo. Ou seja, quanto maior é o R^2 mais ajustado será o modelo para a mostra dos dados, sendo calculado pelo quadro do Coeficiente de Correção R .

Para Candelieri (2017), a ideia básica das medidas de erro é comparar previsões a partir

das observações. E, para ele, a medida mais amplamente utilizada no campo da previsão da demanda de água é o Erro Médio Percentual Absoluto (MAPE). Segundo Candelieri (2017), o MAPE é denotado pela fórmula:

O MAE é calculado com base na média dos erros absolutos, sendo, portanto, menos afetado por pontos extremos (*outliers*). O MSE, por sua vez, é comumente utilizado para medir acurácia de modelos, onde cada erro ($Y_t - Y'_t$) é elevado ao quadrado individualmente; neste caso, a métrica é mais sensível a valores discrepantes e caso tenham muitos erros significativos pode extrapolar o resultado. O RMSE é a raiz quadrada do MSE. Enquanto o R^2 representa o percentual de variância dos dados em relação ao modelo, sendo o resultado entre 0 e 1. Neste caso, quanto mais próximo de 1 melhor, expressando quanto mais explicativo é o modelo em relação aos dados previstos. Por fim, o MAPE é uma métrica cujo resultado expressa o percentual de erros em relação aos valores reais.

A métrica MAPE normaliza os erros relativos ao processo e vem sendo utilizada para comparar a performance de modelos distintos aplicados às mesmas séries temporais. Os estudos que aplicaram o uso da função MAPE para análise de desempenho dos modelos de previsão da demanda de água, podem ser consultados em VanBerlo, Ross e Hsia (2021), Ghiassi, Zimbra e Saidane (2008), Koutiva e Makropoulos (2016), Yan, Kun e Yang, Min-Zhi (2018), Candelieri (2017), Bakker et al. (2014) e Farias et al. (2018). A Tabela A.1, do Apêndice A, demonstra uma classificação dos tipos de métricas de avaliação utilizadas para cada trabalho pesquisado.

TRABALHOS RELACIONADOS

Algumas pesquisas classificam a previsão da demanda de água em curto, médio ou longo prazo. Geralmente, a previsão de demanda de curto prazo diz respeito a previsões com resolução horária, diária ou semanal e sua previsão está mais associada à otimização dos sistemas de abastecimento, enquanto a previsão de longo prazo refere-se a períodos mais longos, como meses e anos. Nesse caso, está mais associado a questões de planejamento do sistema de abastecimento para iniciativas de médio e longo prazos.

O algoritmo de MLP baseada em mínimos quadrados lineares e no algoritmo Simplex (LLSSIM) usa uma abordagem diferente para calibração de ANN, onde os pesos da camada de entrada são treinados usando um algoritmo Simplex multi-entrada e os pesos da camada oculta são identificados usando uma abordagem de estimativa linear condicionada por mínimos quadrados. Para Hsu, Gupta e Sorooshian (1996), a abordagem LLSSIM fornece estimativas de peso, em geral, superiores, e com menos iterações de função, em relação às estratégias convencionais de Retropropagação (BPA, do inglês *Backpropagation Algorithm*) e de Retropropagação Adaptativa (ABPA, do inglês *Adaptive Backpropagation Algorithm*).

Os resultados apresentados por Santos e Filho (2014) demonstram a viabilidade do uso de um modelo MLP baseado em LLSSIM para a previsão de curto prazo da demanda de água na Região Metropolitana de São Paulo (MASP) no Brasil, além da vantagem de reduzir as iterações necessárias para identificar os pesos. No modelo desenvolvido, foram coletados dados de consumo (consumo em m^3/s , por hora, dia e semana) e dados climáticos (temperatura, umidade relativa do ar, precipitação, pressão do ar, direção e velocidade do vento), para definir as variáveis de entrada do modelo. Além disso, o consumo foi classificado por dia da semana, estação do ano e tipo de dia.

Alvisi, Franchini e Marinelli (2007) apresentam a construção de um modelo de previsão de curto prazo, dividido em dois módulos, capaz de fornecer uma previsão precisa das demandas futuras semanais e diárias de água. Eles propuseram um modelo baseado em padrões que considerava ciclos sazonais, semanais e diários. Para analisar o desempenho do modelo proposto, foram utilizados dados de uma das 14 estações de medição da rede de abastecimento de água de Castelfranco Emilia (província de Modena, Itália). Deste conjunto de dados foram retirados 09 postos, o que corresponde a uma área de abastecimento para cerca de 23.000 habitantes. O resultado final do trabalho foi um modelo de rede neural baseado em padrões, que foi estruturado em dois módulos:

- O primeiro módulo faz a predição da média diária de demanda de água no dia, ou dias, utilizando para isso um componente persistente e outro periódico. O componente persistente é a memória de curto prazo, e o periódico é o padrão sazonal e semanal de consumo;

- O segundo módulo, para previsão de demanda horária, também possui um componente persistente e outro periódico. No entanto, o componente periódico é obtido usando o resultado do primeiro módulo e o componente persistente é encontrado através de regressão sobre os erros encontrados nos dados de previsão horária.

Para medir o desempenho do modelo Alvisi, Franchini e Marinelli (2007) utilizaram RMSE, MAE e Variância Estimada (EV).

Bougadis, Adamowski e Diduch (2005) investigaram técnicas de regressão, análise de séries temporais e ANN para previsão de curto prazo da demanda de água. Os três modelos estudados foram usados para fazer previsões semanais de pico de demanda de água. Em seus estudos, a técnica ANN superou substancialmente os métodos de Regressão Linear (LR, do inglês *Linear Regression*) e Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression* em termos de precisão em séries temporais de previsão de demanda de água. Além disso, identificam que os melhores resultados foram obtidos combinando a demanda semanal anterior, com os dados de precipitação e temperatura da semana atual. Os testes foram realizados usando um conjunto de dados de pico de demanda de água, temperatura e dados pluviométricos de 1993 a 2002 em Ottawa, Ontário, Canadá. Esses dados foram divididos em 162 registros para teste e 18 para verificação do modelo. Para avaliar o desempenho dos modelos, foram utilizados R^2 , Erro Relativo Absoluto Médio (AARE, do inglês *Average Absolute Relative Error*) e o Erro Relativo Absoluto Máximo (Max ARE, do inglês *Maximum Absolute Relative Error*).

Em Rees, Clark e Nawaz (2020), é proposto um modelo de previsão de longo prazo para a demanda de água, que usa a previsão de longo prazo do crescimento doméstico como um dos insumos, mais especificamente para o período entre 2011 e 2101, 90 anos, em Londres e no Vale do Tâmsa, Inglaterra. No modelo de previsão de demanda doméstica de água apresentado por eles, os autores identificaram que os principais motivadores, que influenciaram a demanda, foram o tamanho da família, o tipo de propriedade e a etnia da família.

Segundo Bakker et al. (2014), os modelos de previsão de demanda, que utilizam a demanda de água como único insumo, são capazes de gerar previsões muito precisas. No entanto, em condições climáticas variáveis, os erros de previsão são grandes. Usando dados de seis diferentes zonas de abastecimento de água na Holanda, os autores realizaram simulações com os modelos de previsão e mostraram que, ao usar variáveis meteorológicas de entrada, os maiores erros de previsão foram reduzidos em 11% e os erros médios em 7%. Essa melhoria na precisão da previsão é importante quando o modelo é usado como ferramenta de controle do sistema de abastecimento de água e também para detecção de anomalias.

Ghiassi, Zimbra e Saidane (2008) demonstraram o uso de ANN dinâmica para prever a demanda de água de curto, médio e longo prazos (horário, mensal e anual), com a divisão dos modelos em duas seções sazonais, uma para alta demanda que estaria em verão e outro de baixa demanda no inverno. Seu estudo foi baseado em volumes diários de bombeamento e volumes horários de demanda de água do sistema de abastecimento da cidade de San Jose, Califórnia, no período de 1995 a 2004.

Para os autores, a modelagem de séries temporais, utilizando ANN dinâmica, foi

suficiente para obter previsões bastante precisas, com acurácia de até 99%, e embora o uso de variáveis exógenas possa melhorar a previsão de demanda de água para alguns modelos, obter e usar essas informações pode representar um desafio adicional. As previsões dessas informações exógenas (clima, dados econômicos e crescimento populacional, por exemplo) são frequentemente imprecisas e podem introduzir erros adicionais ao modelo. Para Ghiassi, Zimbra e Saidane (2008), as melhorias na precisão, obtidas com a inclusão de variáveis exógenas, como clima, dados econômicos e crescimento populacional, podem não compensar o aumento da complexidade.

Para Bakker et al. (2014), embora a relação entre demanda de água e condições meteorológicas pareça óbvia, alguns artigos relatam que o desempenho do modelo não melhora ao usar dados meteorológicos. E em outros, não é informada a diferença entre a precisão gerada com e sem o uso de dados meteorológicos, o que inviabiliza a análise de uma possível melhoria de desempenho. Para ele, a decisão de usar ou não dados meteorológicos, para determinado modelo, deve ser tomada com base no desempenho, nos custos e nos riscos associados. Ao utilizar dados meteorológicos, o desempenho dos modelos de previsão implementados por Bakker et al. (2014) foi melhorado em 7% em relação aos erros médios e 11% em relação aos maiores erros. Essa melhoria pode ser relevante quando é necessária extrema precisão, como para detecção de anomalias, por exemplo. Ou seja, para cada implementação de um modelo de previsão, deve-se avaliar se os maiores custos e a complexidade do uso de dados meteorológicos valem a pena para obter a maior precisão de previsão.

Outras alternativas às ANNs, usadas para prever a demanda de recursos hídricos, também foram exploradas. Dentre eles, destacam-se os modelos baseados em agentes, como os implementados por Galán, López-Paredes e Olmo (2009), Koutiva e Makropoulos (2016), variações do SVM como visto em Yan, Kun e Yang, Min-Zhi (2018), Candelieri (2017), o método *Prophet* proposto por pesquisadores do Facebook e usado por VanBerlo, Ross e Hsia (2021), que consiste em uma combinação de funções configuráveis ajustadas a dados de séries temporais.

Galán, López-Paredes e Olmo (2009) implementaram um modelo baseado em agentes para simular o comportamento das famílias em relação à demanda pelo consumo de água, na cidade de Valladolid na Espanha, considerando uma série de estímulos ao longo do tempo. Para implementação do agente foram estabelecidos três cenários de estudos e utilizados dados conhecidos para modelagem, considerando:

- Modelo de Dinâmica Urbana: baseada em influência da dinâmica urbana e territorial sobre o consumo;
- Modelo de Difusão de Opinião: baseada no mecanismo de endosso e no reforço positivo em redes sociais e na escolha de comportamentos - ambientalista e não ambientalista;
- Modelo de Difusão Tecnológica: baseada na probabilidade adicional de adoção de inovações que proporcionam economia de água pela população;
- Modelo Estatístico de Consumo: utilizando dados espaciais da base de dados de

consumo de água (empresa de abastecimento de água local) e uma base de dados socioeconômica (da Câmara Municipal local).

O trabalho desenvolvido por Galán, López-Paredes e Olmo (2009) foi focado no entendimento sobre as previsões de longo prazo, onde os autores conseguem demonstrar a importância de vários fatores, como por exemplo, as campanhas de educação, processos de difusão tecnológica e dinâmica social no consumo doméstico de água.

Utilizando à linguagem NetLog, Koutiva e Makropoulos (2016) desenvolveram um agente que simula o comportamento de usuários domésticos na busca de água. O comportamento da demanda de água de famílias urbanas foi simulado usando teorias do domínio da psicologia social, para calcular a evolução da demanda doméstica de água e suas propriedades com o sistema de água urbana. No estudo proposto por eles, o resultado da simulação foi comparado com dados da seca que assolou a cidade de Atenas, na Grécia, entre 1988 e 1994, para captar os efeitos de diferentes políticas sobre a demanda doméstica por água.

Em Yan, Kun e Yang, Min-Zhi (2018), um modelo de previsão de demanda de água é proposto, usando uma SVM baseada em Mínimos Quadrados (LSSVM, do inglês *Least Squares Support Vector Machine*) e um algoritmo de PSO. Dados históricos de 2000 a 2010, da província de Jiangsu, Binhai, China, foram usados como referência para estabelecer previsões de demanda de água de alta precisão. Segundo os autores, o modelo PSO-LSSVM tem boa aplicabilidade e alto valor de aplicação na previsão de demanda de água.

O *Prophet* é um modelo de previsão de código aberto disponível para *download* na internet. Pode ser utilizado para previsão de séries temporais com base em modelo aditivo, permitindo a parametrização da sazonalidade anual, semanal e diária, além dos feriados. Sua implementação pode ser feita utilizando à linguagem R ou Python.

VanBerlo, Ross e Hsia (2021) implementaram o modelo *Prophet* para previsão de demanda de água de longo prazo, com dados de consumo (metro cúbico faturado) para a cidade de London, Canadá, entre 7 de julho de 2009 e 2 de setembro de 2020. Neste caso, faltaram dados de 1º de março de 2014 a 30 de setembro de 2014 e de 25 de março de 2017 a 31 de maio de 2017; e mesmo assim, o modelo *Prophet* foi capaz de fazer previsões, superando os concorrentes. Com o diferencial de produzir resultados intrinsecamente interpretáveis, já que o próprio modelo é a soma dos componentes, para efeitos de tendência, sazonalidade e feriados.

Para fins comparativos, foram implementados outros três modelos: a Regressão Linear dos Mínimos Quadrados Ordinários (OLS, do inglês *Ordinary Least Squares Regression*), Rede Neural com Memória de Curto e Longo Prazo (LSTM, do inglês *Long Short-Term Memory*) e Rede Neural Convolutiva Unidimensional (1D-CNN, do inglês *One-dimensional Convolutional Neural Network*). No entanto, o modelo *Prophet* apresentou as menores taxas de erro, com um MAPE de 2,51%.

Existem pesquisas recentes que utilizam o Modelo de Gestão Integrada de Recursos Hídricos (IUWM, do inglês *Integrated Urban Water Management*) como método para previsão e estimativa de demanda de água. De acordo com Sharvelle et al. (2017), o modelo IUWM é uma abordagem utilizada para construção de um balanço hídrico de massa,

que utiliza como parâmetros de entrada os dados das famílias, população, cobertura de solo e dados climáticos para projetar a demanda de água e a economia de acordo com os cenários projetados.

Segundo Huang, Lou e Li (2016), o método IUWM pode ser usado para modelar e estimar o número de dias que a água doce durará em situações de emergências em sistemas urbanos de água. Em Sharvelle et al. (2017), o modelo IUWM foi utilizado para previsão de demanda de água, considerando diferentes cenários, como previsão do clima, crescimento populacional e reutilização da água. Os autores mostram que o modelo pode ser utilizado como ferramenta útil para gestão e planejamento de sistemas de abastecimento de água. Esse modelo utiliza três métricas para avaliação da performance: o Erro Relativo Médio (MRE, do inglês *Mean Relative Error*), Coeficiente de Eficiência *Nash-Sutcliffe* (NSCE) e a fração de viés (BIAS, do inglês *Bias*).

Segundo Bakker et al. (2014), historicamente o modelo ANN é o mais utilizado para prever a demanda de água. Inicialmente, foram utilizados modelos convencionais de ANN, evoluindo posteriormente para modelos mais complexos, incluindo ANNs dinâmicos e modelos híbridos. Algumas pesquisas têm tentado aproveitar o potencial individual de cada método para compor modelos híbridos de aprendizado de máquina, com o objetivo de aumentar a precisão da previsão de demanda de água.

Um modelo híbrido foi proposto por Jain e Kumar (2007), com a combinação de técnicas convencionais e ANN, onde a remoção de tendências e variação de sazonalidade, pode ser realizada antes de apresentar gradativamente os dados modificados da série temporal à ANN. Os resultados obtidos sugerem que a abordagem de combinar os pontos fortes das técnicas convencionais e de ANN fornecem uma estrutura de modelagem mais robusta, capaz de capturar a natureza não linear de séries temporais complexas e, portanto, gerar previsões mais precisas.

Em Odan e Reis (2012) é apresentado um modelo híbrido para previsão de demanda de água associando ANN à série de *Fourier*. Os autores utilizaram MLP com algoritmo de retropropagação (MLP-BP), DAN2 e duas ANNs híbridas. Os modelos híbridos utilizaram o erro produzido pela previsão da série de Fourier como entrada para a MLP-BP e DAN2, denominadas ANN-H e DAN2-H, respectivamente. O modelo híbrido DAN2-H foi o que apresentou o melhor desempenho em relação aos demais modelos, tanto para previsão da próxima hora quanto para previsão das próximas 24 horas.

Outros autores desenvolveram estudos sobre o uso de modelos híbridos para previsão de séries temporais, não necessariamente para previsão de demanda hídrica, onde a modelagem é realizada de forma a combinar diferentes métodos que possam considerar as características lineares e não lineares das séries temporais. A motivação para o uso de modelos híbridos, segundo Zhang (2003), vem do fato de que nem sempre é possível determinar se uma série temporal em estudo é gerada a partir de um processo subjacente não linear ou linear, ou mesmo se um método particular é mais eficaz do que outro para previsão fora da amostra.

Um modelo híbrido ARIMA-ANN para previsão de séries temporais foi proposto por Babu e Reddy (2014). Eles levaram em consideração a natureza da volatilidade da série temporal usando um filtro de média móvel e, em seguida, aplicaram o modelo ARIMA-ANN. Para o teste e treinamento usaram o conjuntos de dados de manchas solares, dados

de preços de eletricidade e dados do mercado de ações. Os resultados obtidos mostram que, tanto para previsão de uma etapa quanto de várias etapas, o modelo híbrido padrão teve maior precisão de previsão do que os outros modelos individuais ARIMA e ANN e também do que os modelos híbridos ARIMA-ANN propostos por Zhang (2003) e Khashei e Bijari (2011).

No trabalho apresentado por Oliveira e Ludermir (2016), os autores propuseram um sistema híbrido de decomposição evolutiva onde a série temporal é decomposta em seus padrões de baixa e alta volatilidade com um filtro de suavização exponencial. Em seguida, o termo de alta volatilidade é decomposto em padrões lineares e não lineares, aplicando-se os modelos Autorregressivo (AR) e Média Móvel (MA), além do SVR, respectivamente. Além disso, a otimização de parâmetros é realizada usando um algoritmo PSO para produzir previsões mais precisas. O modelo proposto obteve resultados promissores, aumentando, porém, o esforço computacional em relação a outros modelos, considerando o pior caso no desempenho do algoritmo; onde foram utilizados 13 conjuntos de dados da literatura e seis métodos de previsão, incluindo os métodos de Zhang (2003), Babu e Reddy (2014) e ARIMA. Os resultados mostraram que nenhum método foi melhor para todos os conjuntos de dados e, em alguns casos, um modelo linear obteve melhores resultados (menores erros).

Zhang (2003) e Wang et al. (2013) adotaram uma abordagem híbrida do ARIMA-ANN. Enquanto o primeiro utilizou um modelo aditivo, representado pela Equação (2.6), onde o resultado da previsão é dado pela soma da previsão feita no ARIMA mais o resultado da ANN, o segundo utilizou um modelo multiplicativo, representado pela Equação (2.7), para chegar ao resultado da predição, ou seja, a predição é o resultado da multiplicação da predição dada pelo modelo ARIMA pela predição extraída da ANN. Em ambos os modelos (aditivo e multiplicativo), y_t é o valor real da série temporal, L_t que representa a parte linear da série temporal, e N_t representa a parte não linear.

Para realizar a previsão utilizando os modelos aditivo ou multiplicativo, o passo inicial é obter a previsão da parcela linear da série temporal, utilizando o modelo ARIMA e tendo como entrada a série y_t . A série temporal resultante é denominada \hat{L}_t , portanto o valor real y_t deve ser comparado com o valor obtido para extrair a série de componentes não lineares ε_t . Depois, o modelo de ANN pode ser treinado usando os componentes de ε_t como entrada para que possa ser usado para prever a série \hat{N}_t .

As equações (3.1) e (3.2), representam o resultado da série temporal prevista para os modelos aditivo e multiplicativo respectivamente, em que \hat{y}_t é a série temporal prevista, que para Zhang (2003) será a soma de \hat{L}_t e \hat{N}_t , e para Wang et al. (2013) será o produto de \hat{L}_t por \hat{N}_t

$$\hat{y}_t = \hat{L}_t + \hat{N}_t, \quad (3.1)$$

$$\hat{y}_t = \hat{L}_t \times \hat{N}_t. \quad (3.2)$$

Para Candelieri (2017), a ideia básica das medidas de erro é comparar as previsões a partir da observação e a medida mais utilizada no campo da previsão de demanda de água é o MAPE. Neste trabalho, as métricas de erro MSE, RMSE, MAE e MAPE

foram adotadas para análise comparativa dos resultados obtidos com a implementação dos modelos.

Na Tabela A.1, consta o resultado da pesquisa de trabalhos relacionados à previsão de demanda de água para curto, médio ou longo prazo, conforme a finalidade. Esta tabela também informa as métricas de avaliação utilizadas por cada autor, onde é possível observar que a métrica MAPE é uma das mais recorrentes para este tipo de predição.

CAPÍTULO 4

PREVISÃO DE DEMANDA DE ÁGUA PARA A REGIÃO METROPOLITANA DE SALVADOR

4.1 PROPOSTA

A proposta deste trabalho foi seguir um roteiro metodológico de pesquisa para efetuar o estudo e implementação de modelos de Aprendizagem de Máquina que pudessem efetuar a previsão da demanda de água da RMS (com granularidade horária, diária e semanal), utilizando dados de consumo medidos na saída dos reservatórios que compõem o sistema integrado de abastecimento de água da região, como também informações sobre variáveis meteorológicas disponíveis.

Embora o sistema integrado de abastecimento que atende a RMS seja composto por 16 reservatórios, apenas 11 recebem água diretamente das estações de tratamento, os outros 05, portanto, são abastecidos indiretamente pelos primeiros. Por isso, o estudo se concentrará na predição de vazão para os 11 reservatórios que recebem água diretamente das estações de tratamento, pois é a partir destes reservatórios que a água potável é distribuída para as zonas, bairros e reservatórios intermediários garantindo o abastecimento contínuo de água para a população.

Além da análise comparativa sobre a performance alcançada para os modelos implementados, também foi uma proposta deste trabalho, a partir dos resultados alcançados eleger os modelos mais bem ajustados para cada reservatório e publicá-los em produção para disponibilizar as informações em *dashboards* atualizados da PIPPE, para que os operadores possam monitorar as previsões e otimizar o processo de tomada de decisão sobre a operação do sistema.

4.2 METODOLOGIA

O fluxo das ações propostas foi organizado em fases que correspondem aos grupos de atividades que foram realizadas durante o trabalho de pesquisa, iniciando com o mapeamento e coleta dos dados; pré-processamento; preparação e implementação; calibragem e treinamento; predição e coleta dos resultados; e, por fim, análise dos resultados e conclusão.

4.3 MAPEAMENTO E COLETA DOS DADOS

Foi necessário efetuar um estudo prévio sobre o funcionamento do sistema integrado de abastecimento, para ao menos compreender de maneira macro a lógica de funcionamento da rede de abastecimento de água. Como são inúmeros sensores instalados em todo o sistema e cada um deles registra diversas medidas a depender do uso e da sua localização, tornou-se um trabalho árduo identificar corretamente os sensores que correspondessem ao registro de entrada dos reservatórios e, a partir deste ponto, identificar as medidas de vazão para cada um deles.

A Figura 1.1 mostra um panorama geral do Sistema Integrado de Abastecimento de água, que é composto por:

- 05 captações situadas nas barragens: Ipitanga I e II, Joanes I e II, Pedra do Cavallo;
- 04 estações de tratamento de água (ETA Principal, ETA Suburbana, ETA Vieira de Mello e ETA Teodoro Sampaio);
- 16 reservatórios que compõem o Sistema de Distribuição: R1, R7, R25, R14, R17, R18, R21, R3, R19, R15, R5, R20, R10, R12, R23B, R23A.

O mapeamento foi realizado utilizando informações da PIPPE, sobre o sistema de abastecimento integrado da RMS, além da coleta de dados realizada junto à área de Suporte Operacional da EMBASA, em Salvador, Bahia, Brasil e corresponderam ao período de janeiro/2017 a fevereiro/2022, com frequência de um registro por hora, com uma vazão em litros/hora dos reservatórios que abastecem a região. A partir das informações levantadas foi possível identificar todos os reservatórios responsáveis pelo abastecimento da região, com dados valiosos sobre a capacidade e vazão de cada um deles, bem como as zonas e bairros que se beneficiam com a água fornecida por cada reservatório. Os dados da PIPPE são obtidos através de sensores instalados ao longo da rede de abastecimento, alguns situados na entrada de cada reservatório que recebe água das estações de tratamento. Estes dados foram capturados e armazenados em uma base de dados relacional. A Tabela 1.1 mostra a capacidade de vazão e as regiões que se beneficiam com a água fornecida por cada reservatório. A ferramenta da ©SAP, chamada SAP Data Services Designer, foi utilizada para extração, transformação e carga dos dados, sendo disponibilizados em arquivos para etapa seguinte de pré-processamento.

Como o processo de coleta de dados dos reservatórios é realizado através de sensores instalados fisicamente nos pontos de monitoramento, existem fatores que podem interferir na qualidade dos dados coletados, como, por exemplo, a falta ou queda de energia, quebra ou furto de equipamentos e até defeito no funcionamento de sensores. Por isso, apenas a coleta dos dados, em si, não garante a consistência das informações, sendo necessário a realização de uma etapa de análise e pré-processamento dos dados para tratamento de possíveis inconsistências encontradas, como, por exemplo, preenchimento de dados ausentes ou inválidos.

Os dados meteorológicos foram obtidos a partir do banco de dados de informações meteorológicas do Instituto Nacional de Meteorologia do Brasil (INMET), que incluem as variáveis: temperatura em graus Celsius(°C), precipitação em milímetros (mm), pressão

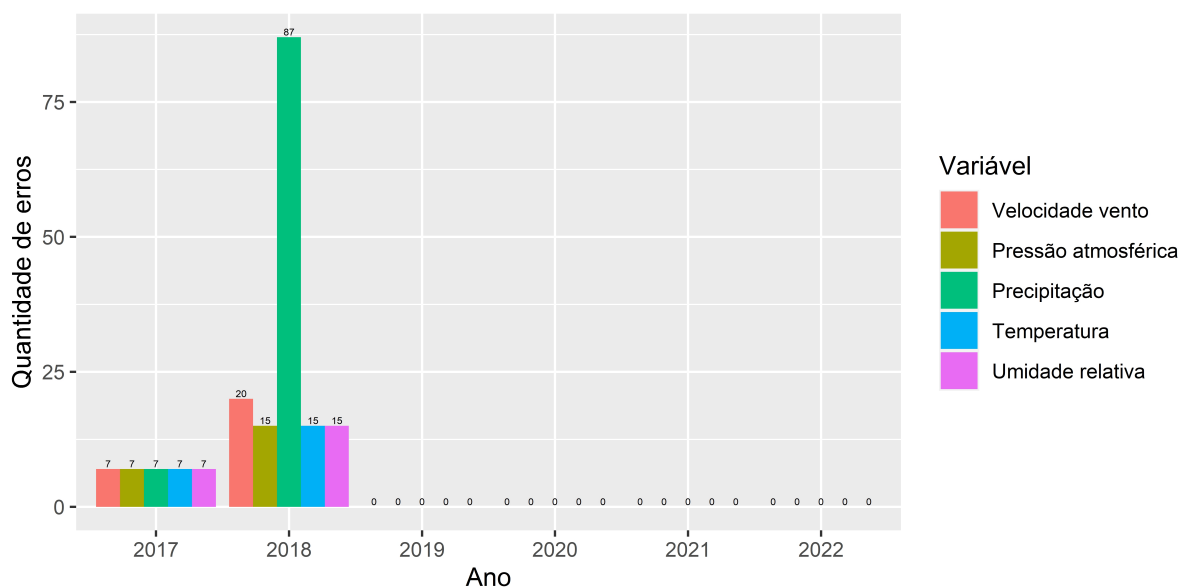
atmosférica em milibares (mB), umidade relativa do ar em percentual (%) e velocidade do vento em metros por segundo (m/s).

4.4 PRÉ-PROCESSAMENTO

A linguagem R foi utilizada para pré-processamento dos dados, utilizando a técnica de interpolação de médias móveis para preenchimento dos dados ausentes e inválidos. Nessa etapa, ocorreu também a junção dos dados de vazão dos reservatórios com os dados meteorológicos, compatibilizando-os para o mesmo padrão horário a ser utilizado nas próximas etapas.

Analisando os dados meteorológicos de 2017 a 2022 (Figura 4.1), observa-se a ocorrência de entradas de dados inválidos nos anos de 2017 e 2018, antes do pré-processamento, onde foram efetuados os tratamentos para substituição dos dados inválidos. Dados inválidos são aqueles cujos valores são negativos ou atípicos, ou seja, estão muito distantes dos demais valores presentes na série e por isso são considerados *outliers*.

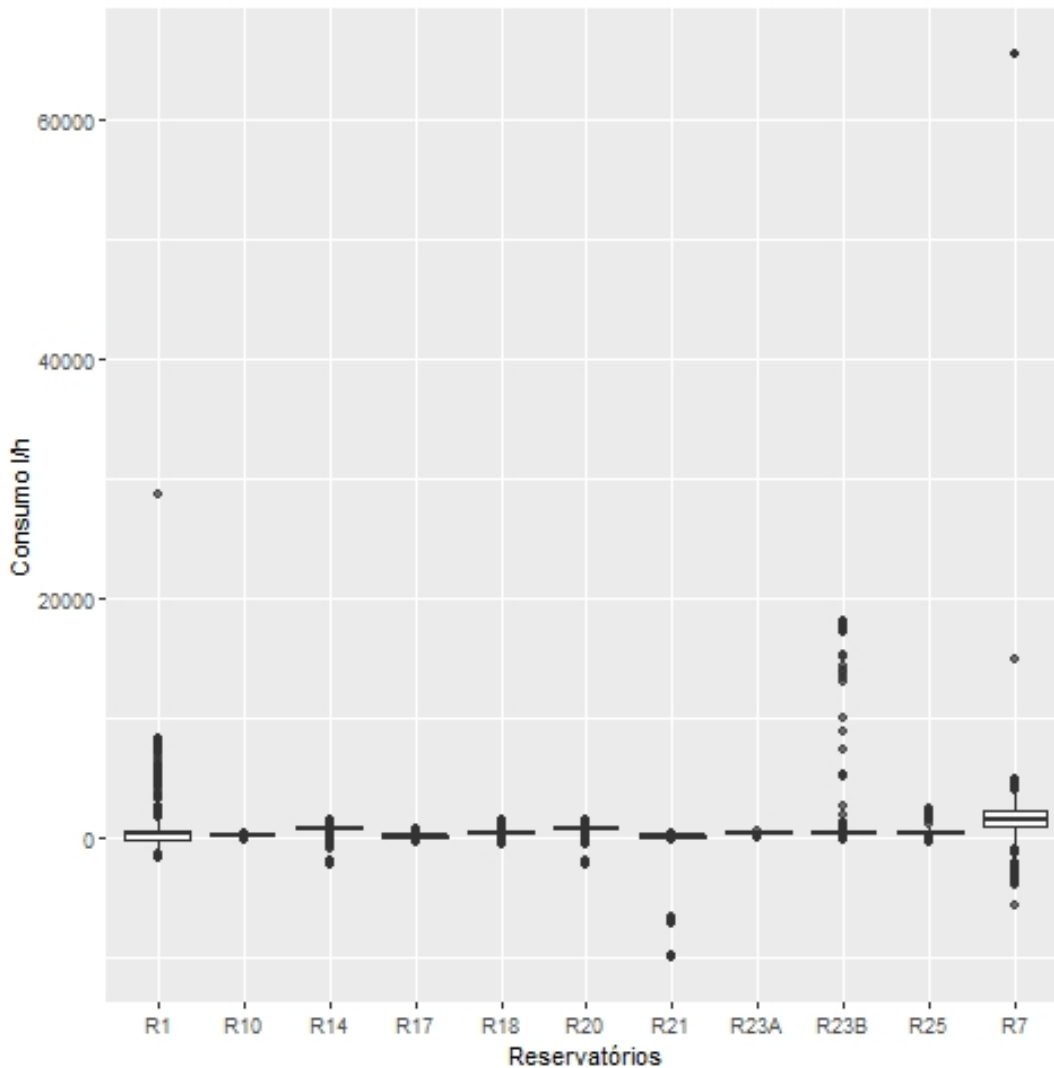
Figura 4.1 Distribuição dos dados meteorológicos inválidos ou ausentes, antes do pré-processamento, por período



Fonte: Próprio autor.

A Figura 4.2 mostra o *boxplot* para os dados de consumo dos 11 reservatórios estudados. É possível observar a existência de *outliers*, como também de pontos com valores negativos.

As Tabelas 4.1 e 4.2 mostram o detalhamento das medidas descritivas, extraídas a partir da análise sobre o conjunto de dados dos reservatórios; com o maior Desvio Padrão (SD, do inglês *Standard Deviation*) registrado nos reservatórios R7, R1 e R23B, onde os valores encontrados estão mais distantes uns dos outros, sendo considerados mais

Figura 4.2 Gráfico *boxplot* para os valores de vazão em litros/hora dos reservatórios.

Fonte: Próprio autor.

heterogêneos; enquanto o menor desvio padrão foi registrado no reservatório R10, onde os dados são mais homogêneos. O Coeficiente de Variação (CV), medida relativa de dispersão, que indica, em termos relativos, o grau de concentração dos dados em torno da média, indicou uma média dispersão nos dados, com o $CV \geq 15\%$ e $CV < 30\%$, para os reservatórios R10, R14, R18, R20 e R23A. Enquanto que os reservatórios R1, R17, R21, R23B, R25 e R7 indicaram alta dispersão, com $CV \geq 30\%$. No aspecto de distribuição, nenhuma das frequências dos reservatórios apresentou uma distribuição simétrica. O Coeficiente de Assimetria (CA) avalia a assimetria de um conjunto de dados e assume valores entre -1 e 1, sendo que os reservatórios R1, R23B e R7 apresentaram distribuição assimétrica positiva à direita e os reservatórios R10, R14, R17, R18, R20, R21, R32A e R25 apresentaram distribuição assimétrica negativa à esquerda.

Apenas para o R23A não havia dados suficientes para predição, pois como mostra a

Tabela 4.1 Mínimos, Máximos e Quartis por reservatório.

RSVR	Mínimo	1º Quartil	2º Quartil	3º Quartil	Máximo
R1	-1.661,2	549,2	630,0	690,6	28.845,3
R10	-119,4	314,7	348,1	368,5	491,2
R14	-2.182,6	782,6	863,6	942,7	1.638,2
R17	-197,3	182,2	282,9	334,1	787,6
R18	-388,7	525,6	557,1	604,7	1.599,1
R20	-2.140,5	790,9	825,0	855,3	1.523,9
R21	-9.950,0	114,3	181,2	210,3	423,5
R23A	0	419,1	558,4	565,4	706,2
R23B	-87,4	416,9	467,9	489,2	18.204,2
R25	-224,2	442,0	546,9	623,7	2.481,0
R7	-2.547,6	3.300,4	3.512,9	3.760,1	67.706,6

Fonte: Próprio autor.

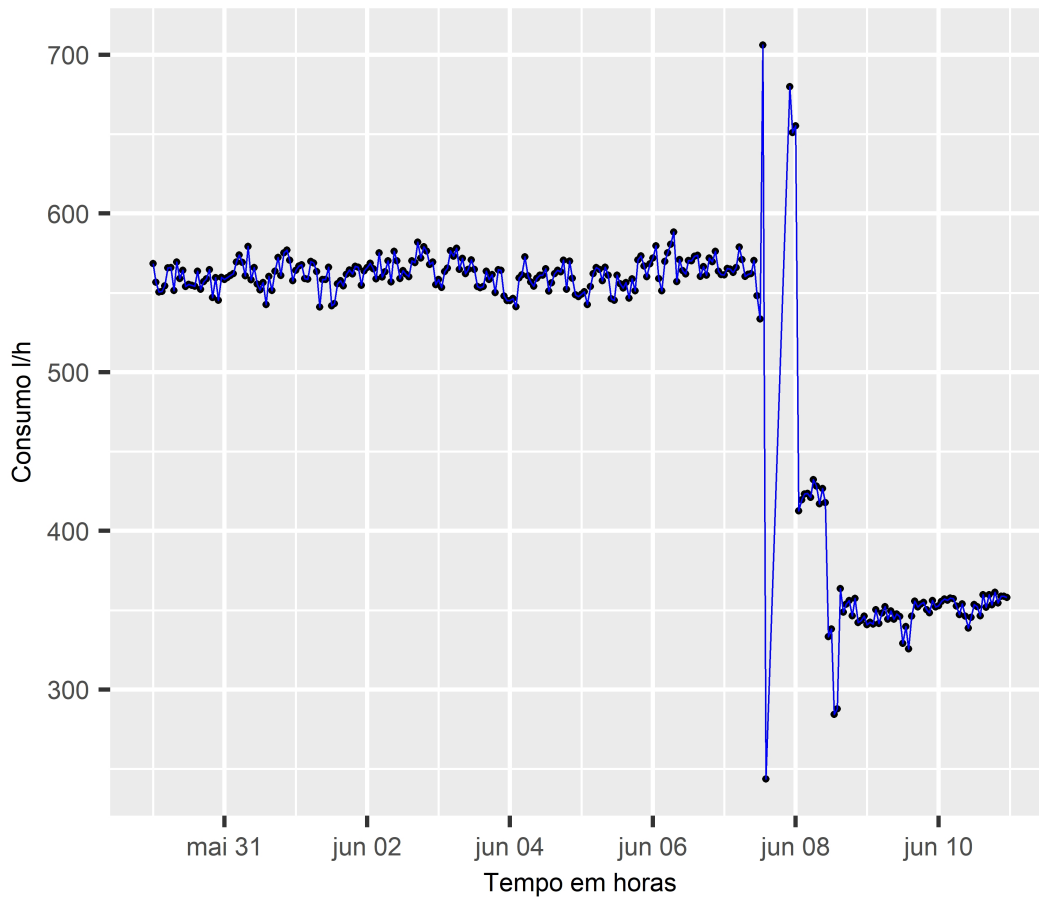
Tabela 4.2 Média, Mediana, SD, Variância, CV e CA por reservatório.

RSVR	Média	Mediana	SD	Variância	CV (%)	CA
R1	757,2	630,0	1.106,8	1.224.957,0	146,2	6,9
R10	333,0	348,1	67,0	4.488,0	20,1	-2,4
R14	848,8	863,6	178,2	31.766,1	21,0	-2,7
R17	258,1	282,9	103,5	10.709,0	40,1	-0,6
R18	551,1	557,1	116,2	13.510,9	21,1	-3,3
R20	804,7	825,0	146,6	21.502,3	18,2	-4,9
R21	152,7	181,2	231,8	53.731,4	151,8	-35,4
R23A	498,3	558,4	121,3	14.717,0	24,3	-2,0
R23B	515,8	467,9	1.059,7	1.122.904,0	205,5	16,1
R25	488,8	546,9	239,5	57.381,2	49,0	-1,4
R7	3.527,1	3.512,9	1.838,0	3.378.094,0	52,1	28,2

Fonte: Próprio autor.

Figura 4.3 só havia dados de vazão para um período de dois meses, ainda assim com um segmento de dados inválidos registrados no início de junho causado por falha nos equipamentos, seguido por um movimento de queda na vazão horária média do reservatório. Os reservatórios R1, R21, R23B e R7 também não possuíam dados a partir de 01/01/2017, mas puderam ser estudados porque tiveram um período maior de coleta, o R1 teve dados a partir de 01/04/2019, o R21 a partir de 27/11/2017, o R23B a partir de 08/11/2019 e, por fim, o R7 desde 08/11/2017.

Para os outros 10 reservatórios analisados, foram encontrados e extraídos dados para um período mínimo de três anos, coletados a partir dos sensores de vazão instalados na entrada dos reservatórios. Estes dados apresentaram diversas situações de invalidez, em função de queda de energia, quebra ou furto de equipamentos e falhas na leitura de sensores, como pode ser observado na Figura 4.4, onde existem diversos registros

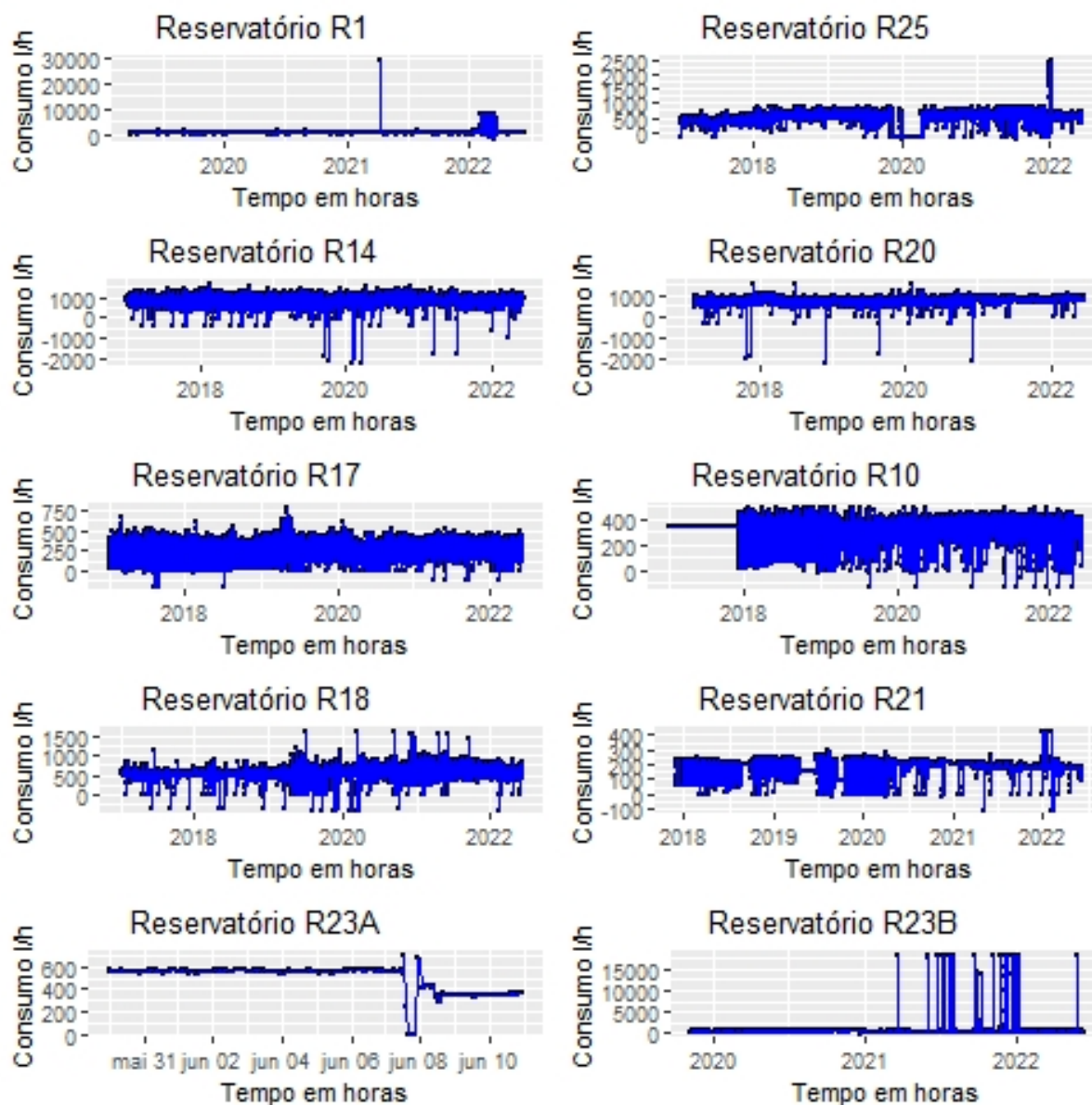
Figura 4.3 Histórico de vazão do reservatório R23A, com apenas 2 meses de registros de vazão.

Fonte: Próprio autor.

de leitura de vazão negativos ou mesmo leituras com valores excessivamente elevados, indicando essas falhas.

Foi necessário o desenvolvimento de um algoritmo para identificação e substituição dos dados inválidos, utilizando a interpolação de média móvel, como regra o cálculo da vazão média entre o registro válido anterior e o registro válido imediatamente posterior à leitura defeituosa.

Figura 4.4 Histórico de vazão dos reservatórios R1, R7, R10, R14, R17, R18, R20, R21, R23B e R25 com dados inválidos, por período.

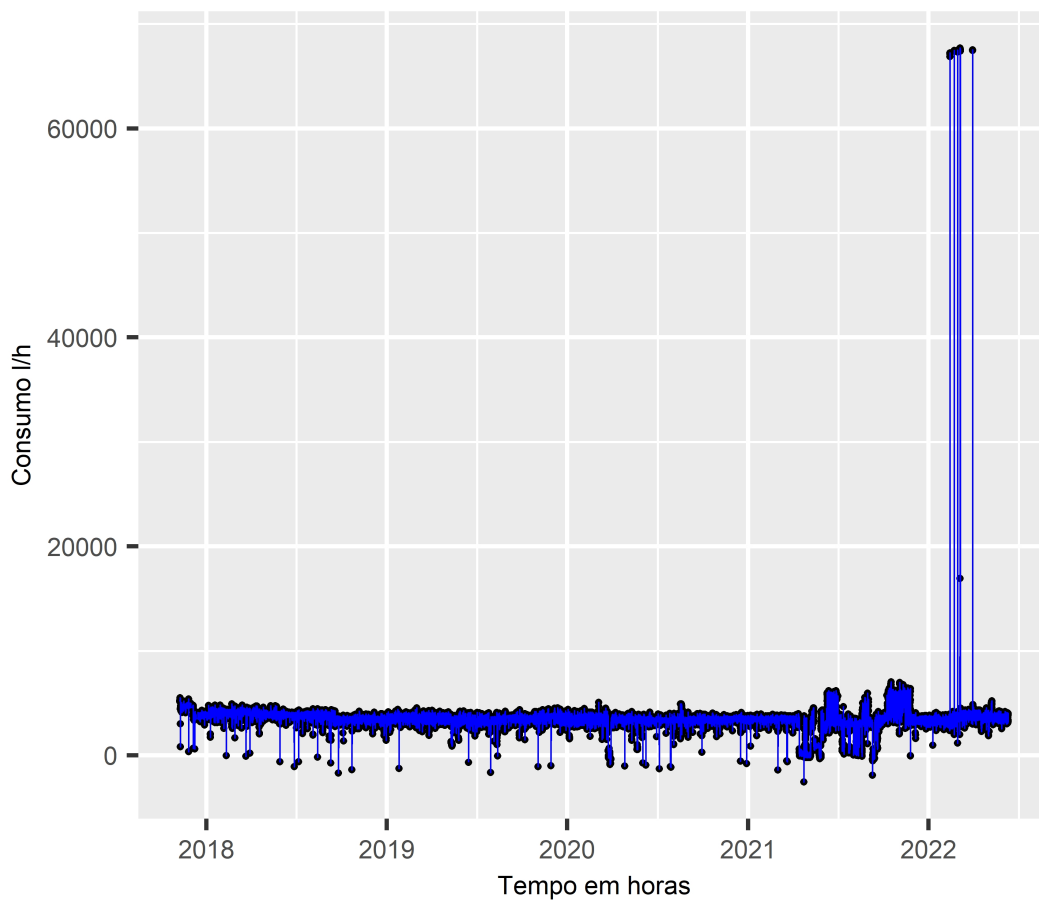


Fonte: Próprio autor.

Observa-se na Figura 4.5 a situação dos dados de vazão do reservatório R7 antes do pré-processamento, onde claramente é possível observar valores negativos e leituras excessivamente altas.

Na Figura 4.6, tem-se a situação após a etapa do pré-processamento, para o mesmo período comparativo de 2017 a 2022. Observa-se que não existem mais registros negativos, pois eles foram substituídos por valores estimados utilizando interpolação de médias móveis durante a fase de pré-processamento dos dados.

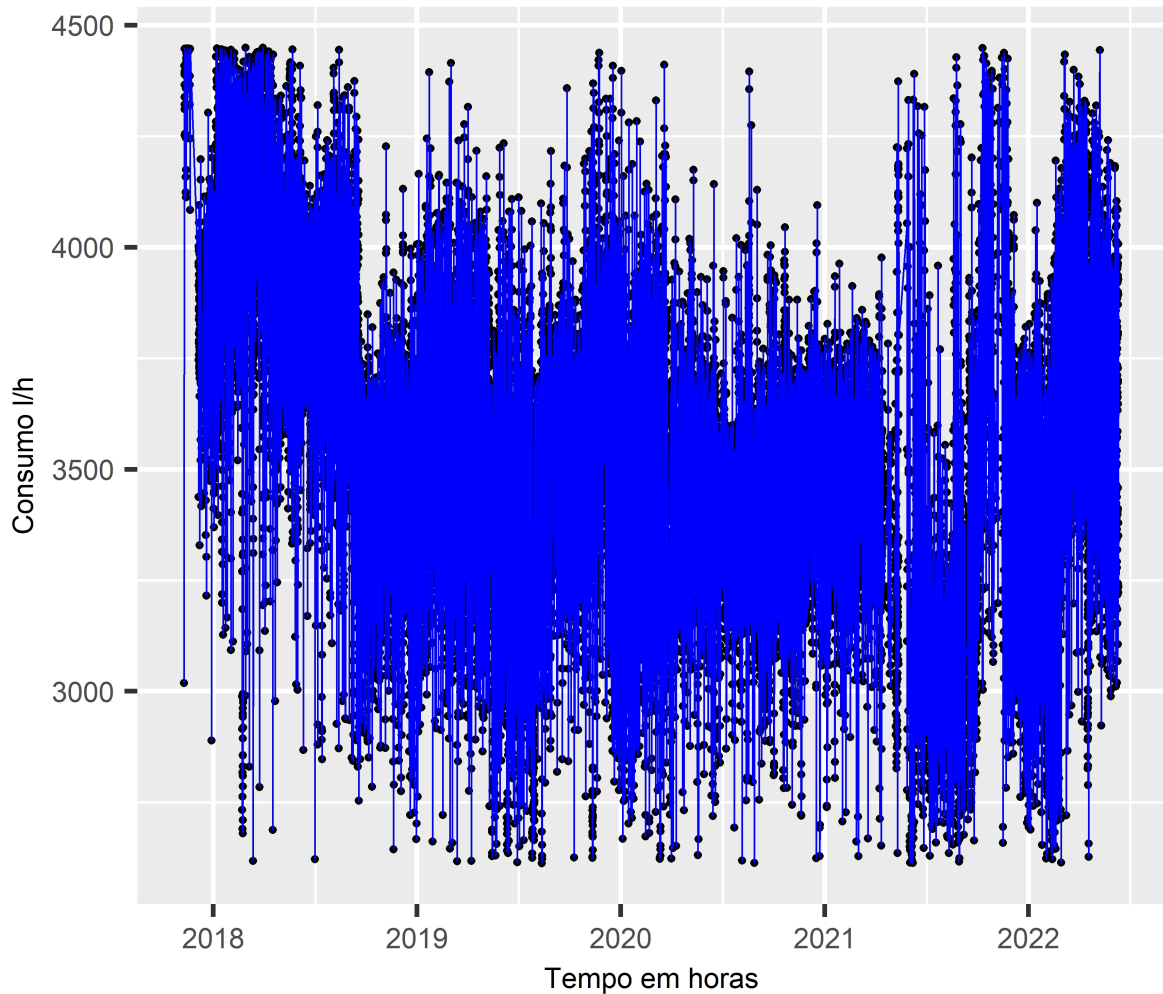
Figura 4.5 Histórico de vazão do reservatório R7, com dados inválidos, por período.



Fonte: Próprio autor.

A mesma função de correção aplicada aos dados inválidos identificados no reservatório R7 também foi aplicada aos demais reservatórios, como pode ser observado na Figura 4.7.

Figura 4.6 Histórico de vazão do reservatório R7, após o pré-processamento, por período.



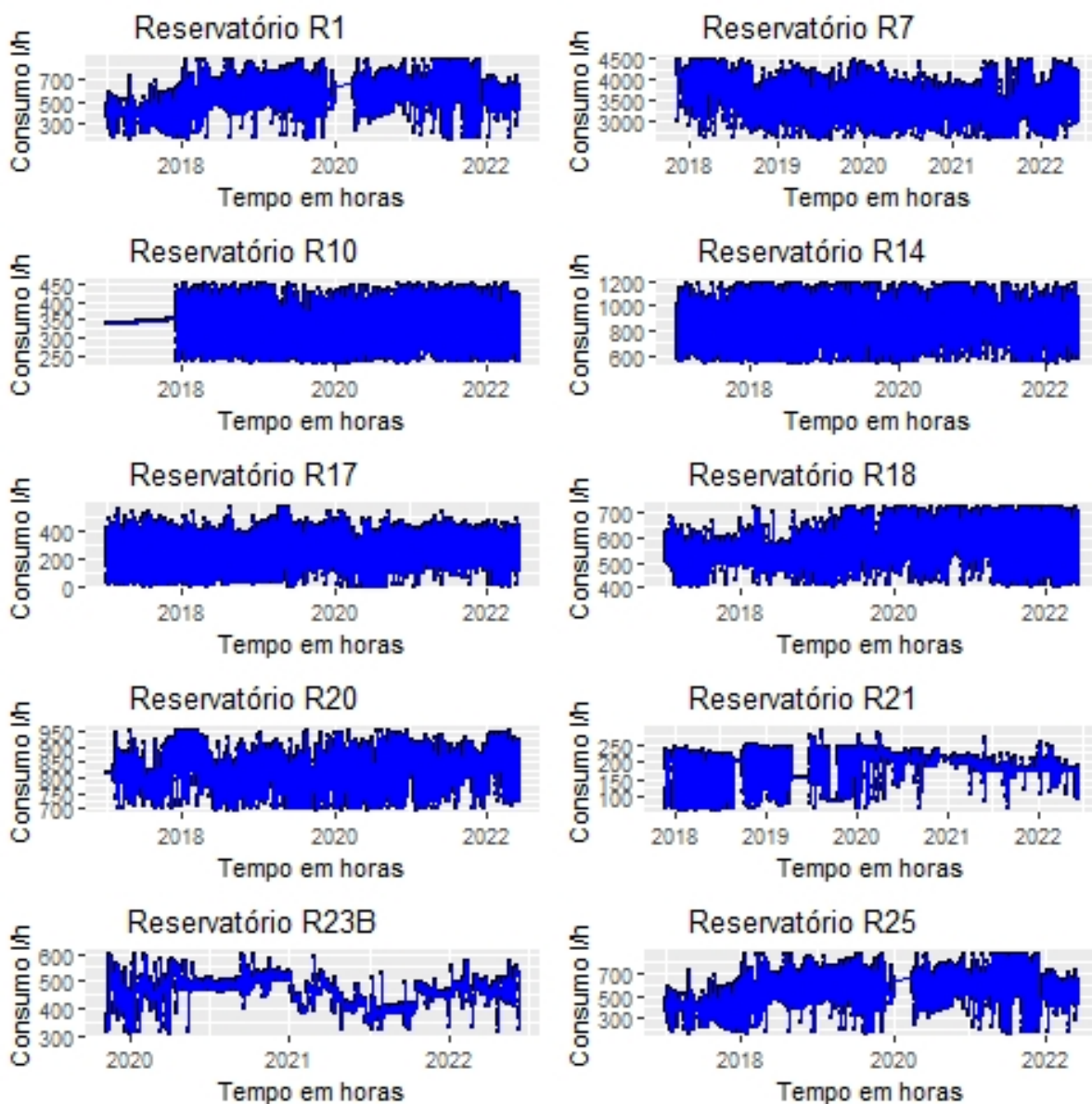
Fonte: Próprio autor.

4.5 CALIBRAGEM E TREINAMENTO

Após a fase de pré-processamento, os dados foram preparados para entrada nos modelos de Aprendizagem de Máquina utilizando linguagem R ou Python, com o objetivo de efetuar as previsões para os períodos horários, diários e semanais utilizando os métodos de aprendizagem baseados em ANN, SVR, ARIMA e modelos híbridos como ARIMA-ANN. Para isso, foi necessário aplicar a técnica de janela deslizante devido à dependência temporal entre as observações, característica peculiar das séries temporais, onde sabe-se que leituras de instantes passados influenciam a leitura de instantes futuros.

Como foram identificados 10 reservatórios responsáveis por abastecer toda a região, foi necessário separar os dados de consumo de cada um deles e combiná-los com os dados meteorológicos da região, utilizando a mesma medida de tempo e granularidade (horário,

Figura 4.7 Histórico de vazão dos reservatórios R1, R7, R10, R14, R17, R18, R20, R21, R23B e R25 após o pré-processamento, por período.



Fonte: Próprio autor.

diário e semanal), com o objetivo de identificar os modelos que mais se adequassem a cada reservatório por granularidade de tempo para predição. Além disso, durante o treinamento, os modelos de aprendizagem implementados também foram calibrados com ajustes de hiperparâmetros e validação cruzada dos dados, sempre em busca da melhor configuração que otimizasse a performance do modelo com tendência a reduzir a taxa de erros na predição.

O conjunto de dados de entrada utilizado para a previsão foi composto pelas variáveis

Tabela 4.3 Combinação das variáveis de entrada usadas para treinamento e teste.

Variáveis de entrada
$P(t), Pa(t), T(t), U(t), W(t), C(t)$
$P(t), Pa(t), T(t), U(t), W(t), C(t), C(t - 1)$
$P(t), Pa(t), T(t), U(t), W(t), C(t), C(t - 1), C(t - 2)$
$P(t), Pa(t), T(t), U(t), W(t), C(t), C(t - 1), C(t - 2), C(t - 3)$
$P(t), Pa(t), T(t), U(t), W(t), C(t), C(t - 1), C(t - 2), C(t - 3), C(t - 4)$
$P(t), Pa(t), T(t), U(t), W(t), C(t), C(t - 1), C(t - 2), C(t - 3), C(t - 4), C(t - 5)$

Fonte: Próprio autor.

precipitação (P), pressão atmosférica (Pa), temperatura (T), umidade relativa do ar (U), velocidade do vento (W), e consumo (C), mais um conjunto de uma a cinco variáveis contendo a defasagem de atraso do consumo ($t - 1$ a $t - 5$). Esse conjunto de variações foi usado para prever o consumo nos períodos horário ($t =$ hora), diário ($t =$ dia) e semanal ($t =$ semana), gerando, assim, 18 combinações de treinamento para cada um dos modelos testados. A Tabela 4.3 mostra as variações do conjunto de dados de entrada usado para treinamento e teste em cada um dos modelos.

Os dados de vazão de consumo (C) foram agrupados por reservatório e separados em treinamento e teste, sendo 75% para treinamento e 25% para teste, submetidos a sete modelos diferentes, sendo cinco modelos mais tradicionais (SVR, MLP, ARIMA1, ARIMA2 e LSTM) e dois modelos híbridos inéditos (SVR-MLP e SVR-LSTM), para previsão horária, diária e semanal.

4.6 PREDIÇÃO E COLETA DOS RESULTADOS

Os métodos empregados neste trabalho foram implementados nas linguagens **R** e **Python** e estão disponíveis no repositório GitHub (<<https://github.com/edmilsondejesus/waterdemand/>>) para permitir a verificação e replicabilidade da pesquisa; foram eles:

4.6.1 MLP

A rede neural MLP foi implementada utilizando à linguagem **Python** e o número ótimo de neurônios da camada oculta foi determinado usando o procedimento exaustivo de tentativa e erro. Os dados foram divididos em cinco subconjuntos para utilização da técnica de validação cruzada específica para séries temporais, usando a função *TimeSeriesSplit* que é uma variação do *KFold* e respeita a temporalidade das observações. Em cada subconjunto, 75% dos dados foram utilizados para treinamento e 25% para teste. A função *GridSearchCV* foi utilizada para otimização dos hiperparâmetros e a configuração ótima da rede neural foi a MLP(6,12,1). O algoritmo *adam* foi escolhido para otimização dos pesos e da função identidade como ativação. Uma vez obtida a melhor arquitetura MLP, ela foi usada para calcular as estatísticas de desempenho nos dados de teste. O menor MAPE registrado com o modelo MLP foi 0,94% para predição horária do reservatório R23B, com *alpha* de 0,01, com o número máximo de 500 iterações.

4.6.2 SVR

O modelo SVR foi implementado na linguagem Python com a função Kernel *rbf*, também dividido em cinco subconjuntos de dados para validação cruzada, respeitando a temporalidade das observações e em cada subconjunto 75% dos dados foram usados para treinamento e 25% para teste. Para a variável de consumo horário medido em litros/hora, denominada VL_MEDICAO, foi aplicada uma defasagem de um passo de atraso (*lag*). Os hiperparâmetros *C*, *epsilon* e *gamma* também foram calibrados usando o procedimento exaustivo de tentativa e erro. O menor MAPE encontrado com o modelo SVR foi 2,70%, para predição diária no reservatório R10 para o *C*=12550, *epsilon*=0,01, *gamma*=1e-08, *kennel*=*rbf* e 5 *lags* de atraso.

4.6.3 ARIMA

Duas variações do modelo ARIMA foram implementadas; o primeiro modelo, denominado ARIMA1, foi implementado na linguagem R e os parâmetros de ordem ótima para AR(*p*), diferenciação *d* e MA(*q*) foram definidos usando a função `auto.arima`. Os dados foram divididos em dois subconjuntos de treinamento e validação, tanto para as variáveis independentes quanto para a variável dependente, sendo que em cada subconjunto 75% dos dados foram utilizados para treinamento e 25% para teste. A configuração ótima encontrada foi o ARIMA(1, 1, 1), escolhida com base no Critério de Informação de Akaike (AIC, do inglês *Akaike Information Criterion*).

O segundo modelo, denominado ARIMA2, também foi implementado na linguagem R utilizando a mesma metodologia aplicada ao ARIMA1, porém, o parâmetro `lambda = "auto"` foi ativado na função `auto.arima` para aplicação da função de transformação Box-Cox, que visa estabilizar a variâncias e com isso tornar os dados mais próximos de uma distribuição normal. O `lambda` encontrado foi 0,1827694 e foi utilizado para instanciar o modelo. Neste caso, a configuração ótima encontrada foi ARIMA(2, 1, 2) e MAPE de 10,83.

4.6.4 Modelo híbrido SVR-ANN

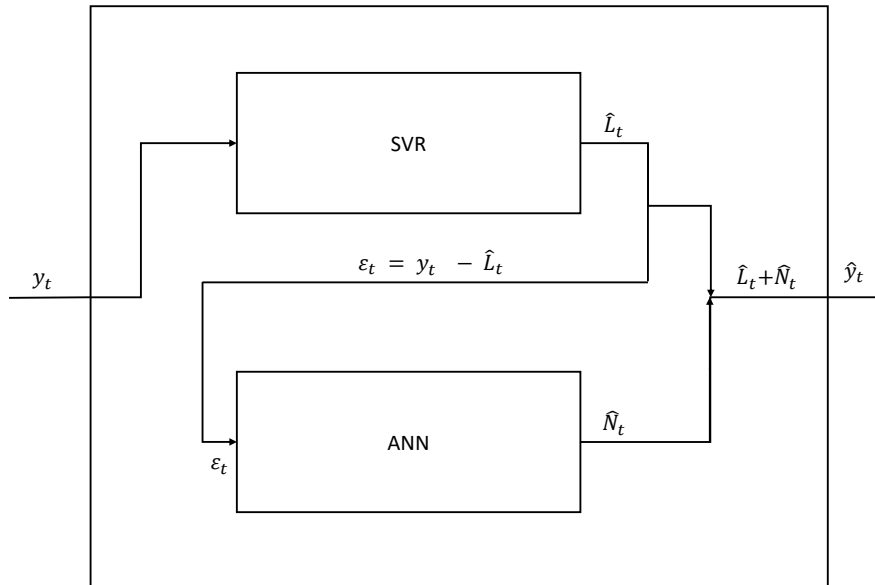
O modelo SVR-ANN foi implementado usando a mesma metodologia proposta por Zhang (2003) para o modelo aditivo, representado pela Equação (2.6), onde L_t representa a parte linear da série temporal, N_t representa a parte não linear.

A Figura 4.8 mostra a arquitetura da SVR-ANN. O passo inicial foi prever a porção linear da série temporal, usando o modelo SVR e a série y_t como entrada. A série temporal resultante foi chamada \hat{L}_t , então subtraída da série real y_t para extrair a porção não linear ε_t , que também pode ser identificada como ruído, conforme Equação (4.1).

A segunda etapa foi analisar a modelagem da componente não linear, usando a série ε_t como entrada para o modelo da ANN. O resultado da predição realizada pela ANN foi denominado \hat{N}_t e foi utilizado para obter o resultado combinado da predição final, na Equação (4.2)

$$\varepsilon_t = y_t - \hat{L}_t, \quad (4.1)$$

$$y_t = \hat{L}_t + \hat{N}_t. \quad (4.2)$$

Figura 4.8 Arquitetura do modelo SVR-ANN.

Fonte: Próprio autor.

4.7 ANÁLISE DOS RESULTADOS E CONCLUSÃO

Após realizada a implementação, ajuste e execução dos modelos de Aprendizagem de Máquina, os resultados de previsão foram coletados e tabulados para análise comparativa sobre o desempenho de cada modelo através das taxas de erros apuradas. É comum a análise comparativa do desempenho dos modelos de previsão, para escolha do método que mais se adequa à necessidade de previsão. Essa prática pode ser verificada nos trabalhos de Ghiassi e Saidane (2005), Alvisi, Franchini e Marinelli (2007), Ghiassi, Zimbra e Saidane (2008), Bakker et al. (2014), Farias et al. (2018) e VanBerlo, Ross e Hsia (2021).

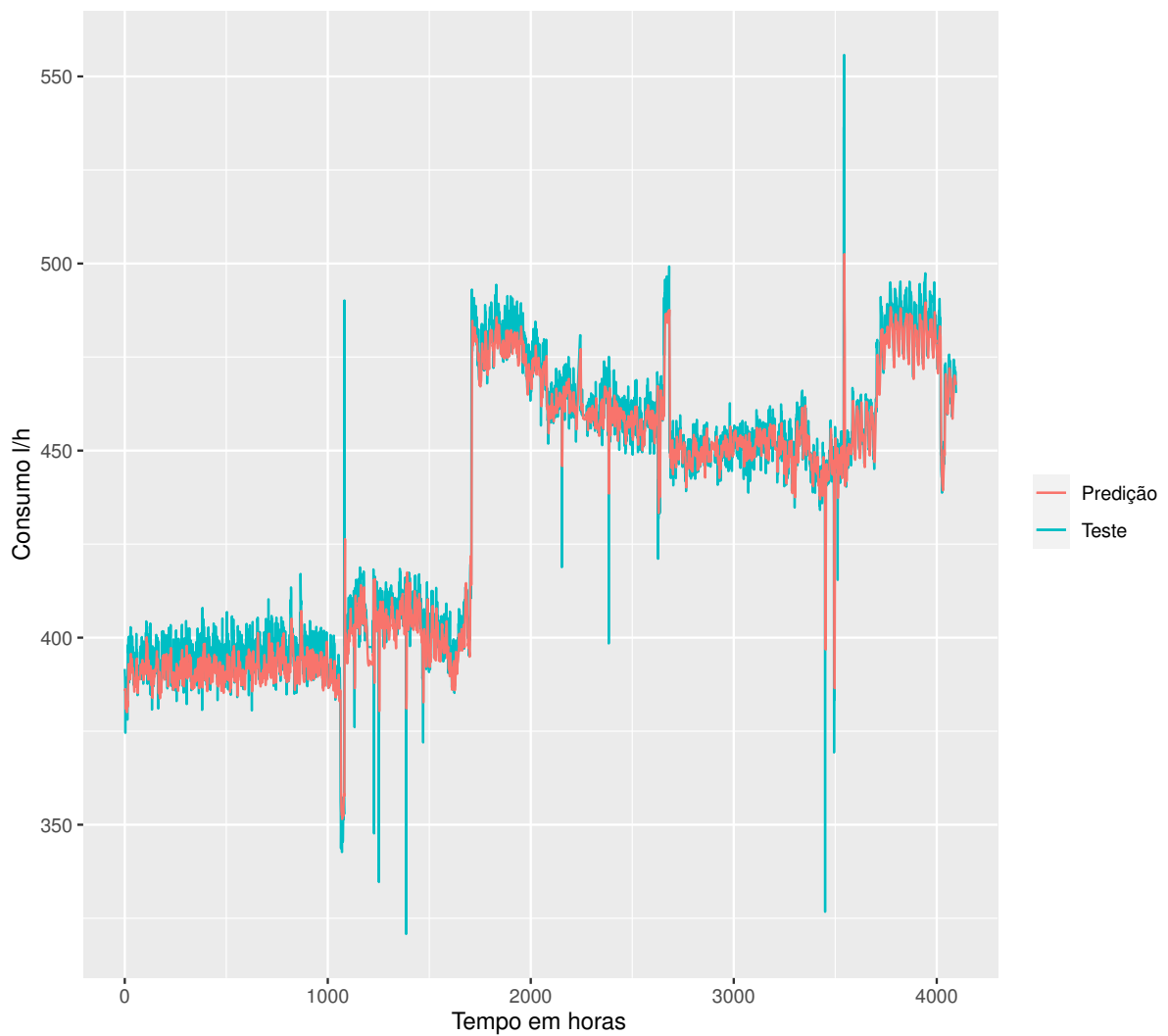
Para Liu, Liu e Tong (2017), a medição dos erros de previsão é crucial para a seleção dos parâmetros dos modelos, bem como para monitorar a precisão e confiabilidade das previsões, uma vez que a degradação do desempenho pode exigir uma atualização dos modelos. Outro ponto importante é o tempo de processamento para previsão, pois como um dos objetivos é disponibilizar um *dashboard* com as previsões de demanda de curto prazo (horas, dias e semanas) as previsões das próximas horas afetam a tomada de decisão sobre a necessidade de manobras operacionais nos sistemas e o atraso nessas previsões anulariam sua funcionalidade.

Para análise comparativa sobre o desempenho dos modelos implementados neste trabalho foram utilizadas como referência as métricas de erro MAPE, MAE, MSE e RMSE. A apresentação e análise dos resultados são apresentados nos capítulos 5 e 6 a seguir.

RESULTADOS ALCANÇADOS

A Figura 5.1 mostra a previsão horária feita pelo modelo LSTM para o reservatório R23B.

Figura 5.1 Previsão horária para o reservatório R23B usando o modelo LSTM com lag de atraso de 3 valores de consumo.



Fonte: Próprio autor.

A Tabela 5.1 mostra a relação dos melhores modelos de previsão horária encontrados para cada um dos dez reservatórios, com a configuração do número de variáveis com

$t - N$ defasagens de atraso na variável consumo. Para a previsão horária, houve predominância na utilização do modelo LSTM, destacando-se o menor MAPE registrado para o reservatório R23B com 0,9% utilizando este modelo. O pior desempenho foi para a série temporal do reservatório R17, que usando o modelo SVR atingiu um MAPE de 8,0%.

Tabela 5.1 Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para a previsão horária.

RSVR	Modelo	Lag	MSE	RMSE	MAE	MAPE (%)	Duração (segundos)
R1	LSTM	1	1.689,4	41,1	5,2	2,3	86,5
R25	LSTM	3	2.372,4	48,7	18,7	4,9	160,8
R14	LSTM	1	3.923,2	62,6	29,6	4,8	118,3
R17	SVR	6	385.887,8	621,2	403,8	8,0	2,5
R18	LSTM	2	1.202,5	34,7	15,7	3,9	314,2
R21	LSTM	5	21,1	4,6	1,4	1,2	120,1
R23B	LSTM	3	50,9	7,1	3,0	0,9	58,1
R20	MLP	3	426,7	20,7	7,7	1,5	246,8
R10	SVR	6	511.559,3	715,2	274,1	6,3	3,3
R7	LSTM	2	34.128,6	184,7	63,4	3,3	125,7

Fonte: Próprio autor.

A Tabela 5.2 mostra a lista dos modelos mais bem ajustados à previsão diária de cada reservatório. E neste caso, não foi evidenciada a predominância de nenhum dos sete modelos em particular, mas os modelos ARIMA se destacaram para este tipo de previsão, sendo a melhor escolha para previsão diária em 50% dos reservatórios. De acordo com a métrica MAPE, o modelo com menor percentual de erro registrado foi o ARIMA1 também para o reservatório R23B, com 1,0%, enquanto a pior métrica MAPE foi registrada para o modelo LSTM na previsão do reservatório R25, com 8,6%.

Para a previsão semanal, também foi aplicada a mesma metodologia de previsão, comparando os resultados obtidos para os sete modelos testados com o conjunto de variáveis meteorológicas e suas variações com defasagens de atraso de consumo. A Tabela 5.3 mostra a lista de modelos que registraram as estatísticas MAPE mais baixas para previsão semanal de demanda de água por reservatório. Há uma predominância em relação aos modelos ARIMA, principalmente o ARIMA1, o que pode ter sido causado pela redução do volume de dados. Por outro lado, também é possível observar que a duração do treinamento, que é medido em segundos, caiu consideravelmente em relação às previsões diárias e horárias, devido à redução do volume de dados, natural devido às necessidades de agregação das variáveis para previsão semanal. As estatísticas MAPE para cada reservatório na ordem semanal também aumentaram em relação às previsões horárias e diárias.

É possível observar na Tabela 5.4 que não houve um modelo que apresentasse predominância em relação à menor estatística MAPE para todos os reservatórios, pois são as características individuais de cada série temporal que determinam o modelo mais adequado para sua previsão. Em relação à duração média das previsões, por modelo, é pos-

Tabela 5.2 Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para previsão diária.

RSVR	Modelo	Lag	MSE	RMSE	MAE	MAPE (%)	Duração (segundos)
R1	ARIMA2	4	3.972,5	63,0	43,6	6,6	146,6
R25	LSTM	2	3.960.186,5	1.990,0	330,7	8,6	8,8
R14	LSTM	3	3.524.767,7	1.877,4	670,0	8,3	9,1
R17	MLP	4	260.528,7	510,4	195,8	5,5	38,6
R18	ARIMA2	1	2.207,5	47,0	36,4	6,4	31,1
R21	SVR	4	58.117,2	241,1	29,0	2,7	73,5
R23B	ARIMA1	4	68,4	8,3	4,6	1,0	78,8
R20	ARIMA1	1	441,4	21,0	13,8	1,6	46,9
R10	MLP	4	510.824,7	714,7	225,0	6,0	36,1
R7	ARIMA1	2	114.778,4	338,8	262,4	7,4	61,2

Fonte: Próprio autor.

sível verificar que apenas o SVR se destaca em relação aos demais, seguido do ARIMA1 e ARIMA2. No entanto, esses modelos não apresentaram predominância entre as menores estatísticas MAPE registradas nas previsões.

A Figura 5.2 mostra o *boxplot* dos resultados MAPE alcançados por cada modelo. É possível observar que os maiores *outliers* foram registrados nos modelos LSTM, SVR e ARIMA2, respectivamente, enquanto os modelos MLP, SVR-MLP e SVR-LSTM apresentaram a menor variabilidade. Embora a implementação do MLP seja mais simples em comparação com os modelos híbridos, o MLP sozinho pode não ser adequado para aplicação em séries temporais com comportamento dependente, ao contrário do modelo híbrido onde o MLP é aplicado aos resíduos que possuem um comportamento mais aleatório.

Os resultados dos testes de Shapiro-Wilk (SHAPIRO; WILK, 1965) aplicados aos valores de MAPE obtidos para cada modelo não apresentaram normalidade, para um nível de significância α de 0,05. Com isso, o teste não paramétrico de Kruskal e Wallis (1952) foi aplicado e demonstrou que existem diferenças nos resultados de três ou mais modelos. O teste Kruskal-Wallis apresentou um p -valor de 0,08359, sendo menor que o α de 0,10, indicando que a hipótese nula de igualdade de variâncias deve ser rejeitada, pois existem diferenças significativas entre os resultados para três ou mais modelos.

A Tabela 5.5 mostra o resultado da aplicação do teste de Dunn (DUNN, 1964) para identificar a diferença entre os modelos. Em cada célula de valor apresentado significa a estatística de teste e seu respectivo p -valor para os modelos da linha e coluna correspondentes. Para o grau de significância α de 0,10, o modelo SVR-MLP apresentou diferença estatisticamente significativa em relação aos modelos ARIMA1, ARIMA2 e SVR. Já o SVR-LSTM também apresentou diferença estatisticamente significativa em relação ao modelo SVR.

Tabela 5.3 Resultados dos melhores modelos de previsão aplicados a cada reservatório (RSVR) para previsão semanal.

RSVR	Modelo	Lag	MSE	RMSE	MAE	MAPE (%)	Duração (segundos)
R1	ARIMA1	3	302.645.886,1	17.396,7	10.687,8	14,2	0,2
R25	ARIMA1	2	230.099.607,0	15.169,0	8.666,1	10,0	0,2
R14	ARIMA1	5	290.432.408,1	17.042,0	7.273,3	5,6	0,3
R17	ARIMA2	6	23.600.185,4	4.858,0	1.479,3	3,39	0,4
R18	ARIMA1	5	97.862.259,8	9.892,5	5.730,9	6,75	0,4
R21	ARIMA2	6	15.436.962,9	3.929,0	1.501,8	4,8	0,5
R23B	SVR	4	150.011.037,4	12.247,9	7.421,0	22,1	0,2
R20	ARIMA1	3	256.936.416,9	16.029,2	6.614,2	4,8	0,2
R10	ARIMA2	2	35.075.147,2	5.922,4	2.320,7	4,3	0,3
R7	ARIMA1	2	13.030.952.288,0	114.153,2	73.111,2	20,2	1,0

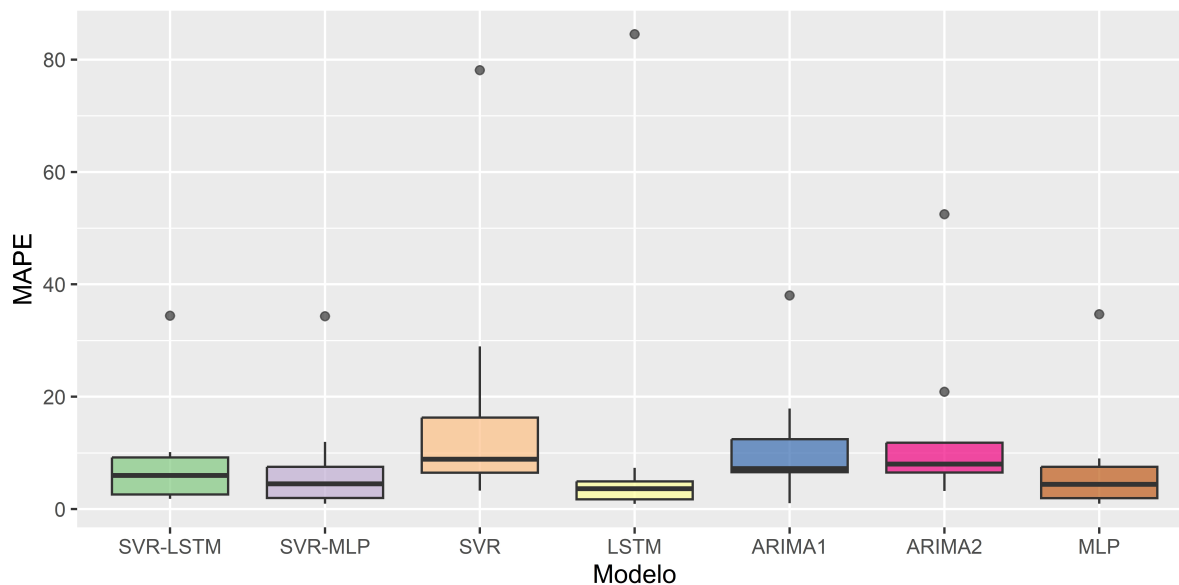
Fonte: Próprio autor.

Tabela 5.4 Resultados das menores estatísticas MAPE por modelo na previsão horária de cada reservatório (RSVR).

RSVR	SVR-LSTM	SVR-MLP	SVR	LSTM	ARIMA1	ARIMA2	MLP
R1	3,54	3,32	28,94	2,31	6,69	6,64	3,25
R25	10,13	11,96	15,70	4,93	11,88	11,85	9,02
R14	7,17	5,82	9,75	4,85	17,89	20,87	6,03
R17	34,41	34,32	8,00	84,53	38,02	52,47	34,69
R18	6,09	4,70	16,46	3,95	6,51	6,45	5,04
R21	2,27	1,33	3,28	1,20	7,02	3,34	1,35
R23B	2,04	0,95	4,39	0,94	1,05	8,52	0,95
R20	1,82	1,54	6,86	1,53	1,63	3,24	1,51
R10	9,87	8,09	6,33	7,34	12,61	11,71	8,01
R7	5,91	4,33	78,10	3,34	7,44	7,56	3,82

Fonte: Próprio autor.

Figura 5.2 Gráfico *boxplot* das estatísticas MAPE por modelo implementado.



Fonte: Próprio autor.

Tabela 5.5 Tabela de diferenças entre modelos, com estatísticas de teste e p -valor respectivamente calculados para cada par.

Modelo x Modelo	ARIMA1	ARIMA2	LSTM	MLP	SVR	SVR-LSTM
ARIMA2	-0.2801					
	0.3897					
LSTM	1.9063	2.1865				
	0.0283°	0.0144°				
MLP	1.4778	1.7580	-0.4285			
	0.0697°	0.0394°	0.3341			
SVR	-0.5383	-0.2582	-2.4447	-2.0162		
	0.2952	0.3981	0.0072°	0.0219°		
SVR-LSTM	0.8680	1.1482	-1.0383	-0.6098	1.4064	
	0.1927	0.1254	0.1496	0.2710	0.0798°	
SVR-MLP	1.4119	1.6920	-0.4944	-0.0659°	1.9503	0.5438
	0.0790°	0.0453°	0.3105	0.4737	0.0256°	0.2933

° apresentou diferença estatisticamente significante ao nível de 10%.

Fonte: Próprio autor.

CONCLUSÃO

Considerando o crescimento da população mundial, segundo Ripple et al. (2019), o que contribui para a necessidade crescente de água potável disponível para consumo humano, torna-se cada vez mais importante o adequado gerenciamento dos recursos hídricos disponíveis, com otimização dos processos e redução de custos. Por isso, torna-se também imperativo aprimorar a precisão na previsão da demanda de água, sendo que a utilização de modelos de Aprendizagem de Máquina, como ANNs e suas variações, podem contribuir de forma decisiva neste processo.

O objetivo deste trabalho foi verificar a viabilidade de um novo modelo híbrido SVR-ANN utilizado para previsão da demanda de água dos reservatórios que abastecem a RMS. O modelo proposto foi implementado utilizando uma adaptação da metodologia proposta por Zhang (2003) para o modelo aditivo representado pela Equação (2.6), com dados históricos de consumo e dados meteorológicos. O modelo se demonstrou viável mas não unânime para utilização em todos os reservatórios. Além disso, foi verificado que a inclusão de variáveis de consumo com N defasagens de tempo também melhorou os resultados da previsão.

Foram implementados os modelos MLP, SVR e ARIMA, que mostraram resultados promissores na previsão de demanda de água, atingindo erro percentual médio absoluto (MAPE) em torno de 3%. Além disso, outros modelos de aprendizado de máquina mais complexos também foram implementados, como os modelos SVR-LSTM e SVR-MLP, que apresentaram estatísticas MAPE competitivas na previsão horária das séries temporais dos reservatórios R21, R23B e R20, atingindo os percentuais de 2,27%, 2,04% e 1,82% para o SVR-LSTM, e 1,33%, 0,95% e 1,54% para SVR-MLP, respectivamente, demonstrando que o modelo híbrido SVR-ANN, seja SVR-LSTM ou SVR-MLP, usando a metodologia proposta por Zhang (2003), pode ser uma solução viável para predição de séries temporais para demanda de água.

Dentre os modelos tradicionais implementados, o LSTM foi um dos que apresentou melhor desempenho na previsão horária, enquanto os modelos ARIMA se destacaram nas previsões diárias e semanais, porém, não houve unanimidade em relação a um modelo específico, o que sugere que são as características individuais de cada série temporal que determinam o melhor modelo para sua previsão. Dentre os modelos escolhidos para previsão (diária, horária e semanal) de cada reservatório, o modelo LSTM foi o que apresentou a melhor estatística MAPE, com 0,9% para o reservatório R23B na previsão horária, e o modelo SVR foi o que registrou a pior estatística com 22,1% na previsão semanal também para o reservatório R23B.

Os resultados deste trabalho de pesquisa foram publicados no artigo científico dos autores Jesus e Gomes (2023), na revista *Neural Computing and Applications*, da Editora Springer, sob o título *Machine learning models for forecasting water demand for the*

Metropolitan Region of Salvador, Bahia.

Todos os códigos-fonte implementados para construção, teste e validação dos modelos, bem como as tabelas de resultados gerados pelos modelos e as tabelas de consolidação das estatísticas, estão disponíveis no repositório GitHub(<<https://github.com/edmilsondejesus/waterdemand/>>) para permitir a verificação e replicabilidade da pesquisa. Para trabalhos futuros, sugere-se a implementação das variáveis tipo de dia e estação do ano para verificar o impacto da sazonalidade nas previsões. O estudo para verificação da dependência georreferenciada entre os reservatórios. Testar o modelo SVR com outras configurações de kernel como *sigmoid* e *precomputed*. Além da implementação de outros modelos híbridos, como o ARIMA-LSTM; e a inclusão de atrasos também para variáveis meteorológicas.

Foram encontrados poucos trabalhos de pesquisa que tratam sobre a utilização de modelos de aprendizagem de máquina híbridos especificamente voltados para previsão da demanda de água e nenhum deles com referência à aplicação para a RMS. Assim, outra contribuição importante deste trabalho foi a implementação de modelos híbridos, como o SVR-MLP e SVR-LSTM para previsão da demanda de água da RMS.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMOWSKI, J.; KARAPATAKI, C. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different learning algorithms. *J. Hydrol. Eng.*, v. 15, p. 729–743, 2010.
- ALVISI, S.; FRANCHINI, M.; MARINELLI, A. A short-term, pattern-based model for water-demand forecasting. *Journal of Hydroinformatics*, v. 9, n. 1, p. 39–50, 01 2007. ISSN 1464-7141. Disponível em: <<https://doi.org/10.2166/hydro.2006.016>>.
- BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid arima-ann model for forecasting time series data. *Applied Soft Computing*, v. 23, p. 27–38, 2014. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494614002555>>.
- BAKKER, M. et al. Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering*, v. 70, p. 93–102, 2014. ISSN 1877-7058. 12th International Conference on Computing and Control for the Water Industry, CCWI2013. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877705814000149>>.
- BOUGADIS, J.; ADAMOWSKI, K.; DIDUCH, R. Short-term municipal water demand forecasting. *Hydrological Processes*, v. 19, n. 1, p. 137–148, 2005. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.5763>>.
- BOX, G.; JENKINS, G. *Time Series Analysis, Forecasting and Control*. [S.l.: s.n.], 1970.
- CANDELIERI, A. Clustering and support vector regression for water demand forecasting and anomaly detection. *Water*, v. 9, n. 3, 2017. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/9/3/224>>.
- CORTES, C.; VAPNIK, V. Support-vector networks. In: *Machine Learning*. [S.l.: s.n.], 1995. v. 20, n. 3, p. 273–297.
- DONKOR, E. A. et al. Urban water demand forecasting: Review of methods and models. *Journal of Water Resources Planning & Management*, v. 140, n. 2, p. 146 – 159, 2014. ISSN 07339496. Disponível em: <<https://search-ebshost-com.ez10.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=93677319&lang=pt-br&site=ehost-live>>.
- DRUCKER, H. et al. Support vector regression machines. In: MOZER, M.; JORDAN, M.; PETSCHKE, T. (Ed.). *Advances in Neural Information Processing Systems*. MIT Press, 1996. v. 9. Disponível em: <<https://proceedings.neurips.cc/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf>>.

DUNN, O. J. Multiple comparisons using rank sums. *Technometrics*, Taylor & Francis, v. 6, n. 3, p. 241–252, 1964. Disponível em: <<https://doi.org/10.1080/00401706.1964.10490181>>.

FARIAS, R. L. et al. Multi-model prediction for demand forecast in water distribution networks. *Energies*, v. 11, n. 3, 2018. ISSN 1996-1073. Disponível em: <<https://www.mdpi.com/1996-1073/11/3/660>>.

GALÁN, J. M.; LÓPEZ-PAREDES, A.; OLMO, R. D. An agent-based model for domestic water management in valladolid metropolitan area. *Water resources research*, Wiley Online Library, v. 45, n. 5, 2009.

GHIASSI, M.; SAIDANE, H. A dynamic architecture for artificial neural networks. *Neurocomputing*, v. 63, p. 397–413, 2005. ISSN 0925-2312. New Aspects in Neurocomputing: 11th European Symposium on Artificial Neural Networks. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231204003327>>.

GHIASSI, M.; ZIMBRA, D. K.; SAIDANE, H. Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning & Management*, v. 134, n. 2, p. 138 – 146, 2008. ISSN 07339496. Disponível em: <<https://search-ebSCOhost-com.ez10.periodicos.capes.gov.br/login.aspx?direct=true&db=aph&AN=29978636&lang=pt-br&site=ehost-live>>.

HAYKIN, S. *Redes Neurais: Princípios e prática*. [S.l.]: Bookman, 2001.

HERRERA, M. et al. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, v. 387, n. 1, p. 141–150, 2010. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169410001861>>.

HSU, K.-L.; GUPTA, H. V.; SOROOSHIAN, S. *A SUPERIOR TRAINING STRATEGY FOR THREE-LAYER FEEDFORWARD ARTIFICIAL NEURAL NETWORKS*. Department of Hydrology and Water Resources, University of Arizona (Tucson, AZ), 1996. Disponível em: <<http://hdl.handle.net/10150/614171>>.

HUANG, J.; LOU, I.; LI, Y. Integrated water resources management for emergency situations: A case study of macau. *Journal of Environmental Sciences*, v. 50, p. 72–78, 2016. ISSN 1001-0742. Sustainable Water Environment and Water Use. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1001074216301991>>.

IBGE, I. B. D. G. E. E. Estimativas da população residente no brasil e unidades da federação com data de referência em 1º de julho de 2021. 2021. Acessado em 04/08/2022. Disponível em: <https://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2021/estimativa_dou_2021.pdf>.

IPEA, I. de P. E. A. Governança metropolitana no brasil - relatório de pesquisa. 2015. Acessado em 03/08/2022. Disponível em: <https://www.ipea.gov.br/redeipea/images/pdfs/governanca_metropolitana/relatorio_1.1_revisao_final_salvador.pdf>.

JAIN, A.; KUMAR, A. M. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, v. 7, n. 2, p. 585–592, 2007. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494606000317>>.

JESUS, E. d. Santos de; GOMES, G. S. d. S. Machine learning models for forecasting water demand for the metropolitan region of salvador, bahia. *Neural Computing and Applications*, 2023. ISSN 1433-3058.

KHASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and arima models for time series forecasting. *Applied Soft Computing*, v. 11, n. 2, p. 2664–2675, 2011. ISSN 1568-4946. The Impact of Soft Computing for the Progress of Artificial Intelligence. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1568494610002759>>.

KOUTIVA, I.; MAKROPOULOS, C. Modelling domestic water demand: An agent based approach. *Environmental Modelling & Software*, v. 79, p. 35–54, 2016. ISSN 1364-8152. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1364815216300056>>.

KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, Taylor & Francis, v. 47, n. 260, p. 583–621, 1952. Disponível em: <<https://doi.org/10.1080/01621459.1952.10483441>>.

LIU, W.; LIU, L.; TONG, F. Least squares support vector machine for ranking solutions of multi-objective water resources allocation optimization models. *Water*, v. 9, n. 4, 2017. ISSN 2073-4441. Disponível em: <<https://www.mdpi.com/2073-4441/9/4/257>>.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, SciELO Brasil, v. 35, p. 85–94, 2021.

ODAN, F. K.; REIS, L. F. R. Hybrid water demand forecasting model associating artificial neural network with fourier series. *Journal of Water Resources Planning and Management*, v. 138, n. 3, p. 245–256, 2012. Disponível em: <<https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0000177>>.

Oliveira, J. F. de; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, v. 180, p. 27–34, 2016. ISSN 0925-2312. Progress in Intelligent Systems Design. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231215016057>>.

REES, P.; CLARK, S.; NAWAZ, R. Household forecasts for the planning of long-term domestic water demand: Application to london and the thames valley. Wiley Online Library, 2020. Disponível em: <<https://doi-org.ez10.periodicos.capes.gov.br/10.1002/psp.2288>>.

RIBEIRO, W. C. *Geografia política da água*. [S.l.]: Annablume Editora, 2008.

RIPPLE, W. J. et al. World Scientists' Warning of a Climate Emergency. *BioScience*, v. 70, n. 1, p. 8–12, 11 2019. ISSN 0006-3568. Disponível em: <<https://doi.org/10.1093/biosci/biz088>>.

SANTOS, C. C. dos; FILHO, A. J. P. Water demand forecasting model for the metropolitan area of são paulo, brazil. 2014.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, [Oxford University Press, Biometrika Trust], v. 52, n. 3/4, p. 591–611, 1965. ISSN 00063444. Disponível em: <<http://www.jstor.org/stable/2333709>>.

SHARVELLE, S. et al. A geospatially-enabled web tool for urban water demand forecasting and assessment of alternative urban water management strategies. *Environmental Modelling & Software*, v. 97, p. 213–228, 2017. ISSN 1364-8152. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1364815217300828>>.

SILVA, W. T. P. da et al. Quota per capita de água, fatores intervenientes e modelagem: estudo de caso para classes socioeconômicas de cuiabá-mt. *Sociedade & Natureza [online]*, v. 20, n. 2, p. 219–230, 2008. ISSN 1982-4513.

SMOLA, A.; SCHÖLKOPF, B. *Learning with Kernels*. [S.l.]: MIT Press, 2002.

SMOLA, A.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*, v. 14, p. 199–222, 08 2004.

VANBERLO, B.; ROSS, M. A. S.; HSIA, D. *Univariate Long-Term Municipal Water Demand Forecasting*. 2021.

VERDÉRIO, A. Sobre o uso de regressão por vetores suporte para a construção de modelos em um método de região de confiança sem derivadas. Programa de Pós-Graduação em Matemática, Universidade Federal do Paraná, 2015.

WANG, L. et al. An arima-ann hybrid model for time series forecasting. *Systems Research and Behavioral Science*, v. 30, n. 3, p. 244–259, 2013. Disponível em: <<https://doi.org/10.1002/sres.2179>>.

WU, B. An introduction to neural networks and their applications in manufacturing. *Journal of Intelligent Manufacturing*, Springer, v. 3, n. 6, p. 391–403, 1992.

Yan, Kun; Yang, Min-Zhi. Water demand forecast model of least squares support vector machine based on particle swarm optimization. *MATEC Web Conf.*, v. 246, p. 01029, 2018. Disponível em: <<https://doi.org/10.1051/mateconf/201824601029>>.

YASSUDA, E. et al. Técnica de abastecimento e tratamento de água. CETESB, v. 1, 1976.

ZHANG, G. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, v. 50, p. 159–175, 2003. ISSN 0925-2312. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0925231201007020>>.

Apêndice

A

TABELA DE TRABALHOS RELACIONADOS

Tabela A.1 Trabalhos relacionados por métrica de avaliação e técnica utilizada.

Nº	Título	Autores	Métricas	Tipo	Técnica(s)
1	Previsão univariada da demanda de água municipal de longo prazo	VanBerlo, Ross e Hsia (2021).	MAPE, MSE e RMSE	Longo prazo (incluiu dias, semanas, meses e anos).	OLS, 1D-CNN, LSTM e o modelo <i>Prophet</i> do Facebook.
2	Previsões familiares para planejamento da demanda doméstica de água a longo prazo: aplicado a Londres e ao Vale do Tâmsa.	Rees, Clark e Nawaz (2020).	Não identificado.	Longo prazo (anos).	Modelo heurístico projetado para previsão do consumo de água doméstica.
3	Modelo de previsão de demanda de água com LSSVM, baseada em PSO.	Yan, Kun e Yang, Min-Zhi (2018).	MAPE.	Longo prazo (anos).	PSO e LSSVM.
4	Ferramenta web geoespacial para previsão de demanda de água e avaliação de estratégias alternativas de gestão de água urbana.	Sharvelle et al. (2017).	MRE e NSCE.	Médio e longo prazo (meses e anos).	IUWM.
5	Clusterização e SVRs para previsão de demanda de água e detecção de anomalias.	Candelieri (2017).	MAPE.	Curto prazo (horas).	Clusterização e SVRs.
6	Modelo de previsão de demanda de água para a Região Metropolitana de São Paulo, Brasil.	Santos e Filho (2014).	ME, MAE, RMSE, R e R^2 .	Curto prazo (horas, dias e semanas).	MLR, ANN-MLP e LLSSIM.
7	Melhorando o desempenho de modelos de previsão de demanda de água usando entrada de clima.	Bakker et al. (2014).	Erro Relativo (RE), MAPE e R^2 .	Curto prazo (dias).	Modelo heurístico adaptativo, ARIMA e MLR.
8	Modelo baseado em agentes para a gestão da água doméstica na área metropolitana de Valladolid	Galán, López-Paredes e Olmo (2009).	Não identificado.	Médio e longo prazo (meses e anos).	Modelo de agentes baseados em comportamento.
9	Previsão de demanda de água urbana com uma DAN2	Ghiassi e Saidane (2005)	MAPE.	Médio prazo (hora, dia, semana e mês).	DAN2, ARIMA, ANN e MLP.
10	Um modelo de curto prazo baseado em padrões para previsão da demanda de água.	Alvisi, Franchini e Marinelli (2007).	RMSE, MAE e EV.	Curto prazo (horas e dias).	Modelo modular baseado em padrões.
11	Previsão de demanda de água municipal de curto prazo.	Bougadis, Adamowski e Diduch (2005).	R^2 , AARE e Max ARE.	Curto prazo (dias e semanas)	ANN, ARIMA, LR, MLR e análise de <i>Fourier</i> .