

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO**

Hiago Martins Borges

**Modelo de previsão de consumo de água combinado com
estratégia de otimização para redução de custos
operacionais em sistemas de distribuição de água.**

São Carlos

2023

Hiago Martins Borges

**Modelo de previsão de consumo de água combinado com
estratégia de otimização para redução de custos
operacionais em sistemas de distribuição de água.**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Afonso Paiva Neto

São Carlos

2023

Hiago Martins Borges

**Modelo de previsão de consumo de água combinado com
estratégia de otimização para redução de custos
operacionais em sistemas de distribuição de água.**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados.

Data de defesa: 18 de março de 2023

Comissão Julgadora:

Prof. Dr. Afonso Paiva Neto
Orientador

Professor
Convidado1

Professor
Convidado2

São Carlos
2023

RESUMO

BORGES, H. M. **Modelo de previsão de consumo de água combinado com estratégia de otimização para redução de custos operacionais em sistemas de distribuição de água.** 2023. 67p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

A crescente demanda por água, os custos operacionais elevados e os desafios impostos pelos cenários econômico-financeiro e ambiental têm levado empresas de saneamento a buscar soluções para melhorar a eficiência e reduzir os custos operacionais. Nesse contexto, o desenvolvimento de modelos para previsão do consumo de água e sua integração em estratégias para tomada de decisão de acionamento de bombas destacam-se como possíveis estratégias para a melhoria operacional. Nesse sentido, este estudo explorou soluções para desenvolvimento de modelos preditivos de consumo de água a curto prazo e otimização do abastecimento de água em sistemas de distribuição, considerando a minimização do consumo de energia elétrica. Foi estudado um sistema de abastecimento operado pela Saneamento de Goiás, cujos dados operacionais foram obtidos por meio do ambiente supervisorio. Após análise exploratória e seleção do modelo preditivo mais adequado com a biblioteca PyCaret do Python, a metodologia proposta foi implementada com o uso da biblioteca Pyomo para otimização do acionamento de bombas. Os resultados mostraram que a estratégia operacional automatizada apresenta vantagens em relação à estratégia atualmente em uso, com reduções significativas de consumo de energia elétrica em até 18%. Verificou-se que o desempenho do modelo preditivo foi adequado para obtenção de resultados satisfatórios. Finalmente, a metodologia proposta neste trabalho se mostrou eficaz na previsão do consumo de água e na otimização do consumo energético e, conseqüentemente, do custo de operação de sistemas de distribuição de água, mostrando-se promissora para a incorporação nas estratégias de automação atualmente existentes nos sistemas operados pela Saneago.

Palavras-chave: saneamento, modelos preditivos, consumo de água, otimização, custos energéticos

ABSTRACT

BORGES, H. M. **Water consumption forecasting model combined with optimization strategy for operational cost reduction in water distribution systems.** 2023. 67p. Monografia (MBA em Ciências de Dados) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

The increasing demand for water, high operational costs, and challenges posed by economic, financial, and environmental scenarios have led sanitation companies to seek solutions to improve efficiency and reduce operational costs. In this context, the development of water consumption forecasting models and their integration into strategies for optimal pump scheduling are recognized as potential strategies for operational improvement. Thus, this study explored solutions for developing short-term water consumption forecasts models and optimizing pump scheduling in a water distribution system, considering the minimization of total electricity costs. A water supply system operated by Saneamento de Goiás was studied, and its operational data were obtained through the supervisory environment. After exploratory analysis and selection of the most appropriate predictive model using the PyCaret library in Python, the proposed optimization methodology was implemented using the Pyomo library. The results showed that the automated operational strategy has advantages over the currently used strategy, with significant reductions in electricity costs of up to 18%. It was found that the predictive model's performance was adequate for achieving satisfactory results. Finally, the proposed methodology in this study was effective in predicting water consumption and optimizing energy consumption and, consequently, the operating costs of water distribution systems, showing promise for incorporation into existing automation strategies in the systems operated by Saneago.

Keywords: sanitation, forecast models, water consumption, energy cost optimization

LISTA DE FIGURAS

Figura 1 – Dados de consumo de água em determinado sistema de abastecimento.	27
Figura 2 – Componentes de uma série temporal hipotética.	28
Figura 3 – Decomposição clássica de uma série temporal aditiva.	30
Figura 4 – Valores de correlação computados para cada <i>lag k</i>	31
Figura 5 – Correlograma de uma série temporal com tendência e sazonalidade.	31
Figura 6 – Etapas propostas para o desenvolvimento, análise e aplicação do modelo.	33
Figura 7 – Dados coletados do ambiente Supervisório <i>Web</i>	34
Figura 8 – Validação cruzada de série temporal.	38
Figura 9 – Abordagem de validação cruzada empregada.	39
Figura 10 – Abordagem empregada para treinamento e predição.	41
Figura 11 – Sistema de Abastecimento de Água estudado.	42
Figura 12 – Tarifa Branca - Postos Tarifários no Estado de Goiás para dias úteis.	43
Figura 13 – Abordagem proposta para operação otimizada das bombas.	44
Figura 14 – Série temporal de consumo de água.	49
Figura 15 – Padrões anuais para o consumo de água.	50
Figura 16 – Variações mensais de consumo horário do sistema estudado.	50
Figura 17 – Padrões semanais para o consumo de água.	51
Figura 18 – Variações do consumo horário ao longo da semana.	51
Figura 19 – Padrões diários para o consumo de água.	52
Figura 20 – Variações do consumo ao longo do dia.	52
Figura 21 – Correlograma dos dados de consumo de água.	53
Figura 22 – Resultado das predições para o conjunto de teste.	55
Figura 23 – Predições finais para o conjunto de teste.	57
Figura 24 – Resultado da aplicação da estratégia de otimização (cenário 2) aos dados observados em janeiro de 2023.	59
Figura 25 – Resultado da aplicação da estratégia de otimização (cenário 2) para os dias 15 e 17 de janeiro.	60
Figura 26 – Comparação entre os custos finais obtidos em cada cenário.	62

LISTA DE TABELAS

Tabela 1 – Relação entre o nível de planejamento, problemas de decisão da concessionária de água e atributos de previsão.	19
Tabela 2 – Vantagens e desvantagens das principais categorias de métodos preditivos para previsão de demanda de água.	22
Tabela 3 – Métodos abordados nos trabalhos analisados para previsão de demanda de água de curto prazo.	26
Tabela 4 – Vazão e energia consumida por cada estado.	45
Tabela 5 – Desempenho dos modelos treinados com o emprego da biblioteca PyCaret.	54
Tabela 6 – Métricas de desempenho apuradas para as previsões realizadas.	56
Tabela 7 – Métricas de desempenho para as previsões finais.	56
Tabela 8 – Violações aos limites de reservação máximo e mínimo admissíveis.	60
Tabela 9 – Tempo de funcionamento das bombas, por estado, e consumo energético total correspondente.	61

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	16
1.2	Objetivos	16
1.3	Estrutura do Trabalho	17
2	REVISÃO BIBLIOGRÁFICA	19
3	CONCEITOS BÁSICOS	27
3.1	Séries Temporais	27
4	METODOLOGIA	33
4.1	Coleta de Dados	34
4.2	Pré-processamento	34
4.2.1	Detecção e remoção de <i>outliers</i>	35
4.3	Análise Exploratória e Descritiva	36
4.4	Modelagem	36
4.4.1	Escolha do Melhor Modelo	39
4.4.2	Modelo de Referência	40
4.4.3	Teste do Modelo	40
4.4.4	Análise do Resultado	41
4.5	Aplicação prática do modelo preditivo	41
4.5.1	Definição de Parâmetros	44
4.5.2	Otimização	45
4.5.3	Implementação	47
5	RESULTADOS	49
5.1	Coleta e Pré-processamento	49
5.2	Análise Exploratória e Descritiva	49
5.3	Modelagem	53
5.3.1	Escolha do melhor modelo	54
5.3.2	Teste e Avaliação do Modelo	55
5.4	Estratégia de operação otimizada	57
6	CONCLUSÃO	63
	REFERÊNCIAS	65

1 INTRODUÇÃO

O cenário que se apresenta para as empresas de abastecimento público de água é desafiador. Questões climáticas e ambientais se somarão a novos desafios econômico-financeiros, tornando crucial a gestão otimizada de recursos naturais e financeiros.

Do ponto de vista ambiental, eventos de seca severa poderão se tornar mais frequentes e intensos nas regiões Nordeste, Centro-Oeste e Leste do Norte do Brasil (LEE *et al.*, 2021). A crise hídrica e energética que antes era pontual e limitada a algumas regiões poderá atingir proporções mais extremas e se tornar uma realidade recorrente nos próximos anos, impondo desafios ainda maiores para os setores de infraestrutura.

No âmbito financeiro, a demanda por energia elétrica para operar os sistemas de abastecimento de água corresponde atualmente a um dos principais custos para as empresas de saneamento no Brasil e no mundo. Segundo o diagnóstico divulgado pelo Sistema Nacional de Informações sobre o Saneamento - SNIS, em 2019, o consumo de energia elétrica representou quase 16% das despesas de exploração das companhias de saneamento no Brasil, sendo este gasto superado somente pelo custo com pessoal (BRASIL, 2020).

Além disso, as empresas de saneamento, especialmente as públicas, enfrentarão novos desafios de sobrevivência de mercado, pois a entrada em vigor do Novo Marco do Saneamento Básico (Lei Nº 14.026, de 15 de julho de 2020) abre espaço para a maior participação de empresas privadas na prestação de serviços de abastecimento público de água, aumentando os desafios neste novo cenário de concorrência pela conquista e manutenção de mercados.

Diante desse cenário, empresas do setor do saneamento que usualmente se limitavam a soluções tradicionais do mercado passaram a dar mais importância aos aspectos de inovação, corroborando com a transformação digital do setor. Sistemas de Supervisão, Controle, Aquisição de Dados e Automação (SCADA) já fazem parte da rotina operacional de companhias de saneamento em todo o país, abrindo caminho para o próximo passo: gerar valor a partir dos dados disponíveis e utilizá-los como suporte na tomada de decisão.

Portanto, com o aumento da quantidade de dados coletados nos sistemas de abastecimento de água e a ampla difusão de técnicas para a geração de valor a partir desses dados, as empresas de saneamento estão percebendo a oportunidade de adotar conhecimentos da área de Ciência de Dados para buscar ferramentas capazes de aumentar consideravelmente sua eficiência operacional.

1.1 Justificativa

A busca por ferramentas que viabilizem a melhoria dos processos e redução de custos representa um desafio para qualquer empresa. Nesse sentido, a Saneago, companhia de Saneamento do Estado de Goiás, tem implementado diversas ações de melhoria operacional, dentre elas se destaca a implementação do Sistema Supervisório.

Atualmente, a companhia conta com o Sistema Supervisório em todos os 226 municípios em que opera. Este sistema permite o acompanhamento em tempo real do nível dos reservatórios, do *status* das bombas (ligadas ou desligadas), das vazões, pressões, entre outras informações importantes para acompanhar a produção e distribuição de água. Além de permitir o monitoramento em tempo real, os dados são coletados e armazenados pelo sistema a cada 15 minutos, originando séries temporais que podem ser utilizadas como fonte importante para estudos e análises diversas.

Com o grande número de sistemas operados, incluindo alguns sistemas deficitários, e o aumento significativo na quantidade de dados coletados, é possível empregar técnicas de mineração de dados aliadas a técnicas de otimização para geração de valor a partir dos dados existentes. Apesar de contar com solução própria que permite o acompanhamento em tempo real e operação automatizada dos sistemas de distribuição de água, a Saneago ainda não adota modelos de previsão de consumo de água como ferramenta de suporte à tomada de decisão.

Nesse contexto, propõe-se estudar técnicas comumente empregadas para a previsão da demanda urbana de água e desenvolver um modelo para determinado sistema operado pela Saneamento de Goiás S.A. Dentre as possíveis aplicações deste tipo de modelo, destaca-se a sua integração ao sistema de operação automatizada, o que possibilitaria o acionamento inteligente de bombas e, conseqüentemente, a redução no gasto com energia, sem prejuízo ao atendimento da demanda de água requerida pelos consumidores. A abordagem proposta é alinhada com a cultura da empresa de desenvolver soluções internas eficientes e aplicáveis em larga escala, não exigindo investimentos consideráveis, baseando-se exclusivamente na geração de valor a partir dos dados existentes.

1.2 Objetivos

Considerando que diversas técnicas têm sido utilizadas para a previsão da demanda de água em sistemas de abastecimento, o objetivo geral deste trabalho é desenvolver um modelo de previsão de consumo de água para um determinado sistema operado pela Saneamento de Goiás S.A e aplicá-lo em um caso prático de otimização do regime de funcionamento de bombas.

Para alcançar este objetivo, são propostos os seguintes objetivos específicos:

- a) Levantar os modelos comumente empregados para previsão de demanda urbana

de água em curto prazo;

b) Realizar análise exploratória e descritiva dos dados de consumo de água em um sistema de abastecimento público;

c) Desenvolver modelos de previsão de demanda de água usando diferentes métodos empregados para a predição de séries temporais;

d) Propor uma metodologia para otimizar o regime de funcionamento de bombas, minimizando o custo de energia total para o bombeamento;

e) Aplicar a metodologia de otimização empregando o modelo preditivo de consumo desenvolvido e avaliar os resultados potenciais da implementação da estratégia proposta.

1.3 Estrutura do Trabalho

Este trabalho está organizado da seguinte forma:

No Capítulo 1, são apresentados a introdução, a contextualização e os objetivos do trabalho. Em seguida, no Capítulo 2, é realizada a revisão bibliográfica, onde são apresentados os conceitos básicos relacionados aos problemas de previsão de demanda de água em curto prazo e os trabalhos relacionados ao tema. O Capítulo 3 apresenta os conceitos fundamentais relacionados ao estudo e análise de séries temporais. Posteriormente, o Capítulo 4 expõe a metodologia proposta para o desenvolvimento do modelo preditivo, bem como a estratégia para integrá-lo na automação otimizada do funcionamento das bombas. No Capítulo 5 são apresentados os resultados encontrados. Finalmente, o Capítulo 6 apresenta as conclusões e recomendações para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

No cenário atual de desafios econômicos e de incertezas no âmbito climático, as concessionárias de água ao redor do mundo enfrentam o desafio de abastecer a crescente população urbana com água em quantidade e qualidade adequadas. A gestão sustentável da demanda de água tem se tornado uma necessidade latente, e as companhias de saneamento estão cada vez mais buscando técnicas para auxílio na tomada de decisão a fim de garantirem o atendimento da população atendida. Dentre as técnicas empregadas, destaca-se o uso de modelos de previsão de demanda de água (GROPPO; COSTA; LIBÂNIO, 2019; NIKNAM et al., 2022).

O estudo de modelos de previsão de demanda de água em sistemas de abastecimento público não se trata de algo novo. Estes modelos, quando confiáveis, auxiliam as companhias de abastecimento na tomada de decisões em nível estratégico, tático e operacional (Tabela 1). Assim, decisões relacionadas à expansão, operação, gerenciamento e otimização dos sistemas públicos de abastecimento de água podem ser tomadas com base em diversos tipos de modelos preditivos, podendo estes terem periodicidade de predição horária, diária, mensal, etc., de acordo com o objetivo para o qual foram desenvolvidos (DONKOR et al., 2014).

Tabela 1 – Relação entre o nível de planejamento, problemas de decisão da concessionária de água e atributos de previsão.

Nível de planejamento	Tipo de problema	Horizonte de previsão	Periodicidade do modelo
Operacional	Gerenciamento, operação e otimização do sistema de abastecimento	Curto prazo (Menor que 1 ano)	Horário
			Diário
			Semanal
			Mensal
Tático	Previsão de receitas e Planejamento de investimentos para melhorias do sistema de abastecimento	Médio prazo (1 – 10 anos)	Mensal
			Anual
Estratégico	Expansão do sistema de abastecimento	Longo prazo (maior que 10 anos)	Anual

Fonte: Adaptado de Donkor et al. (2014).

O consumo de água pela população é influenciado por diversos fatores que podem ou não serem considerados na construção do modelo de previsão de demanda. De acordo com Donkor et al. (2014), os modelos podem incluir, por exemplo, fatores relacionados às características socioeconômicas, bem como às características climáticas da região abastecida. Bich-Ngoc e Teller (2018), categorizam os fatores que influenciam diretamente o consumo residencial de água nos grupos a seguir:

- Fatores econômicos: valor da tarifa de água, renda familiar, etc.;
- Fatores sociodemográficos: número de pessoas no domicílio, idade dos ocupantes, valor do imóvel, etc.;
- Fatores físicos: tipo e tamanho do domicílio, existência de jardim, etc.;
- Fatores climáticos ou meteorológicos: temperatura, umidade, precipitação, etc.;
- Fatores geográficos: densidade urbana, localização, etc.;
- Fatores tecnológicos: precisão dos medidores, existência de medidores inteligentes, sensores, coletores de dados, etc..

De acordo com (NIKNAM et al., 2022), além dos fatores elencados, pode-se também considerar outro grupo denominado *calendar variables*, que reflete as características temporais (dias úteis, fins de semana, feriados, eventos especiais, etc.).

Bich-Ngoc e Teller (2018) e Niknam et al. (2022) destacam que todos os possíveis fatores mencionados podem ser considerados como parâmetros no desenvolvimento de modelos preditivos. Portanto, quanto maior a quantidade de fatores incluídos em um modelo de previsão de demanda, mais complexo ele se torna, o que pode inviabilizar não só a obtenção dos dados necessários, mas também a sua aplicação prática no dia-a-dia das empresas de saneamento.

Embora Donkor et al. (2014) destaque que o estudo de modelos de previsão de demanda de água em sistemas de abastecimento público não se trata de algo novo, a digitalização do setor do saneamento, o crescente número de dados e informações disponíveis e a disseminação de técnicas de mineração de dados tem contribuído para um aumento considerável no interesse pelo tema. Conforme Groppo, Costa e Libânio (2019), o número de artigos publicados nesta área aumentou exponencialmente nos últimos 20 anos, podendo ser reflexo também da crescente escassez de recursos hídricos.

Existe uma quantidade considerável de métodos e técnicas empregadas para o desenvolvimento de modelos de previsão de demanda de água. Diversos autores se propuseram a realizar uma extensiva revisão bibliográfica dos trabalhos publicados na área de forma a fornecer uma orientação abrangente para os profissionais e companhias de saneamento na seleção de um método de previsão a ser empregado dentre a infinidade de opções disponíveis. Diante disso, se atendo aos trabalhos mais recentes nesta área, destacam-se as revisões bibliográficas apresentadas por Groppo, Costa e Libânio (2019), Muhammad, Li e Feng (2019), e recentemente por Niknam et al. (2022).

Muhammad, Li e Feng (2019) analisaram mais de cem artigos de maior impacto à época do estudo. A revisão realizada forneceu uma revisão básica sobre as tendências de pesquisa relacionadas à previsão da demanda urbana de água usando várias abordagens,

desde o emprego de modelos ARIMA até o uso de redes neurais artificiais e modelos híbridos. De acordo com levantamento realizado pelos autores, em relação às variáveis de entradas mais comuns nos modelos analisados destacam-se, respectivamente, dados históricos de demanda de água, temperatura e precipitação. Em relação à escala temporal, os autores relatam que a mais comum dentre os trabalhos revisados é a periodicidade diária, seguida da mensal e horária, predominando, portanto, modelos de curto prazo, conforme apresentado na Tabela 1.

Grosso, Costa e Libânio (2019) revisaram diferentes artigos que abordam a previsão da demanda urbana de água. A revisão abrange métodos e modelos desenvolvidos usando desde técnicas estatísticas, como regressão linear e análise de séries temporais, até técnicas mais sofisticadas e complexas com o emprego de inteligência artificial. A revisão mostra que os estudos analisados são, em sua maioria, focados no gerenciamento e operação dos sistemas de água, concentrando-se, assim, em modelos preditivos de curto prazo, corroborando com o exposto por Muhammad, Li e Feng (2019).

Com foco em modelos preditivos de curto prazo, Niknam et al. (2022) fizeram uma extensiva revisão contemplando mais de cem trabalhos publicados nos últimos dez anos, classificando-os de acordo com os métodos que foram empregados. Os resultados mostram que técnicas tradicionais de análise de séries temporais e redes neurais artificiais estão entre os métodos mais empregados na literatura, usados em 25% e 20% dos artigos revisados, respectivamente.

Além de apresentarem revisões extensivas e compreensivas de trabalhos relacionados a previsão de demanda urbana de água, Muhammad, Li e Feng (2019), Grosso, Costa e Libânio (2019), e Niknam et al. (2022) também buscam elucidar a escolha de qual método empregar dentre a infinidade de opções disponíveis. Nesse sentido, os autores concluem que não existe um método universal que se sobressaia em todas as situações de aplicação e que a escolha do método a ser empregado deve basear-se nos dados disponíveis para incorporarem o modelo, bem como no objetivo de aplicação e alcance pretendido.

Adicionalmente, Niknam et al. (2022) apresenta alguns critérios a serem avaliados para auxiliar na escolha de um método que seja adequado:

- Identificar o escopo temporal e fatores (exógenos) de influência em cada caso particular;
- Determinar o objetivo da análise: predição da demanda média, predição das demandas máximas, detecção de anomalias, etc;
- Levantar em conta a tecnologia disponível: predições em tempo real, quase-real ou *off-line*, integração aos sistemas de tomada de decisão existentes, etc;

A Tabela 2 apresenta uma comparação apresentada por [Niknam et al. \(2022\)](#) dos principais métodos usados para a previsão de demanda de água em termos de quantidade de dados para implementação, precisão, interpretabilidade do modelo, eficiência computacional de implementação e adaptabilidade às mudanças bruscas na demanda de água.

Tabela 2 – Vantagens e desvantagens das principais categorias de métodos preditivos para previsão de demanda de água.

Método	Quantidade de Dados	Precisão	Interpretabilidade	Eficiência Computacional	Adaptabilidade
Redes Neurais e afins	Alta	Alta	Baixa	Baixa	Média
Regressão por Vetores Suporte e afins	Alta	Alta	Média	Baixa	Média
(S)ARIMA(X)	Baixa	Baixa	Alta	Alta	Baixa
Meta-heurística	Alta	Média	Baixa	Baixa	Baixa
Regressão	Baixa	Média	Alta	Alta	Baixa
Híbridos	Alta	Alta	Média	Média	Alta

Fonte: Adaptado de [Niknam et al. \(2022\)](#).

Dentre os métodos destacados na Tabela 2, [Niknam et al. \(2022\)](#) destacam o grande uso de métodos interpretáveis em modelos preditivos de demanda de água: aproximadamente 30% dos artigos analisados basearam sua análise em séries temporais tradicionais (por exemplo, modelos ARIMA) e em modelos de regressão. Conforme os autores, isso se deve a esses modelos terem uma capacidade preditiva suficiente para auxiliar na tomada de decisões na operação dos sistemas de abastecimento de água, como por exemplo no caso prático de programação do acionamento e desligamento de bombas. Além disso, a interpretabilidade de um modelo favorece naturalmente a sua inclusão nos processos de tomada de decisão.

Tendo em vista as extensivas e compreensivas revisões de literatura realizadas por [Grosso, Costa e Libânio \(2019\)](#), [Muhammad, Li e Feng \(2019\)](#), e [Niknam et al. \(2022\)](#), valeu-se dos trabalhos elencados nestes três estudos para a elaboração de um inventário de artigos que abordam modelos preditivos de demanda de água de curto prazo, com periodicidade horária. Excluindo-se os trabalhos em duplicidade, obteve-se uma lista de 37 trabalhos únicos. Adicionalmente, com o objetivo de limitar a abrangência dos artigos a serem abordados, foi realizado o seguinte recorte:

1. Selecionou-se artigos cujos modelos preditivos requeiram apenas dados de demanda de água, tendo sido eliminados da lista cinco trabalhos que incluíam variáveis exógenas como temperatura, humidade, etc;
2. Limitou-se aos artigos publicados até cinco anos antes ao ano de publicação da revisão bibliográfica em que foram abordados;

3. Para casos em que o mesmo autor estivesse envolvido em mais de um trabalho, considerou-se apenas o artigo mais recente caso abordassem o mesmo tema. Após realizadas as análises elencadas, restaram 14 trabalhos dos 37 pré-selecionados.

A análise em relação aos métodos empregados em cada trabalho mostra a grande incidência de técnicas de análise e modelagem de séries temporais, sendo que cerca de 85% dos trabalhos abordam alguma técnica relacionada como (S)ARIMA(X), Suavização Exponencial, entre outras. Devido a sua interpretabilidade, o emprego de métodos de séries temporais se dá, na maioria dos casos, como modelo base para comparação e avaliação com os métodos propostos por cada estudo. Nesse caso, percebe-se o número expressivo de trabalhos que combinam um ou mais métodos diferentes (64%), propondo uma abordagem híbrida para modelagem da previsão de demanda de água de curto prazo, destacando-se a utilização de modelos de RNAs.

[Bata, Carriveau e Ting \(2020\)](#) propõem um método híbrido combinando técnicas de aprendizado de máquina para previsão da demanda de água de curto prazo. Neste modelo, o método de Mapas Auto-Organizáveis foi utilizado para agrupar os dados de entrada de vazão de água em clusters, enquanto o método de Árvores de Regressão previu a demanda hídrica considerando tais grupos. O modelo híbrido melhora consideravelmente o desempenho dos modelos de Árvore de Regressão e SARIMA, porém, se trata de um modelo mais complexo de ser desenvolvido e interpretado, possuindo menor eficiência computacional.

[Candelieri \(2017\)](#) apresenta uma abordagem híbrida em dois estágios (agrupamento de Séries Temporais e Regressão por Vetores Suporte). Primeiro são construídos agrupamentos de séries temporais que refletem o comportamento de uso de água que se repetem em diferentes escalas de tempo. Em uma segunda etapa, vários modelos de Regressão por Vetores de Suporte são treinados, em particular um para cada cluster e para cada hora de consumo, a fim de aumentar a precisão da previsão. A abordagem melhora a precisão das previsões em relação aos métodos amplamente adotados na literatura baseados em apenas um modelo único de Regressão por Vetores de Suporte, porém requer o treinamento de uma quantidade de modelos consideravelmente maior.

Analogamente, [Rangel et al. \(2017\)](#) propõem a utilização de um banco de modelos ao invés de um modelo único. São desenvolvidos 24 modelos para prever a demanda de cada hora específica do dia. Os modelos horários utilizam Redes Neurais Artificiais e Algoritmos Genéticos que foram utilizados para definir a melhor arquitetura e parâmetros para cada modelo. A abordagem é aplicada para dados de demanda de água na cidade de Barcelona, Espanha, e mostra-se mais precisa que modelos tradicionais utilizando Redes Neurais, Holt-Winters e Modelo Ingênuo.

Assim como [Candelieri \(2017\)](#), [Shabani et al. \(2018\)](#) também propuseram uma

abordagem baseada em um processo de aprendizado de dois estágios incluindo agrupamento de séries temporais. Na abordagem proposta, o agrupamento de séries temporais é usado para organizar padrões de demanda diária de água, enquanto Programação de Expressão Gênica (PEG) é empregada para modelar a demanda de tais agrupamentos. Os resultados indicam que métodos de PEG foram capazes de aumentar consideravelmente a precisão de previsões quando associados a algoritmos de aprendizado não-supervisionados como o k-means.

Gagliardi et al. (2017) apresentam um método de previsão de demanda de água de curto prazo baseado no uso da Cadeia de Markov. O método é capaz de fornecer não apenas as estimativas da demanda, mas também inclui a probabilidade de que a demanda esteja dentro de determinado intervalo de previsão. A abordagem proposta é comparada com outros dois modelos usados como referência (Rede Neural e Modelo Ingênuo), tendo oferecido precisão similar ao modelo que utilizou Rede Neural, e superior ao Modelo Ingênuo.

Salloom, Kaynak e He (2021) usaram a aprendizagem profunda com arquitetura de Redes Neurais denominada Gate Recurrent Unit (GRU) como modelo básico para prever a demanda horária de água em Changzhou, China. O método, em comparação aos outros métodos comumente empregados, reduz a complexidade do modelo consideravelmente, conservando a mesma precisão. Hu et al. (2021) também empregam um modelo GRU para prever a demanda horária de água em áreas de Xangai, China, tendo obtido resultados mais precisos em comparação com a aplicação de Redes Neurais tradicionais e Regressão por Vetores de Suporte.

Propondo avaliar o desempenho de diferentes modelos de previsão de demanda de água usando dados coletados em tempo real a partir de medidores inteligentes, Koo et al. (2021) aplicaram modelo de redes neurais *Long Short Term Memory* (LSTM) e o compararam a outros três modelos: ARIMA, Redes Neurais tradicional e Híbrido. O modelo LSTM mostrou melhor resultado quando comparado aos outros modelos estudados, entretanto, indicam que os resultados obtidos poderiam ser melhorados com a inclusão de variáveis exógenas que influenciam a demanda de água.

Pacchin et al. (2019) comparam seis modelos de previsão de demanda de água de curto prazo baseados em redes neurais, padrões, padrões com janela deslizante, cadeia de Markov e modelo ingênuo. Foi verificado que os modelos mais complexos (redes neurais e cadeia de Markov) forneceram resultados similares e com precisão de previsão média a alta. Por outro lado, os modelos mais simples como os baseados em padrões foram mais robustos e consideravelmente mais fáceis de configurar e parametrizar.

Pandey et al. (2021) apresentam uma abordagem híbrida e concluem que se trata de uma abordagem melhor que outros modelos comumente empregados com base em redes neurais em termos de precisão de previsão, sem comprometer consideravelmente a

eficiência computacional.

[Ponte et al. \(2016\)](#) abordam uma metodologia de modelo de suporte a decisão baseado em Redes Neurais Artificiais em conjunto com outros quatro métodos de previsão (modelo ingênuo, médias móveis, alisamento exponencial e ARIMA). Para o caso em estudo, verificou-se que a introdução de técnicas de inteligência artificial no processo de previsão pode diminuir significativamente o erro quando comparadas com outras técnicas tradicionais de séries temporais.

[Sardinha-Lourenço et al. \(2018\)](#) demonstram uma combinação de um modelo heurístico com um modelo ARIMA e um método de cálculo de ponderação capaz de melhorar significativamente o desempenho do modelo preditivo ARIMA e heurístico quando aplicados separadamente.

[Shirkoohi, Doghri e Duchesne \(2021\)](#) abordam modelos de redes neurais para previsões de demanda de água de curto prazo (15 min) com otimização dos hiperparâmetros através do emprego de Algoritmos Genéticos. Com base nos resultados, indicaram que a otimização dos hiperparâmetros no modelo de RNA com o algoritmo genético pode melhorar a precisão da predição.

[Yousefi et al. \(2020\)](#) investigaram técnicas de previsão para várias escalas temporais com aplicação da Teoria do Caos e técnica de Aproximação Não-Linear comparando os resultados com os obtidos por meio de Modelo de Regressão Linear Múltipla. De acordo com os autores, os resultados obtidos abrem caminho para investigações mais aprofundadas acerca do comportamento caótico dos valores de consumo urbano de água, em diferentes escalas temporais.

A Tabela 3 apresenta o resultado da análise quanto aos principais métodos abordados pelos 14 trabalhos elencados.

Tabela 3 – Métodos abordados nos trabalhos analisados para previsão de demanda de água de curto prazo.

REFERÊNCIAS	MÉTODOS ABORDADOS					
	Séries Temporais	Redes Neurais e Afins	Regressão	Métodos Híbridos	Regressão por Vetores Suporte e Afins	Meta Heurística
Bata, Carriveau e Ting (2020)	X		X	X		
Candelieri (2017)	X				X	
Gagliardi et al. (2017)		X				
Hu et al. (2021)	X	X		X	X	
Koo et al. (2021)	X	X		X		
Pacchin et al. (2019)	X	X				
Pandey et al. (2021)	X			X		
Ponte et al. (2016)	X	X		X		
Rangel et al. (2017)	X	X		X		X
Salloom, Kaynak e He (2021)	X			X		
Sardinha-Lourenço et al. (2018)	X					X
Shabani et al. (2018)	X			X		X
Shirkoochi, Doghri e Duchesne (2021)	X	X		X		X
Yousefi et al. (2020)			X			X

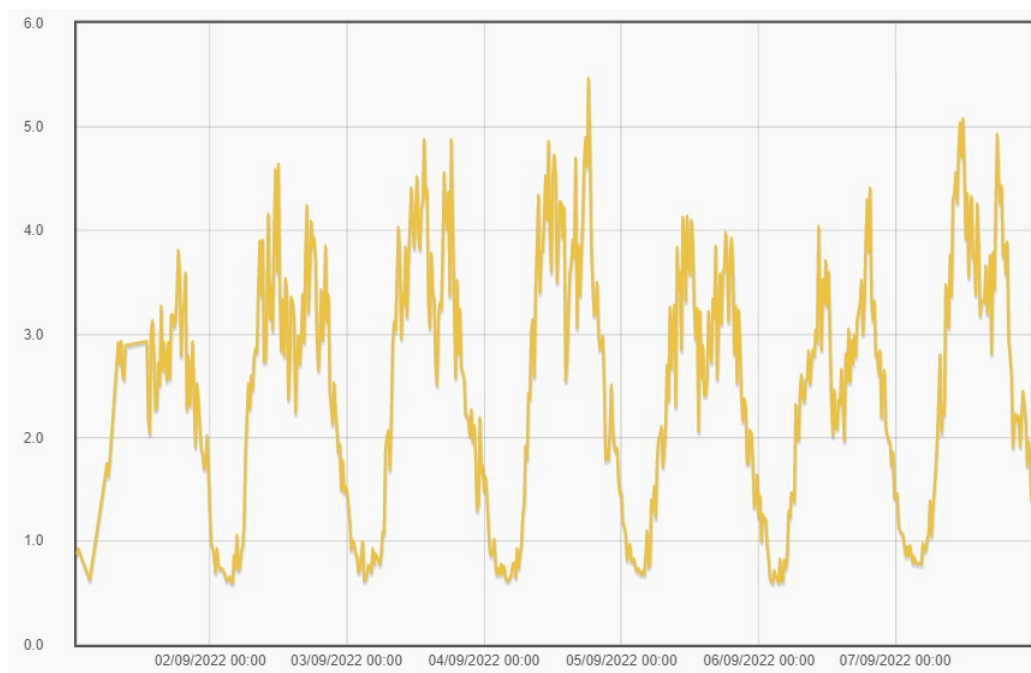
Fonte: o Autor (2023).

3 CONCEITOS BÁSICOS

3.1 Séries Temporais

Uma série temporal é qualquer conjunto de observações ordenadas no tempo (MORETTIN; TOLOI, 2018). Por exemplo, dados de vazão de água na saída de um reservatório registrados ao longo do tempo compõem séries temporais, conforme ilustra a Figura 1.

Figura 1 – Dados de consumo de água em determinado sistema de abastecimento.



Fonte: o Autor (2022)

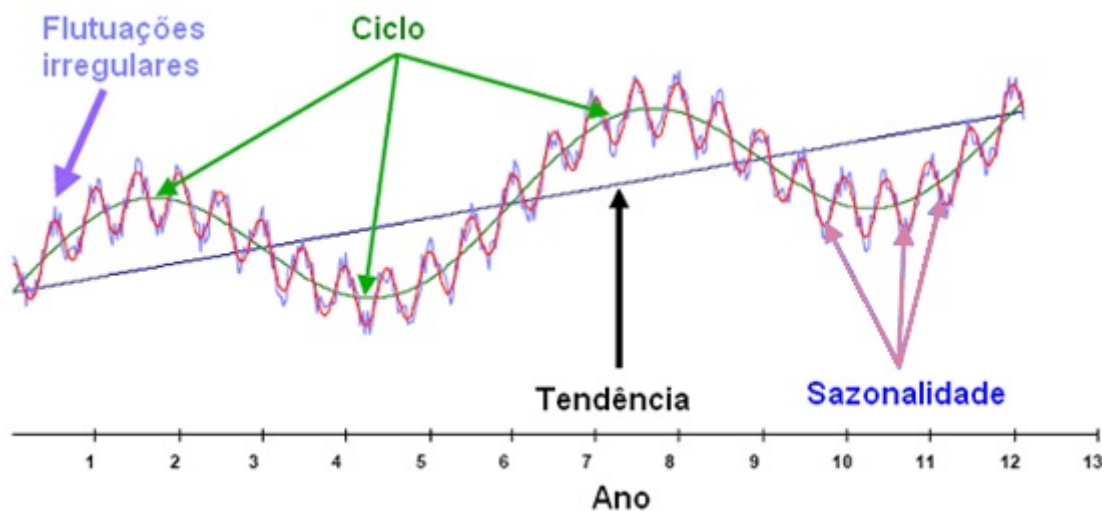
Séries temporais como a ilustrada na Figura 1 são denominadas séries discretas, pois as observações são feitas em tempos específicos, sendo os dados geralmente coletados em períodos equidistantes. Adicionalmente, a série apresentada também pode ser dita univariada e unidimensional pois, neste caso, apresenta a relação de uma única componente (vazão) a um parâmetro único (tempo).

Os dados de séries temporais podem exibir uma variedade de padrões e geralmente é útil dividir uma série temporal em vários componentes, cada um representando uma categoria de padrão subjacente. Conforme Hyndman e Athanasopoulos (2021), podemos pensar em uma série temporal como compreendendo três componentes: um componente de tendência-ciclo, um componente sazonal e um componente aleatório. Nesse sentido, para uma dada série temporal pode-se observar os seguintes componentes:

- **Tendência:** mostra a forma geral da série temporal refletindo aumentos (tendência crescente), diminuições (tendência decrescente) ou estabilidades (tendência estacionária) de longo prazo nas observações. Uma série pode apresentar tendência tanto linear quanto não linear.
- **Ciclicidade:** Um ciclo ocorre quando os dados exibem subidas e descidas que não são de frequência fixa. Essas flutuações geralmente são devidas a condições econômicas e muitas vezes estão relacionadas ao “ciclo de negócios”. A duração dessas flutuações é geralmente de pelo menos 2 anos.
- **Sazonalidade:** variações sazonais, por exemplo, variações relacionadas à época do ano ou o dia da semana. A sazonalidade é sempre de um período fixo e conhecido. A sazonalidade de uma série pode ser aditiva (amplitudes regulares) ou multiplicativa (amplitudes crescentes) dependendo do nível global da série.
- **Erro aleatório:** também conhecida por ruído, é a componente de variáveis aleatórias quando retiradas as demais componentes da série.

A Figura 2 apresenta uma representação dos componentes de uma série temporal.

Figura 2 – Componentes de uma série temporal hipotética.



Fonte: (MOREIRA et al., 2010, p.3)

Dessa forma, séries temporais aditivas podem ser decompostas conforme a Equação 3.1. Já as séries multiplicativas podem ser decompostas de acordo com a Equação 3.2.

$$y_t = S_t + T_t + R_t \quad (3.1)$$

$$y_t = S_t \times T_t \times R_t \quad (3.2)$$

onde:

- y_t são os dados da série;
- S_t é a componente sazonal;
- T_t é a componente de ciclo-tendência;
- R_t denota o erro aleatório (ruído), todos correspondentes ao período de observação t .

A decomposição de uma série temporal para extração de suas componentes pode ser realizada por meio do método clássico baseado em médias móveis, originado na década de 1920, ou por meio de outros métodos mais robustos (HYNDMAN; ATHANASOPOULOS, 2021). Atualmente, pacotes como o statsmodels do *Python* e stats do R possuem funções implementadas para a decomposição de uma série, dividindo-a em vários componentes, cada um representando uma categoria de padrão subjacente de tendência, sazonalidade e ruído. O processo de decomposição facilita, por exemplo, na análise visual de cada componente e entendimento do comportamento da série temporal a ser modelada. A Figura 3 ilustra um exemplo de série aditiva decomposta pelo método clássico.

Além da decomposição, outra importante propriedade que pode caracterizar uma série temporal é a autocorrelação. Conforme Hyndman e Athanasopoulos (2021), assim como a correlação mede a força de uma relação linear entre duas variáveis, a autocorrelação mede a relação linear entre os valores defasados de uma série temporal.

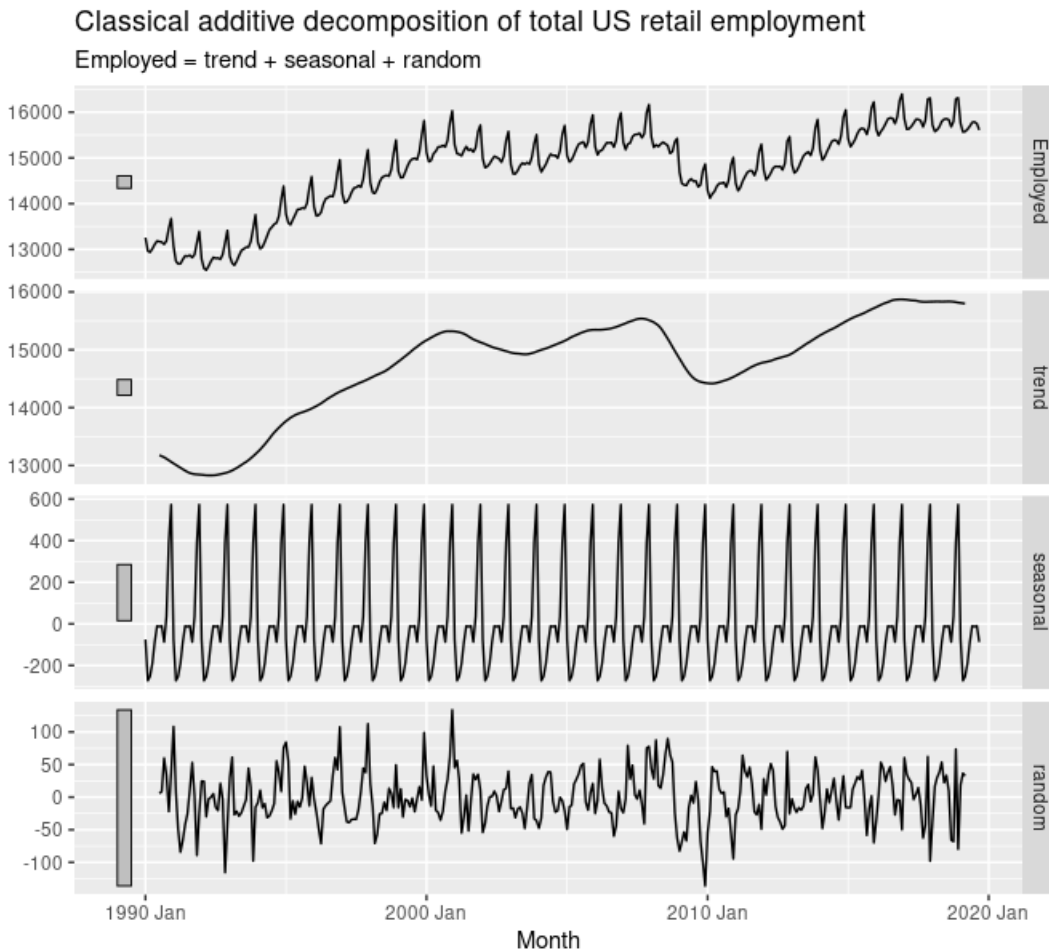
Em uma série temporal podemos obter vários coeficientes de autocorrelação r_k , correspondentes a cada intervalo de defasagem k considerado. Por exemplo, r_1 mede a relação entre y_t e y_{t-1} , r_2 mede a relação entre y_t e y_{t-2} , assim por diante. A equação abaixo mostra como o valor r_k é determinado:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (3.3)$$

onde T é o comprimento da série temporal e \bar{y} é a média das observações.

A representação gráfica (Figura 4) dos valores r_k correspondentes a cada defasagem k é denominada correlograma. Um correlograma se trata portanto de uma representação de como as correlações mudam de acordo com a defasagem. Nesse tipo de gráfico, coeficientes de correlação r_k que ultrapassam as linhas azuis tracejadas indicam a existência de correlação significativa entre observações y_t e y_{t-k} .

Figura 3 – Decomposição clássica de uma série temporal aditiva.

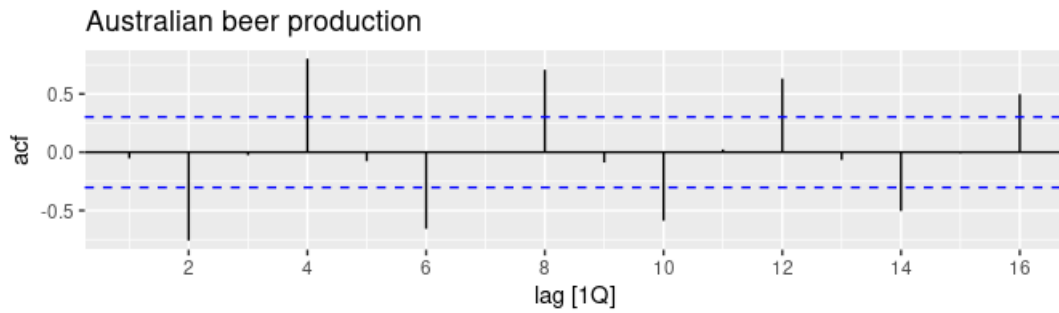


Fonte: (HYNDMAN; ATHANASOPOULOS, 2021)

Finalmente, de acordo com Hyndman e Athanasopoulos (2021), quando os dados apresentam tendência, as autocorrelações para pequenas defasagens costumam ser grandes e positivas porque as observações próximas no tempo também estão próximas em valor. Assim, o correlograma de uma série temporal de tendência costuma ter valores positivos que diminuem lentamente à medida que as defasagens aumentam. Já em relação à sazonalidade, serão observadas correlações maiores para as defasagens correspondentes (múltiplas) ao período sazonal do que para outras defasagens.

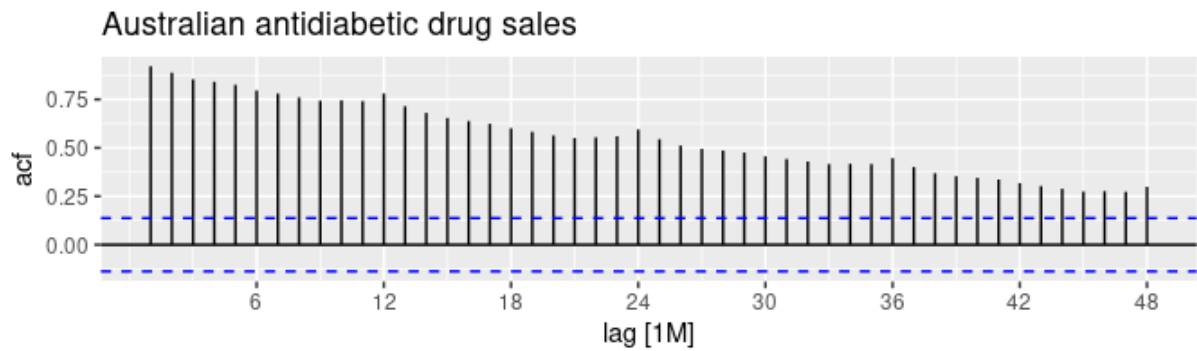
A Figura 5 exemplifica ambos os efeitos, neste caso, combinados: o decaimento lento dos valores de correlação é devido à tendência, por outro lado, os picos observados entre intervalos equidistantes refletem a sazonalidade dos dados.

Finalmente, considera-se que os conceitos básicos aqui apresentados são fundamentais para realizar análise crítica e refinamento dos modelos obtidos por meio dos pacotes desenvolvidos para modelagem de séries temporais. Assim, espera-se obter um modelo

Figura 4 – Valores de correlação computados para cada $lag\ k$ 

Fonte: (HYNDMAN; ATHANASOPOULOS, 2021)

Figura 5 – Correlograma de uma série temporal com tendência e sazonalidade.



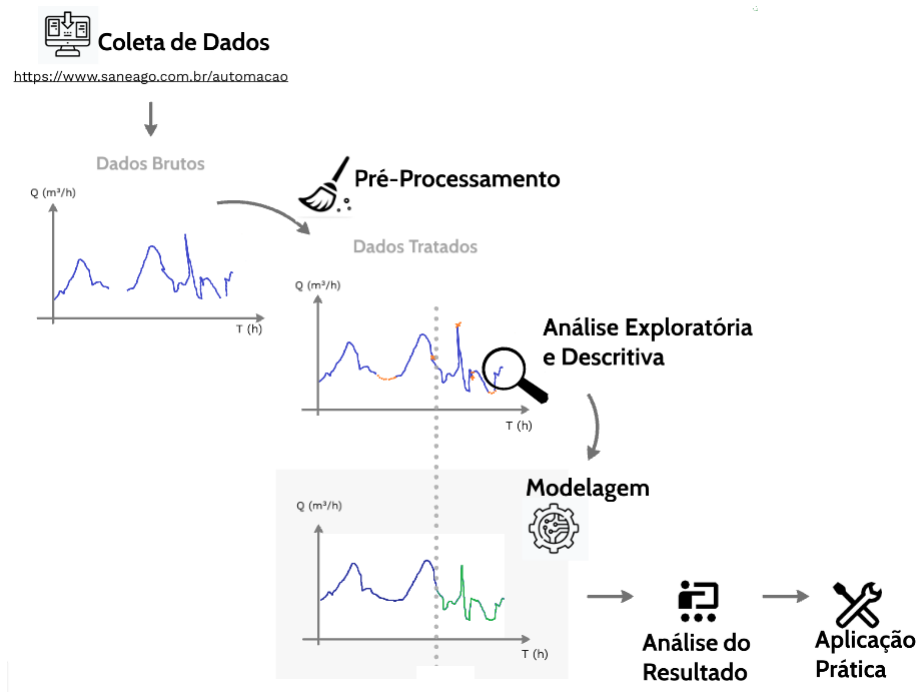
Fonte: (HYNDMAN; ATHANASOPOULOS, 2021)

que reproduza adequadamente os padrões relacionados aos dados observados e que tenha, portanto, capacidade preditiva adequada.

4 METODOLOGIA

De acordo com Ehlers (2009), a análise de séries temporais pode ter diversos objetivos, destacando-se, dentre eles a identificação de padrões, de *outliers* (observações discrepantes), bem como para prever valores futuros com base nas observações passadas. Nesse sentido, a Figura 6 apresenta o processo abordado na análise e modelagem dos dados históricos de consumo de água em um determinado sistema de abastecimento público operado pela Saneamento de Goiás.

Figura 6 – Etapas propostas para o desenvolvimento, análise e aplicação do modelo.



Fonte: o Autor (2023)

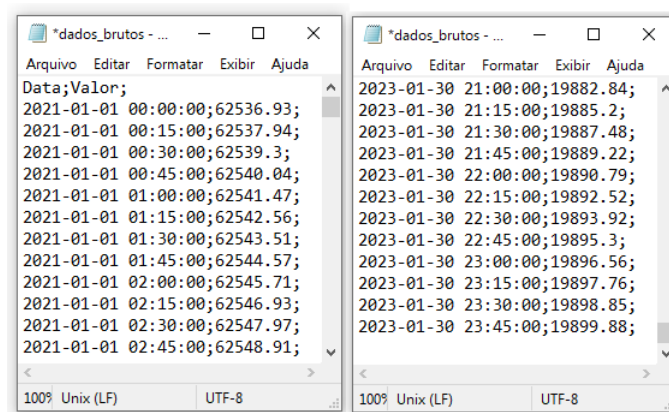
A metodologia adotada pode ser descrita da seguinte forma: iniciou-se a coleta de dados a partir das informações disponíveis na base de dados do sistema supervisor da Saneago, um sistema de supervisão e aquisição de dados desenvolvido por técnicos da empresa. Com os dados em mãos, realizou-se o pré-processamento, tratamento e análise exploratória. Na etapa de modelagem, utilizou-se a biblioteca PyCaret (ALI, 2020) em sua configuração padrão para selecionar, dentre os vários modelos disponíveis, o que apresentou os melhores resultados para os dados estudados. Posteriormente, buscou-se aprimorar o modelo escolhido e comparar seu desempenho preditivo com o de um modelo ingênuo utilizado como referência na análise. Por fim, propôs-se uma abordagem para aplicação do modelo preditivo, com o objetivo de incorporá-lo à prática operacional da Saneago.

4.1 Coleta de Dados

Os dados foram obtidos no ambiente Supervisório *Web* da Saneago (<<https://www.saneago.com.br/automacao>>), o qual possui acesso restrito, permitindo a consulta em tempo real e a realização de *download* do histórico de diversos parâmetros para os sistemas de abastecimento operados pela companhia.

Dessa maneira, para o sistema de distribuição de água analisado, obteve-se a série histórica do volume de água disponibilizado pelo reservatório, com registros a cada 15 minutos, abrangendo o período entre 01/01/2021 e 31/01/2023 (Figura 7).

Figura 7 – Dados coletados do ambiente Supervisório *Web*.



Fonte: o Autor (2023)

4.2 Pré-processamento

O primeiro passo no desenvolvimento de modelos de previsão é preparar os dados no formato correto. Esse processo pode envolver o carregamento dos dados, a identificação de valores faltantes e discrepantes e outras tarefas relacionadas ao pré-processamento.

O pré-processamento dos dados foi realizado em *notebook* escrito em linguagem de programação Python com a finalidade de realizar a leitura dos dados e estruturá-los para as análises posteriores.

Os dados de leitura do macromedidor foram originalmente obtidos no formato CSV (Arquivo de Valores Separados por Vírgulas) e apresentavam uma periodicidade de 15 minutos (Figura 7). Durante essa etapa, utilizou-se a interpolação linear para preencher os valores em casos de dados faltantes.

Com o conjunto de dados completo para cada intervalo de 15 minutos, o volume disponibilizado durante o período de uma hora foi determinado a partir da diferença entre as leituras do macromedidor a cada quatro observações (intervalo de 60 minutos). Por fim,

após a reamostragem dos dados, obteve-se a série temporal com a vazão média horária (m^3/h).

4.2.1 Detecção e remoção de *outliers*

O processo de detecção de valores discrepantes teve início a partir do estabelecimento de valores máximos e mínimos aceitáveis do ponto de vista físico e técnico. Com relação ao limite inferior, considerou-se admissível apenas vazões superiores a $0 \text{ m}^3/\text{h}$ (condição de abastecimento). Já em relação ao limite superior, definiu-se a vazão máxima admissível de $25 \text{ m}^3/\text{h}$, que corresponde ao consumo horário máximo que o sistema de distribuição é capaz de atender, de acordo com sua concepção de projeto.

Após a censura dos dados fora dos limites estabelecidos, realizou-se a detecção e remoção de *outliers* a partir da dispersão da Amplitude Interquartil (IQR - *Interquartile Range*) das observações horárias. Foram determinados os quartis para as observações correspondentes a cada hora do dia, sendo eles:

$Q_{1(h)}$: primeiro quartil, relativo aos primeiros 25% dos dados observados para h;

$Q_{2(h)}$: segundo quartil (mediana), relativo aos primeiros 50% dos dados observados para h;

$Q_{3(h)}$: terceiro quartil, relativo aos primeiros 75% dos dados observados para h;

O IQR foi então calculado por meio da Equação 4.1:

$$IQR_h = Q_3 - Q_1, \quad (4.1)$$

Assim, para cada hora h, determinou-se o limite superior (LS) e o limite inferior (LI) conforme as Equações 4.2 e 4.3:

$$LI_h = Q_1 - 1,5 \cdot IQR_h \quad (4.2)$$

$$LS_h = Q_3 + 1,5 \cdot IQR_h \quad (4.3)$$

Dessa forma, os valores que estavam fora do intervalo entre o limite superior (LS) e o limite inferior (LI) foram substituídos pela mediana das vazões correspondentes àquele respectivo horário do dia.

Com a série temporal tratada em mãos, deu-se início à etapa de análise exploratória e descritiva dos dados.

4.3 Análise Exploratória e Descritiva

Hyndman e Athanasopoulos (2021) e Ehlers (2009) destacam a importância da visualização dos dados na tarefa de análise de dados. De acordo com os autores, os gráficos permitem que muitos recursos dos dados sejam visualizados, incluindo padrões, observações incomuns, mudanças ao longo do tempo e relações entre variáveis.

Para visualização dos dados da série temporal, foi utilizado inicialmente um gráfico temporal, no qual as observações foram plotadas em relação ao tempo. Além disso, foram utilizados gráficos sazonais, gráficos de correlação (correlogramas) e diagramas de caixa (*boxplots*).

4.4 Modelagem

Para se desenvolver um modelo é necessário separar os dados disponíveis em duas partes, dados de treinamento e dados de teste, onde os dados de treinamento são usados para estimar os parâmetros do método de previsão e os dados de teste são usados para avaliar sua precisão.

Nesse estudo, após realizado o tratamento descrito em 4.2, os dados foram separados em dois subconjuntos. Dessa forma, dividiu-se o conjunto de dados da seguinte maneira:

- Conjunto de treinamento (validação cruzada): observações compreendidas entre 01/01/2021 e 31/12/2022 foram utilizadas para a determinação do modelo mais adequado com o emprego de validação cruzada apropriada para séries temporais.
- Conjunto de teste: após determinado o modelo mais adequado aos dados, foram feitas previsões para o período compreendido entre 01/01/2023 e 30/01/2023. As observações deste período foram utilizadas para avaliar a precisão do modelo e compará-lo a um modelo de base (ingênuo).

De acordo com Hyndman e Athanasopoulos (2021), diante da vasta gama de modelos disponíveis, especificar um modelo adequado para os dados é essencial para produzir previsões apropriadas.

Conforme observado na revisão bibliográfica realizada, percebe-se que não existe um método universal que se sobressaia em todas as situações de aplicação e que a escolha do método a ser empregado deve basear-se nos dados disponíveis, bem como no objetivo de aplicação e alcance pretendido.

Ainda de acordo com a revisão de literatura, depreende-se que a previsão de séries temporais em geral, pode ser dividida em modelos clássicos/estatísticos, de aprendizado de máquina (*machine learning*) e aprendizado profundo (*deep learning*). Dentre alguns exemplos desses tipos de modelos, destacam-se:

- Modelos Clássicos / Estatísticos:
 - Médias Móveis (*Moving Averages*);
 - Alisamento Exponencial; (*Exponential Smoothing*);
 - Modelo Autorregressivo Integrado de Médias Móveis - ARIMA (*AutoRegressive Integrated Moving Average*);
 - Modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade - SARIMA (*Seasonal AutoRegressive Integrated Moving Average*).

- Modelos de aprendizado de máquina:
 - Regressão Linear;
 - *Extreme Gradient Boosting (XGBoost)*;
 - Floresta Aleatória (*Random Florest*);

- Modelos de aprendizado profundo:
 - Redes neurais artificiais.

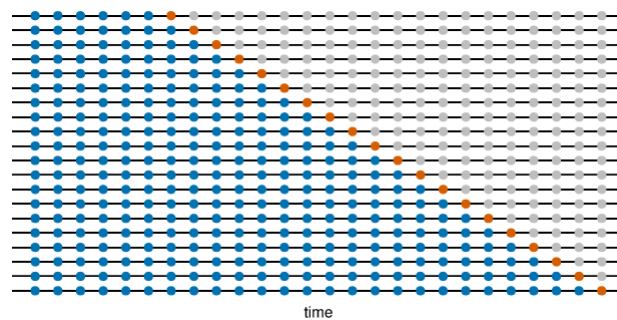
Diante da infinidade de modelos disponíveis, e do grau de complexidade envolvido na implementação dos mais diversos tipos de modelos, julgou-se oportuna a utilização do PyCaret ([ALI, 2020](#)), uma biblioteca de Aprendizado de Máquina Automatizado (AutoML) desenvolvida em Python e que não requer uma quantidade grande de códigos para o processo de modelagem de dados. A biblioteca, cuja versão atual é a PyCaret 2.3.10 de abril de 2022, conta com um módulo específico para séries temporais onde é possível empregar de forma simples e prática desde técnicas de modelagem tradicionais, bem como modelos clássicos de aprendizado de máquina tais como modelos de regressão, *Gradient Boosting* e árvores de regressão.

De acordo com [Ali \(2020\)](#), em comparação com outras bibliotecas de aprendizado de máquina de código aberto, o PyCaret é uma alternativa *low-code* para um processo de modelagem mais prático e eficiente por englobar uma variedade de outras bibliotecas e *frameworks* implementados em *python* como scikit-learn, XGBoost, LightGBM, CatBoost, Optuna, Hyperopt, Ray, entre outros.

No PyCaret, a modelagem pode ser feita de maneira intuitiva e prática, iniciando-se com a inserção da série completa, a especificação dos subconjuntos de treino e teste, do horizonte de previsão para avaliar e comparar os modelos disponíveis na biblioteca. Adicionalmente, a biblioteca também permite a inclusão de variáveis exógenas para o treinamento dos modelos. Além disso, pode-se optar pelo emprego da validação cruzada para o treinamento dos modelos.

De acordo com [Hyndman e Athanasopoulos \(2021\)](#), a validação cruzada em modelagem de séries temporais se trata de uma abordagem mais sofisticada dos conjuntos de treinamento/teste. Neste procedimento, há uma série de conjuntos de testes, e os conjuntos de treinamento correspondentes consistem apenas de observações que ocorreram antes da observação utilizada no teste. Assim, nenhuma observação futura pode ser usada na construção da previsão. A Figura 8 mostra o funcionamento desta técnica de validação, onde as observações azuis formam os conjuntos de treinamento e as observações laranja formam os conjuntos de teste.

Figura 8 – Validação cruzada de série temporal.



Fonte: ([HYNDMAN; ATHANASOPOULOS, 2021](#))

Dessa forma, o PyCaret conta com diversos modelos implementados e opção pelo emprego de técnicas de validação cruzada, como a descrita por [Hyndman e Athanasopoulos \(2021\)](#). Entretanto, de acordo com [Ali \(2020\)](#), muitas vezes não é viável a aplicação de todos os modelos disponíveis no PyCaret pois a depender do horizonte de predição e do número de *folds* especificados, pode-se levar bastante tempo para que todos eles sejam treinados. Nesse sentido, a biblioteca possui como padrão o parâmetro *turbo*, que exclui da análise modelos que requeiram longos períodos de treinamento.

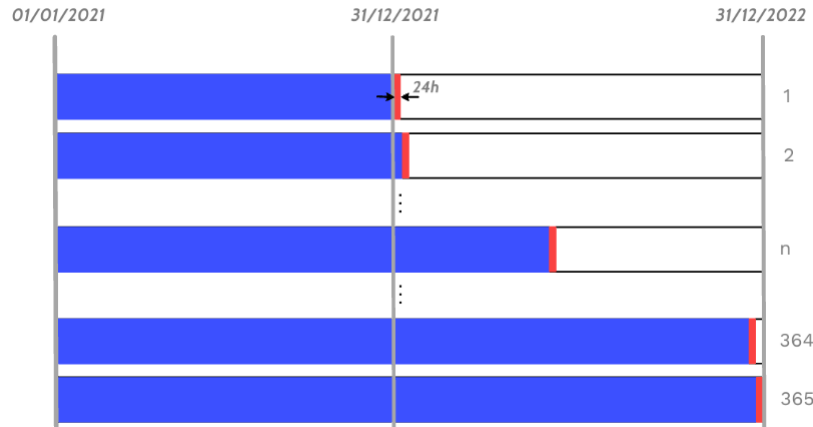
Diante disso, no processo de escolha do melhor modelo do PyCaret empregou-se a validação cruzada temporal com janela expansiva (*Expanding Window*) com 365 *folds* e a utilização do parâmetro padrão `Turbo = True`.

Conforme ilustra a figura 9, na abordagem empregada, cada modelo objeto da análise no PyCaret foi treinado 365 vezes. O conjunto de dados utilizado no treinamento do *fold* 1 compreendeu todas as observações de 2021, para o qual foram geradas previsões para as 24 horas sucessivas para avaliação do seu desempenho. Em seguida, os dados das 24 horas seguintes foram incorporados ao conjunto de dados para o próximo treinamento (*fold* 2), sucessivamente até que todo o conjunto de dados fosse utilizado.

Ao final de todas as iterações (treinamento, predição e avaliação) dos 365 *folds* do processo de validação cruzada, as métricas de avaliação foram calculadas como a média

das métricas obtidas em cada iteração. Isso forneceu uma medida geral da performance de cada modelo em todo o conjunto de dados, que foi utilizada para comparar os diferentes modelos e escolher o melhor deles.

Figura 9 – Abordagem de validação cruzada empregada.



Fonte: o Autor (2023)

4.4.1 Escolha do Melhor Modelo

Após realizado todo o processo descrito, obteve-se como resultado a relação de todos os modelos analisados do PyCaret, ordenados de acordo com o seu respectivo desempenho preditivo médio. Para isso, a biblioteca apresenta diversas métricas de avaliação baseadas no erro das previsões para o conjunto de teste. De acordo com [Hyndman e Athanasopoulos \(2021\)](#), o erro de uma previsão é a diferença entre esta previsão e o valor efetivamente observado. Nesse sentido, as métricas consideradas para a avaliação da precisão das previsões neste estudo foram:

- Erro Médio Absoluto (MAE)

$$MAE = média(|e_t|) \quad (4.4)$$

- Raiz do Erro Quadrático Médio (RMSE)

$$RMSE = \sqrt{média(e_t^2)} \quad (4.5)$$

Com base nas métricas de desempenho, escolheu-se dentre todos os modelos aquele que apresentou o melhor desempenho.

4.4.2 Modelo de Referência

De acordo com [Hyndman e Athanasopoulos \(2021\)](#), alguns métodos de previsão são extremamente simples e surpreendentemente eficazes, dentre eles destacam-se aqueles conhecidos por Modelos Ingênuos. Algumas vezes, esses modelos podem ser o melhor método de previsão disponível, porém, em muitos casos, servem como referência (base para comparação) e não como o método de escolha. Ou seja, recomenda-se que quaisquer métodos de previsão proposto sejam comparados a esses métodos simples para garantir que o modelo proposto tenha desempenho superior. Caso contrário, não vale a pena considerar um novo método mais complexo.

Apesar da biblioteca PyCaret contar com a implementação de alguns modelos deste tipo, optou-se neste trabalho por utilizar uma adaptação do modelo ingênuo apresentado por [Salomons e Housh \(2020\)](#) proposto especificamente para previsão de consumo horário de água.

Dessa forma, o modelo ingênuo proposto calcula a mediana das quatro últimas observações correspondentes àquela hora e dia da semana, nas últimas quatro semanas. Assim, no intervalo de quatro semanas temos 672 observações ($4 \times 7 \times 24$). Considerando t como o tempo absoluto atual em horas, as observações correspondentes às quatro semanas anteriores são denotadas, respectivamente, por y_{t-168} , y_{t-336} , y_{t-504} e y_{t-672} . Portanto, a demanda \hat{y}_t prevista para a próxima hora é dada pela Equação 4.6.

$$\hat{y}_t = \text{mediana}(y_{t-168}, y_{t-336}, y_{t-504}, y_{t-672}) \quad (4.6)$$

Acredita-se que o Modelo Ingênuo apresentado pela Equação 4.6, apesar de simples pode retornar predições adequadas por incorporar dados refletindo a tendência das últimas semanas. Diferentemente do modelo baseado na média aritmética empregado por [Salomons e Housh \(2020\)](#), considerou-se aqui um modelo com base na mediana por acreditar que este seja mais adequado para casos práticos (menos sensível a valores extremos).

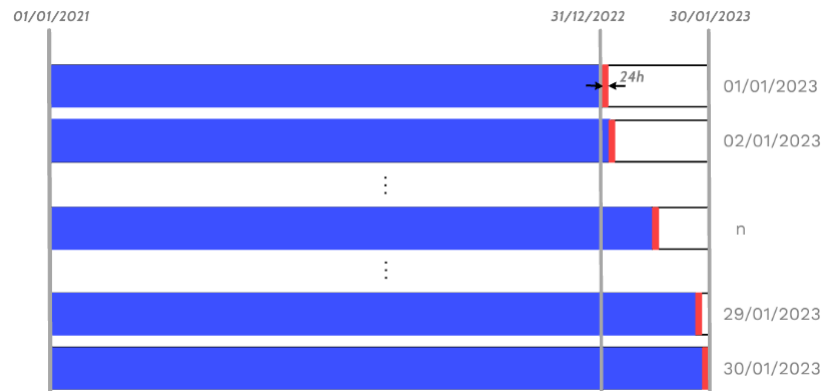
4.4.3 Teste do Modelo

Foram realizadas predições para os trinta primeiro dias do ano de 2023. Na abordagem proposta (fig. 10), cada uma das predições realizadas foi obtida por meio do uso do melhor modelo conforme descrito em 4.4.4.

Para o treinamento e predição, foram considerados os parâmetros padrão do PyCaret e o uso da função de refinamento denominada "tune_model". Essa função possibilita o ajuste fino de um modelo de previsão por meio de uma abordagem *grid search* para testar diferentes combinações de hiper-parâmetros para encontrar a melhor configuração disponível. Finalmente, manteve-se o parâmetro padrão da função denominado "choose_better",

garantindo-se assim que o PyCaret retornasse automaticamente o melhor resultado obtido para as diferentes combinações de hiper-parâmetros.

Figura 10 – Abordagem empregada para treinamento e predição.



Fonte: o Autor (2023)

4.4.4 Análise do Resultado

Conforme descrito em 4.4.3, foram realizadas predições para janeiro de 2023 empregando-se o melhor modelo obtido com o emprego da biblioteca PyCaret. Adicionalmente, realizou-se predições para o mesmo período para o modelo de referência (ingênuo - 4.4.2). Ambos os modelos foram comparados quanto ao desempenho preditivo, suas vantagens e desvantagens. Em seguida, foi proposta uma abordagem para aplicação prática desses modelos objetivando otimizar a operação do sistema de abastecimento de água objeto deste estudo.

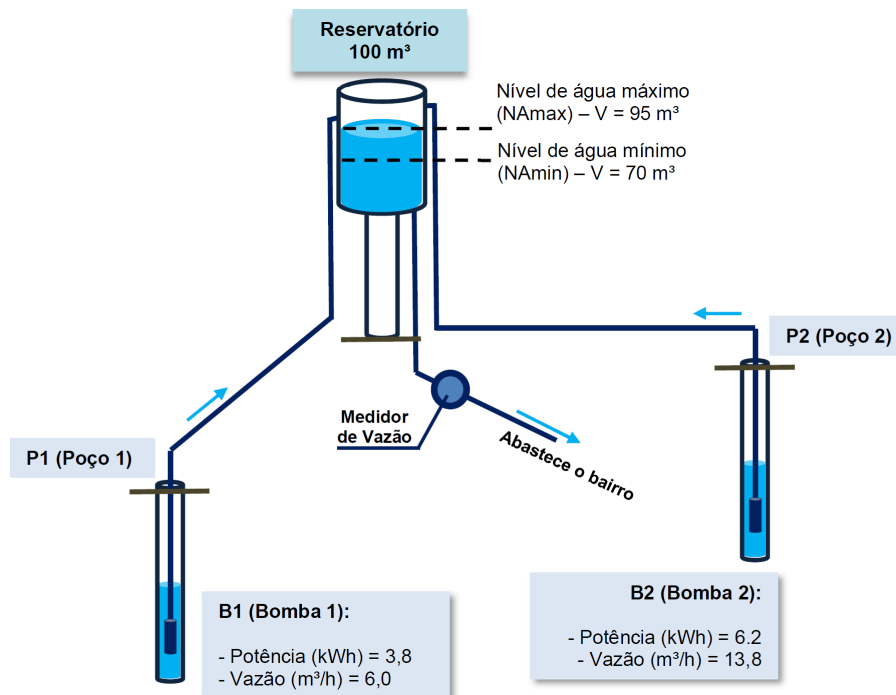
4.5 Aplicação prática do modelo preditivo

Dentre as principais aplicações de modelos preditivos de consumo de água a curto prazo destaca-se a otimização da operação do sistema de abastecimento (DONKOR et al., 2014). Essa otimização pode ter como objetivo, por exemplo, reduzir as despesas operacionais relacionadas ao consumo de energia pelos equipamentos de bombeamento.

Os sistemas de distribuição de água, em sua maioria, contam com bombas que são utilizadas para o transporte da água para os pontos mais altos, onde geralmente encontram-se instalados reservatórios para armazenamento de água. A Figura 11 apresenta um modelo esquemático com informações do sistema de abastecimento abordado nesse estudo.

Atualmente, a Saneago opera inúmeros sistemas que possuem configuração similar ao apresentado. Todos os sistemas em operação pela empresa são automatizados e

Figura 11 – Sistema de Abastecimento de Água estudado.



Fonte: o Autor (2023)

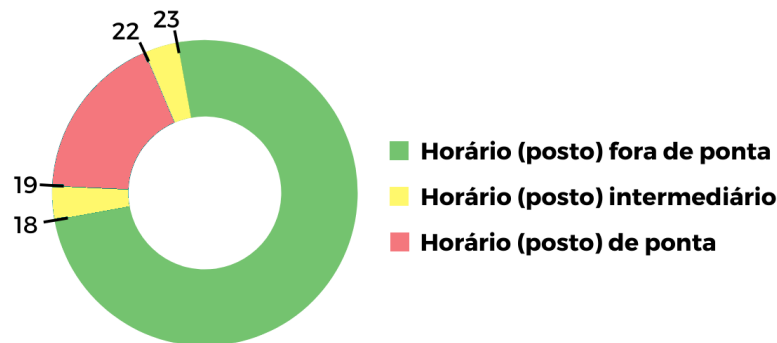
interligados a um sistema supervisorio em nuvem desenvolvido por técnicos da empresa e que vem evoluindo com o passar dos anos. O acionamento das bombas nesses sistemas é geralmente ditado pelo nível observado no reservatório. Nesse caso, se o reservatório atinge um determinado nível mínimo permitido, as bombas são acionadas garantindo assim o enchimento do reservatório. No sentido oposto, ao se atingir um determinado nível máximo estabelecido, as bombas são desligadas a fim de que não haja extravasamentos.

Atualmente, existem soluções disponíveis no mercado (ex: [SCUBIC \(2023\)](#), [Bentley \(2023\)](#) e [Suez \(2023\)](#)) que propõem o uso de diversas técnicas de aprendizagem de máquina e de otimização para determinar o melhor regime de funcionamento de bombas. Essas soluções tiram proveito das tarifas elétricas variáveis ao longo do dia para que as bombas sejam acionadas no melhor momento possível, com garantia de atendimento ao consumo demandado.

Em relação a tarifa elétrica variável, desde 2020 encontra-se em vigor no Brasil a possibilidade de escolha pela opção tarifária denominada “Tarifa Branca”. De acordo com a [AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL \(2023\)](#), “diferente da modalidade Convencional, que tem um único valor de tarifa, a Tarifa Branca possui valores diferentes ao longo do dia. Nos dias úteis, existem 3 valores de tarifa, aplicados de acordo com os períodos (postos)”. A tarifa é mais elevada nos horários tidos como

“Ponta”, assume valor intermediário nos postos “Intermediários”, e o valor mínimo nos horários “Fora Ponta”. Em finais de semana e feriados nacionais, o valor mínimo aplica-se às 24 horas. A Figura 12 apresenta os postos tarifários da distribuidora Equatorial Energia Goiás (Equatorial Energia Goiás, 2023).

Figura 12 – Tarifa Branca - Postos Tarifários no Estado de Goiás para dias úteis.



Fonte: o Autor (2023)

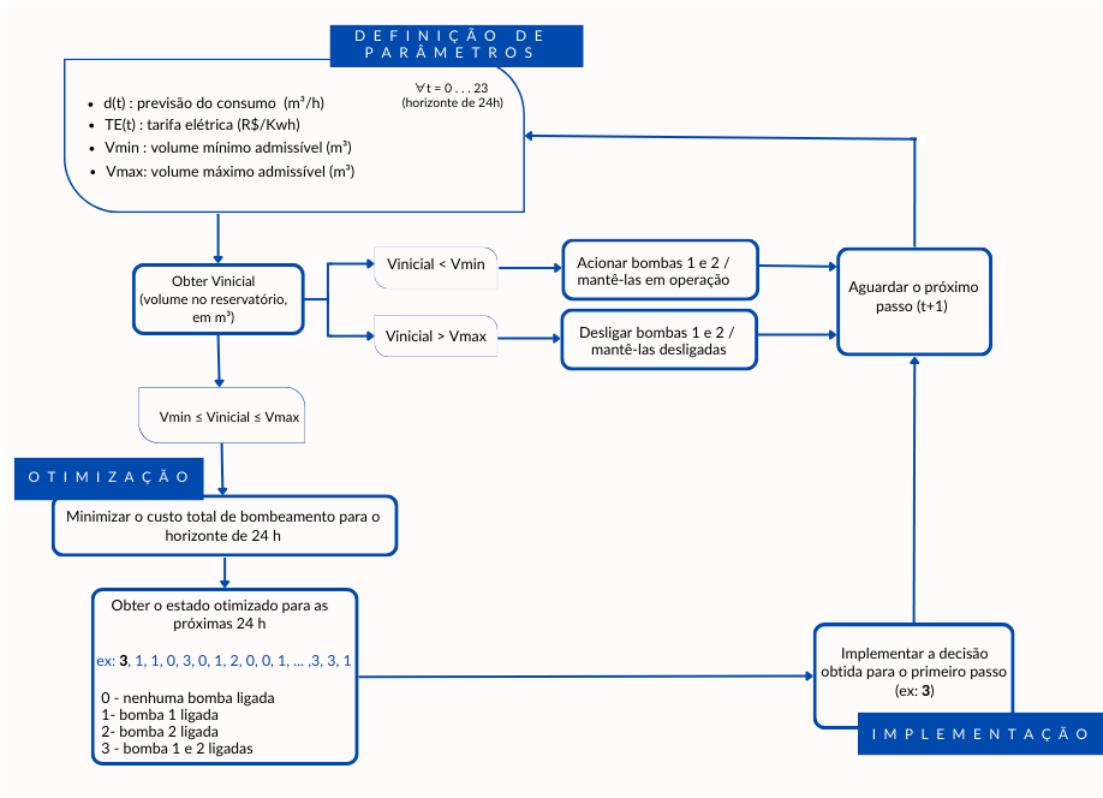
Conforme exposto, o sistema de automação atualmente em operação na Saneago, apesar de bastante eficaz, não integra valores gerados pelo grande número de dados produzidos. Dessa forma, atualmente não são desenvolvidos nem aplicados modelos preditivos de demanda de água. Além disso, a empresa ainda continua adotando a tarifa convencional para a maior parte de suas unidades consumidoras, perdendo muitas vezes a chance de aproveitar horários de menor custo para o bombeamento.

Em busca por abordagens na literatura que nortearassem as melhorias vislumbradas, foram encontrados alguns estudos que propõem a minimização dos gastos com energia elétrica por meio da busca pelo melhor regime de bombeamento dentro das condições limitantes impostas. As possíveis soluções para o problema geralmente são condicionadas ao princípio físico de balanço hídrico, bem como no estabelecimento de níveis operacionais máximo e mínimo.

Nesse sentido, Salomons e Housh (2020) exploram o tema e propõem metodologia para otimização do funcionamento de bomba em sistemas de distribuição de água, com base nos critérios já discutidos. Para este estudo, foi proposta uma abordagem adaptada de Salomons e Housh (2020). A figura 13 apresenta a abordagem proposta para o sistema de abastecimento em estudo.

A seguir, as principais etapas do processo de otimização proposto são descritas em mais detalhes.

Figura 13 – Abordagem proposta para operação otimizada das bombas.



Fonte: o Autor (2023)

4.5.1 Definição de Parâmetros

A primeira etapa da abordagem proposta consiste na definição dos parâmetros que serão utilizados no processo de otimização. Dentre eles pode-se destacar a necessidade de um modelo preditivo para o horizonte de otimização considerado e os limites impostos para o volume mínimo e máximo admissível (figura 11). Além disso, é necessário conhecer o valor da tarifa elétrica ao longo do horizonte estabelecido (ex: 24 horas). Os valores tarifários (sem tributos) de energia em vigor no estado de Goiás em janeiro de 2023 para as duas modalidades de tarifa (tarifa convencional e tarifa branca) são:

- **Tarifa convencional**
 - 0.671 R\$/kWh.
- **Tarifa branca**
 - 0.564 R\$/kWh (horário fora de ponta);
 - 0.783 R\$/kWh (horário intermediário);
 - 1.165 R\$/kWh (horário de ponta).

Adicionalmente, dados referentes aos equipamentos de bombeamento (vazão fornecida - m³/h e energia consumida - kWh), apresentados na figura 11, foram imprescindíveis para a determinação dos volumes no reservatório conforme os possíveis cenários de acionamento/desligamento das bombas. Com essas informações foi possível formular o problema de otimização objetivando minimizar o gasto com energia devido o bombeamento.

4.5.2 Otimização

O problema de otimização aqui considerado teve como objetivo minimizar o custo de operação (R\$) devido o processo de bombeamento de água. Para isso, utilizou-se as informações necessárias para os cálculos, e que foram apresentas na figura 11 e na seção 4.5.1.

Para o processo de otimização, definiu-se a representação da variável S como o conjunto dos possíveis estados de funcionamento (st) das bombas ao longo de 24 horas. Os valores que essa variável pode assumir foi convencionado da seguinte maneira:

- 0 : nenhuma bomba em funcionamento;
- 1: apenas a bomba 1 em funcionamento;
- 2: apenas a bomba 2 em funcionamento;
- 3: as duas bombas funcionando.

Dessa forma, a variável pode assumir a seguinte forma:

$$S = (st(0), st(1), \dots, st(22), st(23)) \quad (4.7)$$

sendo $st(t) = [0,1,2 \text{ ou } 3]$ para $t = 0, \dots, 23$.

Os parâmetros de vazão e consumo de energia relacionados a cada possível estado de funcionamento das bombas (0, 1, 2 ou 3) são apresentados na tabela 4

Tabela 4 – Vazão e energia consumida por cada estado.

st	Q (m ³ /h)	E (kW)
0	0	0
1	6.0	3.8
2	13.8	6.2
3	19.8	10

Fonte: o Autor (2023).

Levando em conta as considerações apresentadas, o problema de otimização a ser resolvido foi definido da forma apresentada na expressão 4.8:

$$\min_{st(0), st(1), \dots, st(23)} \sum_{t=0}^{23} TE(t) \cdot E[st(t)] \quad (4.8)$$

sujeito a:

$$V_{\min} \leq V(t) \leq V_{\max} :$$

$$\text{- Para } t = 0: V(0) = V_{\text{inicial}} + Q[st(0)] - d(0)$$

$$\text{- Para } 0 < t \leq 23: V(t) = V(t-1) + Q[st(t)] - d(t)$$

onde:

- $TE(t)$ é a tarifa elétrica, em R\$/Kwh, correspondente ao período t ;
- $E[st(t)]$ é o consumo energético, em kWh, correspondente ao estado de funcionamento da bomba st (0,1,2 ou 3) atribuído para o período t ;
- V_{\min} é o volume mínimo admissível, em m^3 , para o reservatório;
- V_{\max} é o volume máximo admissível, em m^3 , para o reservatório;
- V_{inicial} é a condição inicial, correspondente ao volume existente no reservatório em m^3 ;
- $V(t)$ é o volume, em m^3 , simulado para o reservatório para o período t ;
- $V(t-1)$ é o volume, em m^3 , determinado para o reservatório para o período anterior a t .
- $Q[st(t)]$ é a vazão, em m^3/h , correspondente ao estado de funcionamento da bomba st (0,1,2 ou 3) atribuído para o período t ;
- $d(t)$ é a demanda de água, em m^3/h , predita para o período t , obtida por meio de um modelo de previsão de consumo.

O problema formulado foi implementado em *python* com o emprego do pacote Pyomo (HART et al., 2020). O Pyomo é um pacote de modelagem matemática de problemas de otimização de maneira intuitiva e flexível. Com o uso do pacote o modelo de otimização pode ser definido, incluindo-se as funções objetivo, variáveis e restrições. O problema pode então ser resolvido por meio de *solvers* de otimização, pois o Pyomo pode se integrar com diversos tipos de *solvers* de código aberto ou comerciais.

Dentre os *solvers* de código aberto possíveis de serem integrados ao Pyomo, destacam-se os implementados e mantidos pelo *COmputational INfrastructure for Operations Research* - COIN-OR (COIN-OR FOUNDATION,). Neste problema de otimização,

seguiu-se a mesma abordagem empregada por [Salomons e Housh \(2020\)](#), tendo sido utilizado o *solver COIN-OR Branch-and-Cut*, conhecido como CBC, uma implementação do método *Column Generation Branch-and-Cut* ([CORNUÉJOLS; LIU, 2001](#)).

Assim, com o emprego das técnicas descritas anteriormente, o problema de otimização (caso haja solução viável) retorna a solução ótima na forma apresentada em 4.7. A implementação da rotina de otimização proposta incluiu condições adicionais para reduzir o gasto computacional envolvido, dentre elas, destacam-se a verificação do volume V_{inicial} e o estabelecimento de tempo limite na busca pela solução ótima.

Para o primeiro caso (verificação do volume V_{inicial}), caso este fosse inferior ao V_{min} , retornou-se de imediato como solução o estado de funcionamento $st=3$ (duas bombas acionadas). Analogamente, caso $V_{\text{inicial}} > V_{\text{max}}$, obteve-se como solução o estado de funcionamento $st=0$ (ambas as bombas desligadas). Finalmente, caso V_{inicial} estivesse dentro dos limites mínimo e máximo impostos, prosseguiu-se com o processo de otimização por meio do *solver* CBC.

Finalmente, dentre as configurações do *solver* disponibilizadas pelo Pyomo, destaca-se o emprego do tempo máximo (parâmetro *timelimit*) disponível para busca da solução ótima. Dessa forma, caso ultrapassado este tempo estabelecido, o *solver* retorna a solução encontrada até aquele momento. Como a abordagem de otimização proposta neste estudo é destinada ao controle em tempo-real, um limite de tempo prático para cada execução de otimização foi definido em 5 minutos, consoante com o proposto por [Salomons e Housh \(2020\)](#).

4.5.3 Implementação

A estratégia de otimização descrita possibilita obter os estados de funcionamento das bombas para um período de 24 horas. Entretanto, não se tem garantia que o sistema irá se comportar em conformidade com as condições impostas na otimização ao longo das 24 horas. Dessa forma, em busca de garantir que os estados de bombeamento propostos correspondam ao estado ótimo dentro das condições apresentadas, propõe-se a implementação apenas do estado de bombeamento correspondente ao primeiro passo t . Em seguida, aguarda-se o próximo passo $t+1$ e é verificada a nova condição do sistema (volume observado no reservatório). O parâmetro V_{inicial} então assume este valor, e a busca pela solução ótima é novamente realizada. Dessa forma, o processo se repete a cada passo de tempo, garantindo assim que o estado do sistema seja atualizado a cada hora.

A fim de comparar os resultados potencialmente obtidos com o emprego da estratégia de otimização discutida, foi proposta a aplicação da abordagem para o mês de janeiro de 2023, onde se considerou cenários com o emprego da tarifa convencional e da tarifa branca para o modelo preditivo desenvolvido de acordo com o exposto em 4.4. Adicionalmente, empregou-se a mesma estratégia, porém considerando um modelo preditivo ideal (caso

hipotético de modelo com 100% de acerto nas predições).

Dessa forma, foi possível comparar não só a vantajosidade da estratégia de otimização do bombeamento proposta, mas também avaliar o impacto da precisão do modelo preditivo nos resultados obtidos.

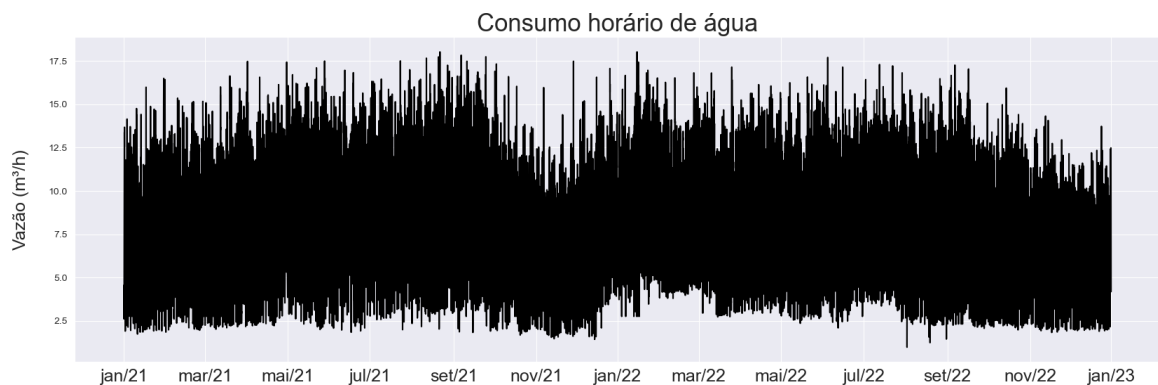
5 RESULTADOS

5.1 Coleta e Pré-processamento

Os registros de consumo de água referentes ao período entre janeiro de 2021 e janeiro de 2023 foram coletados da base de dados do sistema supervisorio da Saneago. Como esperado, o conjunto de dados apresentou algumas inconsistências, tais como registros ausentes para determinados períodos, presença de registros duplicados e valores discrepantes em algumas observações. Para solucionar tais problemas, realizou-se um processo de pré-processamento e tratamento dos dados, como descrito na seção 4.2.

Após a conclusão das etapas de pré-processamento e tratamento dos dados, que incluíram a reamostragem para intervalos horários, a detecção e remoção de *outliers* e a atribuição de valores para dados faltantes, obteve-se o conjunto final de dados tratados, conforme ilustrado na figura 14.

Figura 14 – Série temporal de consumo de água.

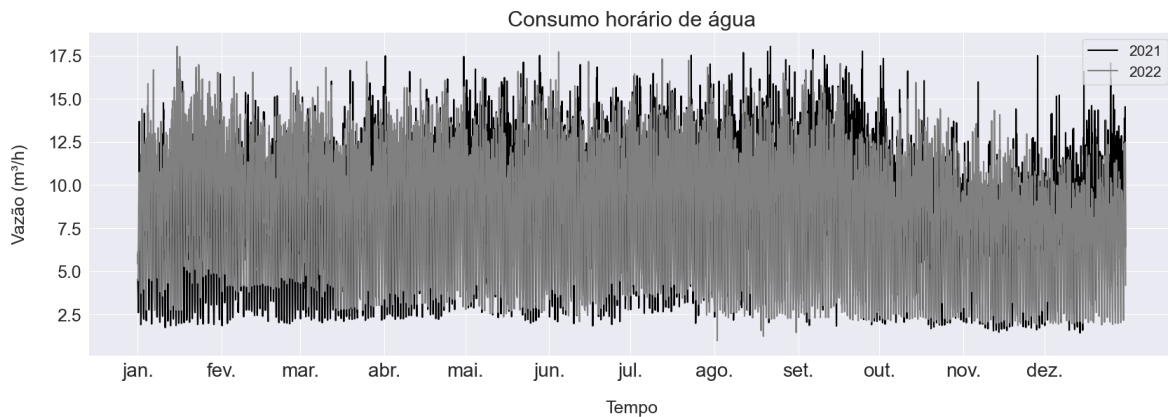


Fonte: o Autor (2023)

5.2 Análise Exploratória e Descritiva

Devido à complexidade das séries temporais de consumo de água, que geralmente apresentam mais de um padrão sazonal, existem métodos específicos que podem ser utilizados para analisar cada um desses padrões. Assim, foi gerado inicialmente um gráfico sazonal que reflete os padrões anuais de consumo de água no sistema de abastecimento estudado (figura 15).

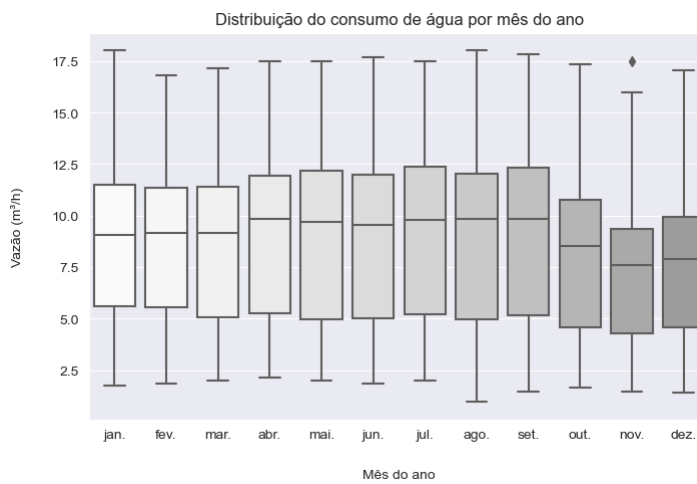
Figura 15 – Padrões anuais para o consumo de água.



Fonte: o Autor (2023)

A análise das Figuras 14 e 15 possibilita observar a variação anual da demanda horária de água, mostrando a ocorrência de queda do consumo no último trimestre de cada ano. A Figura 16, por sua vez, apresenta o diagrama de caixas das variações observadas ao longo dos meses.

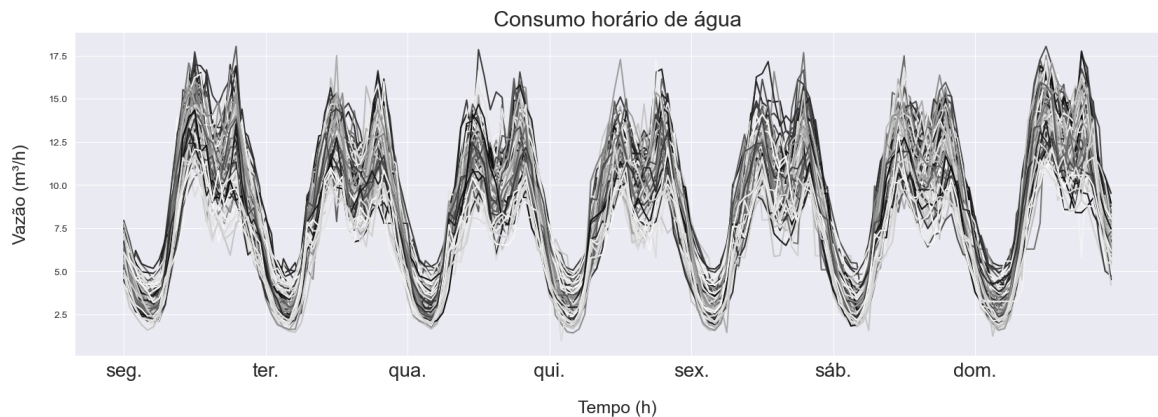
Figura 16 – Variações mensais de consumo horário do sistema estudado.



Fonte: o Autor (2023)

Realizada a análise da sazonalidade anual do consumo de água, a próxima etapa consistiu na análise da sazonalidade semanal. Para tal, foi plotado o gráfico sazonal que apresenta os padrões semanais de consumo de água no sistema de abastecimento estudado, conforme ilustrado na Figura 17.

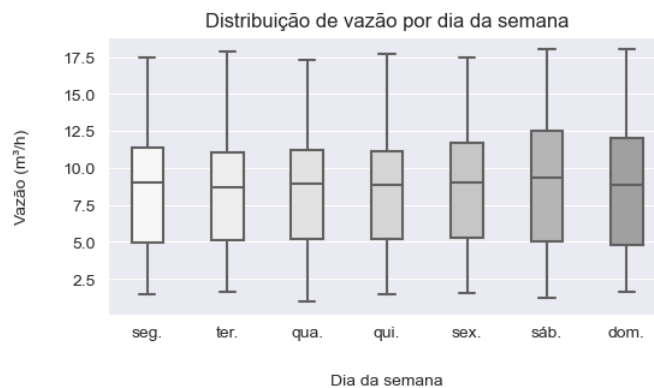
Figura 17 – Padrões semanais para o consumo de água.



Fonte: o Autor (2023)

Os padrões semanais identificados mostram que há um aumento no consumo de água nos fins de semana. Isso é reforçado pelo diagrama de caixas apresentado na Figura 18, que mostra um aumento no consumo nos dias de sábado e domingo.

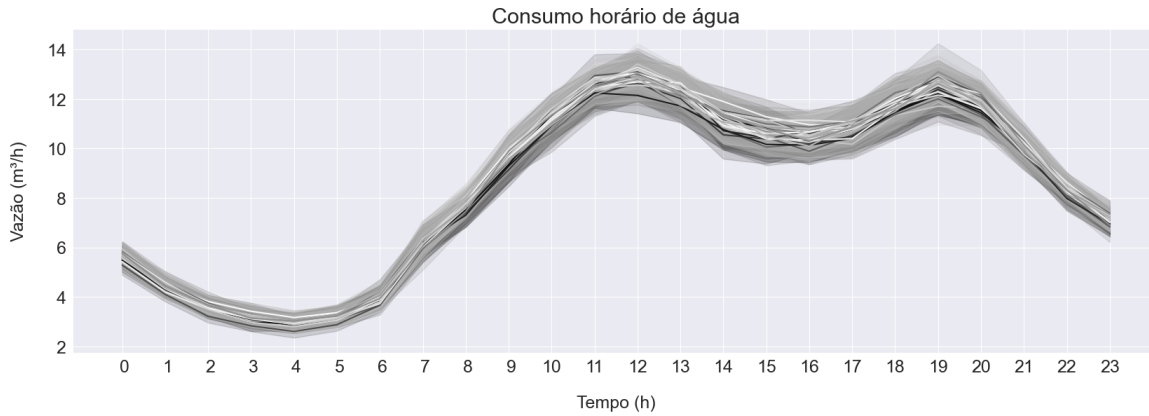
Figura 18 – Variações do consumo horário ao longo da semana.



Fonte: o Autor (2023)

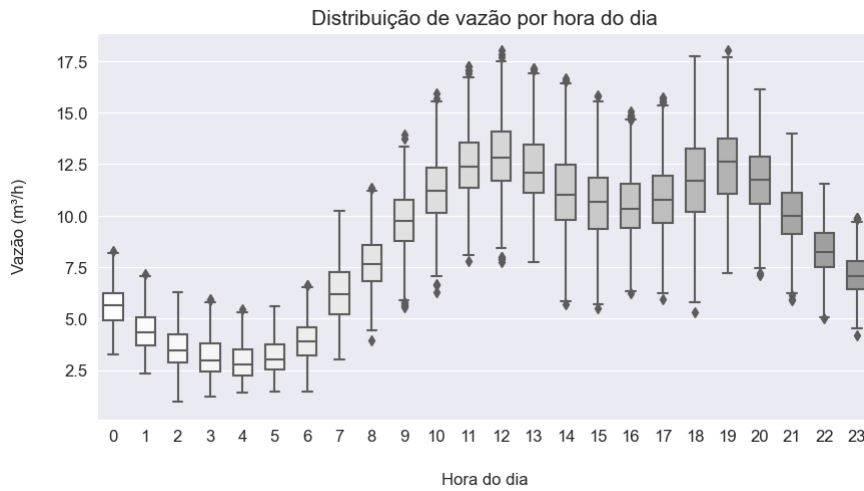
A figura 19 apresenta os resultados obtidos para a sazonalidade horária observada ao longo dos dias. É possível observar um padrão de consumo de água que apresenta variação ao longo do dia, com um pico no final da manhã e outro no início da noite. Por outro lado, durante a madrugada, verifica-se que o consumo de água é mínimo. O diagrama de caixas para as variações horárias, apresentado na figura 20, confirma esse comportamento típico da curva de consumo em sistemas de abastecimento de água.

Figura 19 – Padrões diários para o consumo de água.



Fonte: o Autor (2023)

Figura 20 – Variações do consumo ao longo do dia.



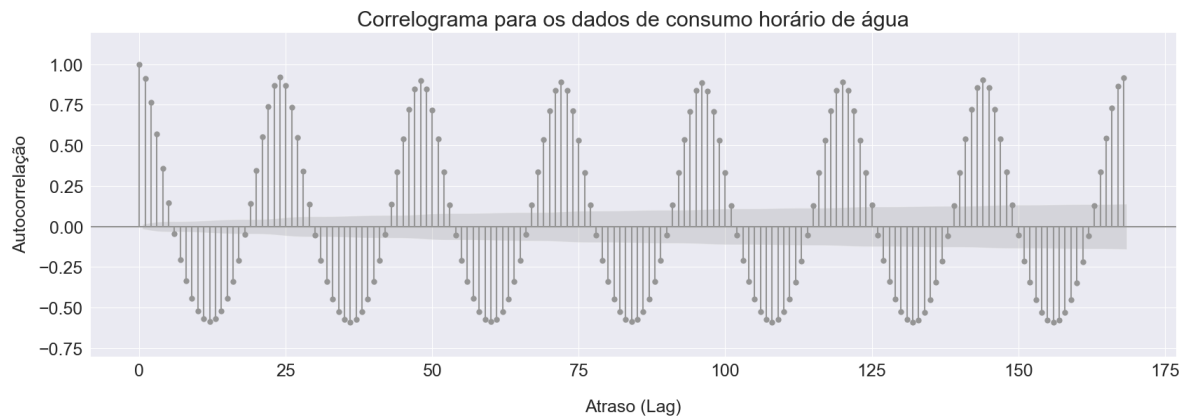
Fonte: o Autor (2023)

Após as análises realizadas para os diferentes tipos de padrões sazonais, constatou-se a existência de múltiplos padrões, tais como anual, semanal e diário. Entretanto, é possível observar que o padrão diário se destaca dos demais, uma vez que as variações do consumo ao longo do dia (Figura 19) apresentam maior relevância em relação às variações do consumo entre os dias da semana (Figura 18) e os meses do ano (Figura 16).

Para investigar a relação entre as observações, foi construído o correlograma das mesmas (Figura 21). Verifica-se que há uma forte correlação entre as observações, mesmo

para atrasos (*lags*) de até 168 horas. É importante destacar que as observações com atraso de 12 horas apresentam correlação negativa, sugerindo um ponto de inflexão com inversão na tendência da série temporal.

Figura 21 – Correlograma dos dados de consumo de água.



Fonte: o Autor (2023)

5.3 Modelagem

Após o treinamento dos modelos, foi obtida uma lista em ordem decrescente de desempenho preditivo, considerando as métricas de avaliação estabelecidas em 4.4.1. A tabela 5 apresenta os resultados obtidos.

Os resultados da modelagem de série temporal com o uso do PyCaret indicaram que os modelos de regressão linear com condicionamento para remoção da sazonalidade e tendência apresentaram a melhor performance entre os 29 modelos avaliados. É importante destacar que esses modelos superaram tanto modelos clássicos, como ARIMA e Suavização Exponencial (*Exponential Smoothing*), quanto modelos mais sofisticados, como o Prophet - um *framework open source* desenvolvido pela equipe do Facebook (TAYLOR; LETHAM, 2018). Os resultados são apresentados na tabela 5.

O fato de que os melhores modelos apresentaram condicionamento para remoção da sazonalidade e tendência evidencia a forte influência desses dois atributos na série temporal em questão. O uso do condicionamento para remoção da sazonalidade e tendência pode ser aplicado para o preparo de séries temporais para análise e modelagem. As técnicas empregadas auxiliam na remoção de componentes indesejadas da série temporal, tornando mais simples a identificação de padrões e tendências significativos.

As técnicas de remoção da sazonalidade e tendência aplicadas pelo PyCaret são métodos automáticos que utilizam modelagem de tendência por meio de regressão poli-

Tabela 5 – Desempenho dos modelos treinados com o emprego da biblioteca PyCaret.

Nome do modelo implementado no PyCaret	MAE	RMSE
Huber w/ Cond. Deseasonalize & Detrending	0.882	1.132
Bayesian Ridge w/ Cond. Deseasonalize & Detrending	0.884	1.133
Ridge w/ Cond. Deseasonalize & Detrending	0.884	1.133
Least Angular Regressor w/ Cond. Deseasonalize & Detrending	0.884	1.133
Linear w/ Cond. Deseasonalize & Detrending	0.884	1.133
Gradient Boosting w/ Cond. Deseasonalize & Detrending	0.891	1.144
Light Gradient Boosting w/ Cond. Deseasonalize & Detrending	0.899	1.153
Extra Trees w/ Cond. Deseasonalize & Detrending	0.910	1.165
Random Forest w/ Cond. Deseasonalize & Detrending	0.912	1.166
Orthogonal Matching Pursuit w/ Cond. Deseasonalize & Detrending	0.923	1.174
Extreme Gradient Boosting w/ Cond. Deseasonalize & Detrending	0.929	1.186
K Neighbors w/ Cond. Deseasonalize & Detrending	0.998	1.263
Prophet	1.012	1.259
Seasonal Naive Forecaster	1.045	1.367
Elastic Net w/ Cond. Deseasonalize & Detrending	1.086	1.332
Theta Forecaster	1.110	1.402
ETS	1.132	1.454
Lasso w/ Cond. Deseasonalize & Detrending	1.153	1.401
Lasso Least Angular Regressor w/ Cond. Deseasonalize & Detrending	1.162	1.408
Exponential Smoothing	1.193	1.500
Decision Tree w/ Cond. Deseasonalize & Detrending	1.327	1.678
AdaBoost w/ Cond. Deseasonalize & Detrending	1.333	1.610
Croston	3.055	3.693
Polynomial Trend Forecaster	3.095	3.607
Grand Means Forecaster	3.128	3.609
Naive Forecaster	3.374	3.807
Passive Aggressive w/ Cond. Deseasonalize & Detrending	7.923	13.068
ARIMA	1.040	1.362
Naive Forecaster	3.374	3.807

Fonte: o Autor (2023).

nomial, regressão exponencial, suavização exponencial, entre outros. Para modelagem da sazonalidade, são empregadas técnicas como média móvel sazonal, regressão sazonal e modelo aditivo/multiplicativo. Além disso, o PyCaret realiza análises de resíduos para avaliar a qualidade da remoção de tendência e sazonalidade. Essas técnicas são úteis para remover componentes indesejadas da série temporal e facilitar a identificação de padrões e tendências significativas.

5.3.1 Escolha do melhor modelo

O modelo "Huber w/ Cond. Deseasonalize Detrending" obteve o melhor desempenho dentre os modelos treinados pelo PyCaret. Ele utiliza a função de perda de Huber e emprega técnicas de condicionamento para remover a sazonalidade e a tendência da série temporal.

A função de perda de Huber foi proposta pelo matemático suíço Berthold Huber em 1964 ((HUBER, 1964), como uma extensão da regressão linear. Em uma regressão linear comum, a função de perda usada é a soma dos erros quadráticos. Isso significa que a regressão é altamente sensível a valores discrepantes, o que pode levar a estimativas imprecisas dos parâmetros do modelo. Já na regressão de Huber, a função de perda é uma combinação ponderada de perdas absolutas e perdas quadráticas. A ponderação dos termos é determinada por um parâmetro delta, que controla o grau de sensibilidade a valores discrepantes.

De acordo com Huber (1981) o emprego da função de perda Huber é especialmente útil quando os dados apresentam valores discrepantes ou quando a suposição de normalidade dos erros da regressão não é satisfeita.

Portanto, optou-se pelo modelo "Huber w/ Cond. Deseasonalize Detrending" para efetuar as previsões e avaliações no conjunto de teste (observações de janeiro/2023).

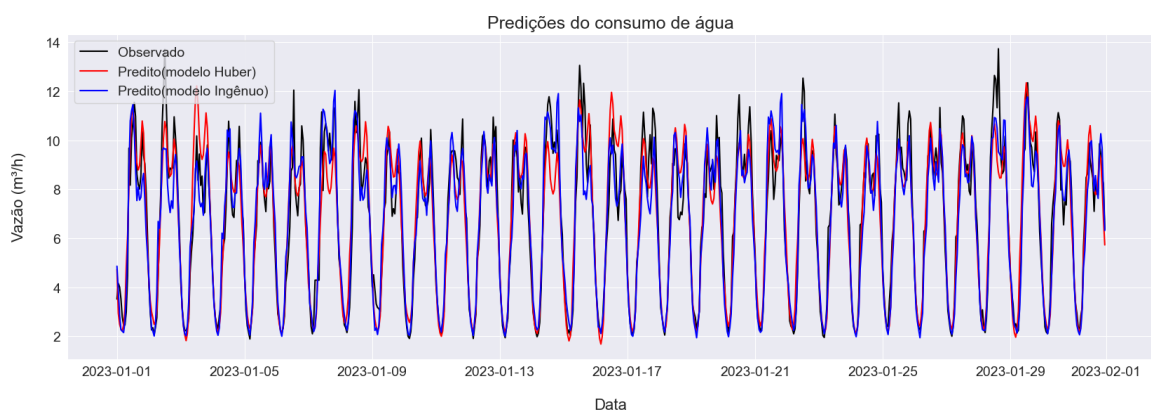
5.3.2 Teste e Avaliação do Modelo

Nesta etapa de teste, utilizou-se o modelo selecionado em 5.3.1 para realizar previsões de consumo para cada hora do dia entre 01/01/2023 e 30/01/2023.

Além disso, a fim de avaliar o desempenho preditivo do modelo, foram geradas previsões utilizando o modelo de referência (ingênuo).

Os valores observados para o consumo horário e as respectivas previsões realizadas são apresentados na figura 22.

Figura 22 – Resultado das previsões para o conjunto de teste.



Fonte: o Autor (2023)

As métricas de avaliação para os modelos são apresentadas na tabela 6. O modelo de regressão de Huber, apesar de ser mais robusto e sofisticado, apresentou desempenho

inferior ao modelo de referência (ingênuo), o que foi contrário às expectativas. Uma possível razão para esse resultado é o fato de o modelo ingênuo levar em consideração apenas observações das últimas quatro semanas em suas previsões. Isso sugere que, para previsões de curto prazo (24 horas), limitar o treinamento às observações mais recentes pode ser benéfico para o desempenho do modelo.

Tabela 6 – Métricas de desempenho apuradas para as previsões realizadas.

Modelo	MAE	RMSE
Huber	0.728	0.995
Ingênuo	0.685	0.948

Fonte: o Autor (2023).

Para dar continuidade à análise do desempenho dos modelos, optou-se por realizar novas previsões para janeiro de 2023 utilizando o modelo de regressão de Huber. No entanto, neste experimento, o modelo foi treinado apenas com os dados do mês anterior (dezembro de 2022). Os resultados obtidos estão apresentados na figura 24.

Apesar de o modelo treinado apenas com os dados do mês anterior ter apresentado melhores resultados do que o modelo treinado com todo o conjunto de treinamento, a comparação entre o modelo Huber e o modelo de referência (ingênuo) revela que o primeiro não apresentou uma vantagem significativa em termos de desempenho preditivo, conforme apresentado na tabela 7.

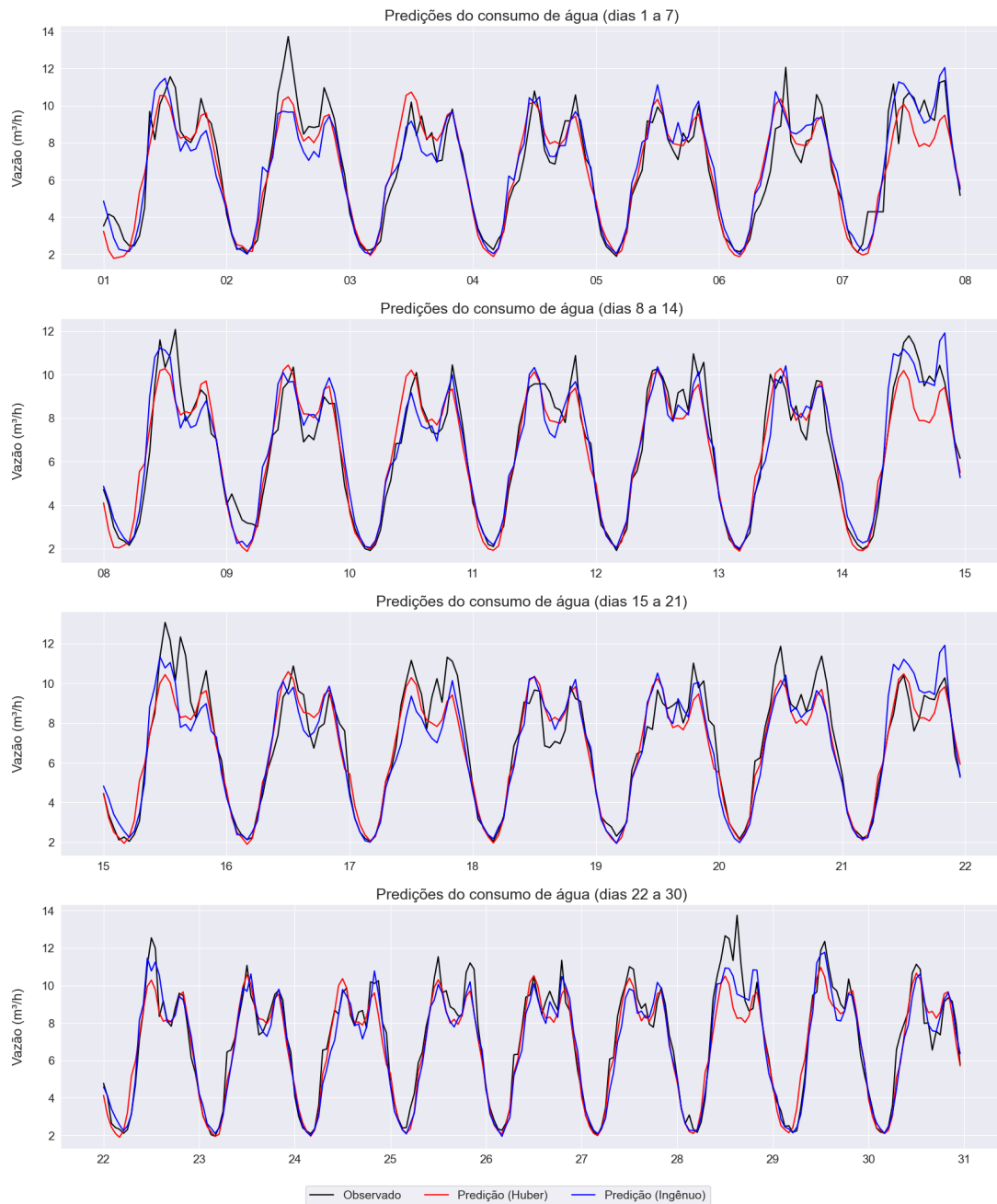
Tabela 7 – Métricas de desempenho para as previsões finais.

Modelo	MAE	RMSE
Huber	0.679	0.943
Ingênuo	0.685	0.948

Fonte: o Autor (2023).

Ao se comparar os resultados de um método de previsão com um modelo de referência, como o método ingênuo, é possível avaliar se o método proposto é de fato eficaz e traz ganhos em relação a um modelo mais simples. Nesse sentido, o modelo ingênuo desenvolvido neste estudo, além de apresentar resultados semelhantes ao melhor método obtido pela biblioteca PyCaret, apresenta simplicidade na implementação e baixa exigência em termos de quantidade de dados. Sua simplicidade pode torná-lo uma alternativa viável para a previsão do consumo de água, facilitando a tomada de decisões estratégicas para a operação otimizada de sistemas de distribuição de água.

Figura 23 – Predições finais para o conjunto de teste.



Fonte: o Autor (2023)

5.4 Estratégia de operação otimizada

Foi proposta a aplicação prática do modelo ingênuo para predição de consumo de água como meio de integrar uma metodologia de otimização do funcionamento de bombas em um sistema de distribuição de água, conforme descrito em 4.5. A implementação dessa abordagem específica para o sistema em estudo foi realizada através do uso da linguagem de programação Python.

Com o objetivo de avaliar a eficácia da estratégia operacional em relação a estratégia atualmente existente, a estratégia de otimização foi aplicada retroativamente (*backtesting*) às observações registradas durante o mês de janeiro. Foram realizadas buscas para as soluções ótimas em relação aos dois possíveis tipos de tarifação elétrica: tarifa convencional e tarifa branca. Essa análise permitiu avaliar como a estratégia operacional teria se comportado em cada um desses cenários, caso tivesse sido aplicada durante o período em questão.

Além disso, o procedimento foi repetido com a busca das decisões ótimas de bombeamento considerando a substituição do modelo ingênuo por um modelo ideal, teórico, que apresentasse uma acurácia de 100% nas predições. Dessa forma, tornou-se possível analisar a sensibilidade da estratégia em relação ao desempenho do modelo preditivo de consumo que é utilizado para integrá-la.

Com a aplicação da estratégia operacional conforme detalhado, e considerando como "cenário 0" o regime de funcionamento das bombas efetivamente observado para o período analisado, determinou-se o regime de funcionamento ótimo das bombas para cada um dos quatro possíveis cenários descritos a seguir:

- Cenário 1: Modelo ingênuo e tarifa convencional;
- Cenário 2: Modelo ingênuo e tarifa branca;
- Cenário 3: Modelo ideal e tarifa convencional;
- Cenário 4: Modelo ideal e tarifa branca.

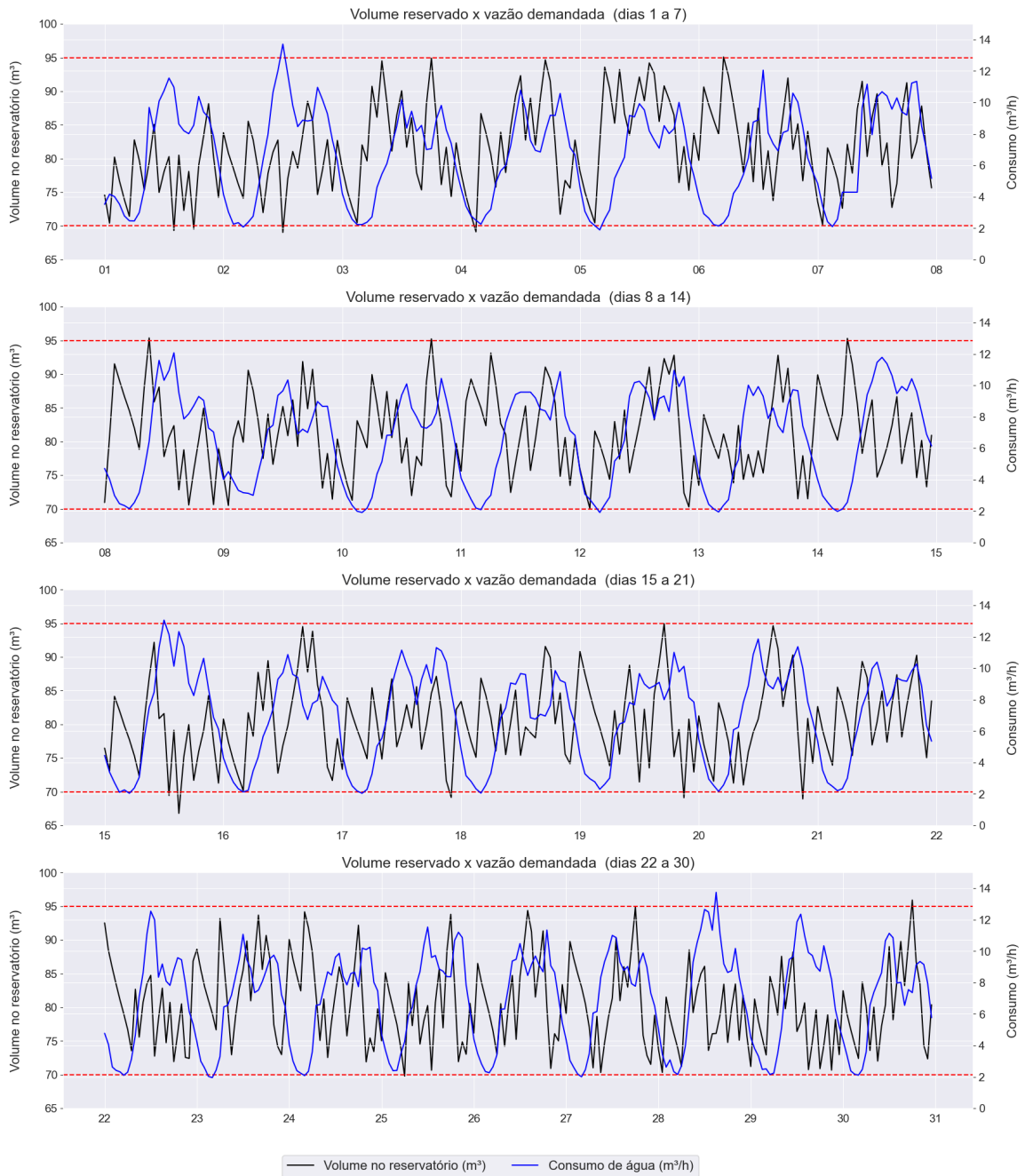
Com o intuito de ilustrar os resultados obtidos, a figura 24 apresenta a curva de variação do volume do reservatório simulada para o período de teste resultante da aplicação da estratégia de otimização proposta pelo Cenário 2. É possível perceber que o acionamento das bombas acompanha a tendência observada no consumo, buscando antecipar as ações necessárias para que o nível do reservatório esteja dentro dos limites estabelecidos (linhas vermelhas).

A figura 25 apresenta um recorte selecionado para visualizar com mais detalhes o funcionamento da estratégia aplicada, que abrange os dias 15/01/2023 (domingo) a 17/01/2023 (terça-feira). Nesse cenário onde aplicou-se a tarifa branca, é possível observar a influência da tarifa elétrica na tomada de decisão de bombeamento, onde as ações de bombeamento durante os horários intermediários e de pico (18:00 h às 23:00 h) foram limitadas ao menor custo de bombeamento possível, a fim de manter os limites do reservatório dentro das faixas admissíveis.

Em relação às restrições impostas para os volumes no reservatório, verifica-se nas figuras 24 e 25 alguns momentos específicos onde ocorreram violações. Nesse sentido, foi feita a verificação das violações nesses limites para cada cenário simulado. O resultado dessa análise é apresentado na tabela 8.

No que se refere às limitações de volume impostas, foram verificadas possíveis

Figura 24 – Resultado da aplicação da estratégia de otimização (cenário 2) aos dados observados em janeiro de 2023.

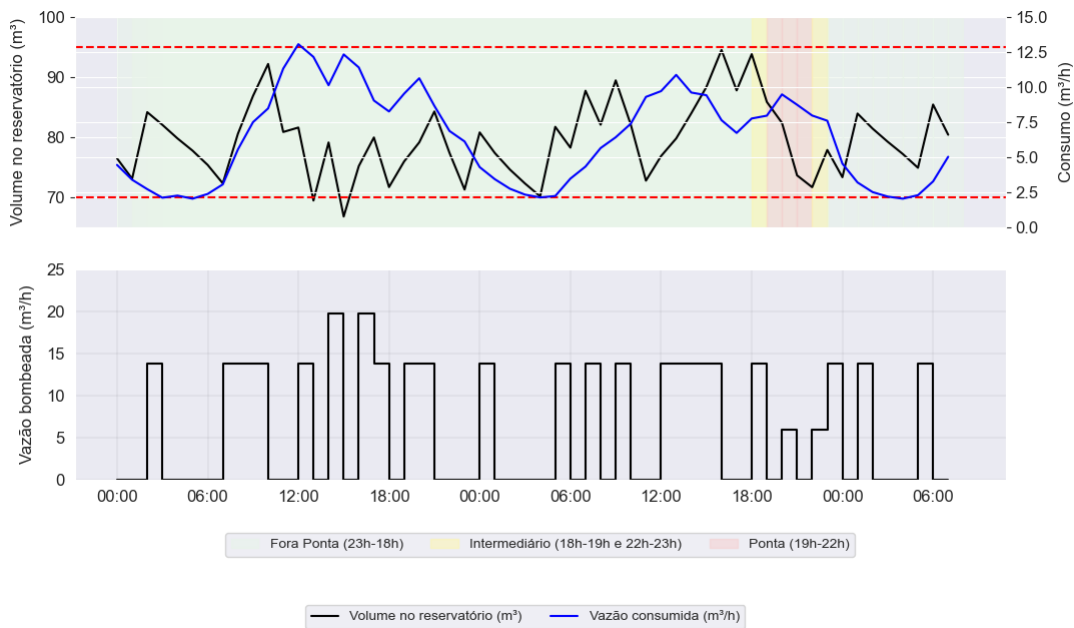


Fonte: o Autor (2023)

violações em todos os cenários, principalmente devido ao uso de modelos preditivos sujeitos a erros para a tomada de decisão de acionamento ou não das bombas. Em todos os casos, houve violação da condição de volume mínimo.

Além disso, mesmo nos cenários que utilizaram modelos preditivos teóricos (cenários 3 e 4), ocorreram violações das condições de otimização. Isso pode ter sido devido ao

Figura 25 – Resultado da aplicação da estratégia de otimização (cenário 2) para os dias 15 e 17 de janeiro.



Fonte: o Autor (2023)

Tabela 8 – Violações aos limites de reservação máximo e mínimo admissíveis.

Cenário	Violação ao V_{min} ($70m^3$)		Violação ao V_{max} ($95 m^3$)	
	Número de ocorrências	Menor volume observado (m^3)	Número de ocorrências	Maior volume observado (m^3)
Cenário 0	24	40.00	28	97.00
Cenário 1	13	67.93	0	94.99
Cenário 2	10	66.79	5	95.91
Cenário 3	2	67.63	0	94.99
Cenário 4	4	68.08	0	94.99

Fonte: o Autor (2023).

tempo limitado de busca pela solução ótima, que foi definido em até 5 minutos no processo de otimização. Dessa forma, algumas das ações de bombeamento implementadas podem não ter sido as melhores soluções possíveis para aquele instante.

Em relação aos estados de funcionamento das bombas selecionados, a tabela 9 apresenta os resultados referentes ao tempo total de funcionamento para cada estado e, conseqüentemente, o consumo de energia envolvido.

Observa-se que, de acordo com a estratégia de automação atual (cenário 0), que se baseia apenas nos níveis do reservatório para acionar as bombas, ambas as bombas são sempre acionadas em conjunto, tendo resultando em um consumo energético de 2560 kWh para o mês de janeiro de 2023. Em contrapartida, nos cenários otimizados, o acionamento das bombas seria realizado de maneira estratégica, resultando em reduções de cerca de

Tabela 9 – Tempo de funcionamento das bombas, por estado, e consumo energético total correspondente.

Estado de Funcionamento	Cenário 0	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Duas bombas desligadas	209 h	367 h	345 h	361 h	353 h
Apenas bomba n ^o 1 ligada	0 h	15 h	52 h	18 h	32 h
Apenas bomba n ^o 2 ligada	0 h	311 h	299 h	326 h	320 h
Duas bombas ligadas	256 h	27 h	24 h	15 h	15 h
Consumo energético (kWh):	2560	2255.2	2291.4	2239.6	2255.6

Fonte: o Autor (2023).

12% no consumo de energia.

Ainda ao observar a tabela 9, é possível notar que há uma maior prevalência do estado de funcionamento em que apenas a bomba n^o2 é acionada. Isso provavelmente ocorre devido à diferença entre o "custo-benefício" de acionamento de cada bomba, que pode ser quantificado pelo consumo específico de energia (kWh/m³). Enquanto a bomba n^o1 consome 3,8 kW para fornecer 6 m³/h (consumo específico de 0,63 kWh/m³), a bomba n^o2 apresenta um consumo específico de 0,45 kWh/m³ (consumo de 6,2 kW para uma vazão de 13,8 m³/h). A maior prevalência de acionamento da bomba n^o1 nos cenários 2 e 4 (tarifa branca), somada à análise realizada com base na figura 25, sugere que essa escolha de bombeamento ocorre em momentos em que é vantajoso optar por um menor consumo energético pontual, característico dos horários de pico na modalidade de tarifa branca.

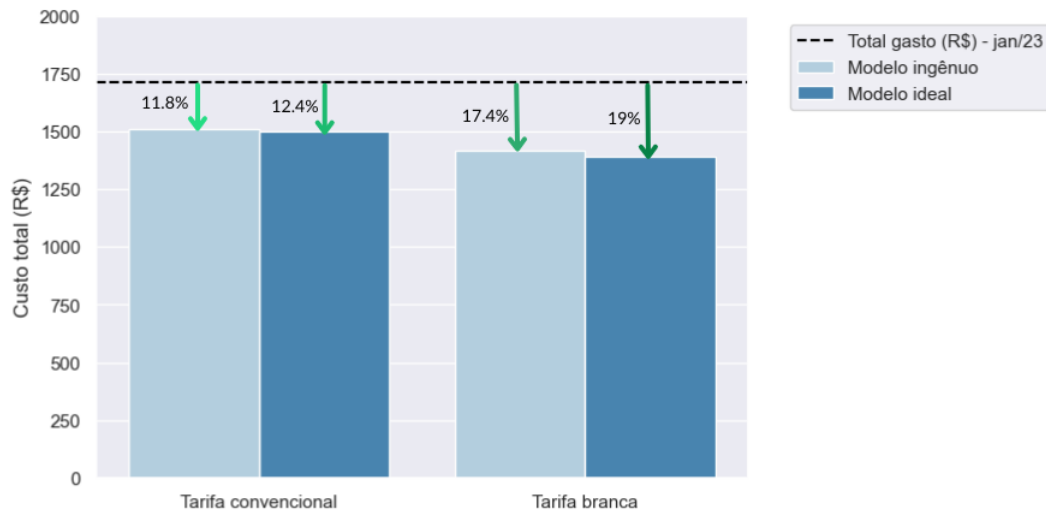
Por fim, uma vez que o objetivo proposto na otimização foi minimizar o custo decorrente do funcionamento das bombas, a figura 26 apresenta a comparação entre os custos totais de energia para o mês de janeiro para cada cenário, destacando, sobretudo, a diferença em relação ao cenário 0 (total gasto em jan/23), que representa a situação efetivamente ocorrida.

A análise dos resultados apresentados na figura 26 indicam que todos os cenários simulados foram capazes de reduzir os custos, variando entre cerca de 12% a 19% em relação aos valores observados no período. Verifica-se que a modalidade tarifária Tarifa Branca apresentou maior potencial de ganhos, com redução de custos em torno de 19%.

Quanto à influência do modelo preditivo nos resultados obtidos, observou-se que não houve grandes diferenças entre os custos para os dois casos (modelo ingênuo e modelo ideal), independentemente do tipo de tarifa aplicado. Assim, conclui-se que o modelo ingênuo é apropriado para a estratégia proposta, dada sua maior facilidade de implementação e bons resultados na otimização de custos de consumo de energia elétrica.

Finalmente, ao se confrontar os resultados obtidos para o consumo energético (kWh), apresentados na 9, com os resultados para os custos correspondentes (R\$), verifica-se que as reduções obtidas foram proporcionais (cerca de 12%) para cenários com modalidade de

Figura 26 – Comparação entre os custos finais obtidos em cada cenário.



Fonte: o Autor (2023)

tarifa convencional. Para cenários com modalidade de tarifa branca, enquanto houve uma redução de até 19% do custo, a redução equivalente em consumo energético foi na casa dos 12%.

6 CONCLUSÃO E CONSIDERAÇÕES FINAIS

Ao longo deste trabalho, foi possível explorar e desenvolver soluções para o desenvolvimento de modelos preditivos de consumo de água, bem como para a otimização do abastecimento de água em sistemas de distribuição. A aplicação de modelos preditivos para previsão do consumo de água a curto prazo, juntamente com a estratégia operacional automatizada de tomada de decisão de acionamento de bombas, mostrou-se eficaz para a redução do consumo de energia elétrica e, conseqüentemente, do custo operacional.

No desenvolvimento do trabalho, destacam-se o uso do PyCaret e Pyomo, bibliotecas do Python que simplificaram o processo de modelagem, garantindo o sucesso obtido. O PyCaret se mostrou uma ferramenta essencial, permitindo a comparação de vários modelos de machine learning e a escolha do mais adequado para o problema em questão. O Pyomo, por sua vez, permitiu a obtenção de soluções precisas e eficientes para a estratégia de otimização proposta.

É importante destacar que, embora o modelo preditivo de Huber com condicionamento para remoção de sazonalidade e tendência tenha apresentado melhor desempenho entre os modelos avaliados por meio do PyCaret, não foi observada uma diferença significativa em relação ao modelo ingênuo proposto neste estudo, que foi escolhido como o mais apropriado para ser utilizado na estratégia de otimização.

O modelo ingênuo, escolhido como o mais apropriado para a estratégia de otimização, possibilitou uma redução expressiva do consumo de energia elétrica. A aplicação desse modelo indicou uma significativa diminuição nos custos, variando entre 12% para a tarifa convencional e 18% para a tarifa branca, o que apresenta resultados promissores se comparados aos da startup portuguesa SCUBIC, que obteve uma redução de 19% no gasto de energia elétrica em um sistema de abastecimento no estado do Paraná, por meio da aplicação de sua solução comercial de inteligência artificial e otimização (SCUBIC, 2023).

Assim, os resultados obtidos permitiram constatar que a estratégia operacional automatizada apresenta vantagens em relação à estratégia atualmente em uso, com reduções significativas de consumo de energia elétrica, principalmente em cenários de tarifa branca. A aplicação de modelos preditivos, por sua vez, mostrou-se importante para aprimorar a precisão da previsão de consumo de água, mas não se mostrou essencial para a obtenção de resultados satisfatórios de redução do consumo de energia elétrica.

Como propostas de melhorias e trabalhos futuros, sugere-se a inclusão da minimização do consumo energético como uma das condições da otimização, uma vez que, para alguns casos quando empregada a tarifa branca, pode ocorrer de o consumo não ser reduzido mesmo com a redução do custo total. Além disso, propõe-se a inclusão do número

de acionamentos/desligamentos das bombas como restrição, a fim de aumentar a vida útil das mesmas.

Conclui-se, portanto, que a metodologia empregada neste trabalho se mostrou eficaz na previsão do consumo de água e na otimização do consumo energético e, consequentemente, do custo de operação de sistemas de distribuição de água. Os resultados apresentados evidenciam a importância do uso de modelos preditivos para a tomada de decisão e a necessidade de se considerar não apenas o custo financeiro, mas também o impacto ambiental. Espera-se que este trabalho possa contribuir para o desenvolvimento de estratégias operacionais mais sustentáveis e eficientes para a gestão de sistemas de distribuição de água.

REFERÊNCIAS

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA - ANEEL. **Tarifa Branca**. 2023. Url<https://www.gov.br/aneel/pt-br/assuntos/tarifas/entenda-a-tarifa/postos-tarifarios>.
- ALI, M. **PyCaret: An open source, low-code machine learning library in Python**. [S.l.], 2020. PyCaret version 1.0.0. Disponível em: <<https://www.pycaret.org>>.
- BATA, M.; CARRIVEAU, R.; TING, D. S.-K. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. **Smart Water**, Springer, v. 5, n. 1, p. 1–18, 2020.
- BENTLEY. **Bentley WaterGEMS CONNECT Edition Help**. 2023. <<https://docs.bentley.com/LiveContent/web/Bentley%20WaterGEMS%20SS6-v1/en/GUID-965C64F3-C894-4BAB-BDD1-0DE35F277ACE.html>>. Acesso em: 18 de fevereiro de 2023.
- BICH-NGOC, N.; TELLER, J. A review of residential water consumption determinants. In: SPRINGER. **International Conference on Computational Science and Its Applications**. [S.l.], 2018. p. 685–696.
- BRASIL. Sistema nacional de informações sobre saneamento: 25º diagnóstico dos serviços de Água e esgotos – 2019. In: _____. Brasília: Ministério do Desenvolvimento Regional. Secretaria Nacional de Saneamento – SNS, 2020. p. 183.
- CANDELIERI, A. Clustering and support vector regression for water demand forecasting and anomaly detection. **Water**, MDPI, v. 9, n. 3, p. 224, 2017.
- COIN-OR FOUNDATION. **COIN-OR: Computational Infrastructure for Operations Research**. <<https://www.coin-or.org>>. Acesso em: 12 de fevereiro de 2023.
- CORNUÉJOLS, G.; LIU, L. **Branch and Cut: Studies in Integer Programming**. [S.l.]: Springer, 2001.
- DONKOR, E. A. et al. Urban water demand forecasting: review of methods and models. **Journal of Water Resources Planning and Management**, American Society of Civil Engineers, v. 140, n. 2, p. 146–159, 2014.
- EHLERS, R. Apostila análise de séries temporais, 5ª edição. **UFPR, Curitiba, 94p**, 2009.
- Equatorial Energia Goiás. **Postos Tarifários em Goiás**. 2023. Url<https://go.equatorialenergia.com.br/consumo-na-tarifa-branca/>.
- GAGLIARDI, F. et al. A probabilistic short-term water demand forecasting model based on the markov chain. **Water**, MDPI, v. 9, n. 7, p. 507, 2017.
- GROPPO, G. de S.; COSTA, M. A.; LIBÂNIO, M. Predicting water demand: A review of the methods employed and future possibilities. **Water Supply**, IWA Publishing, v. 19, n. 8, p. 2179–2198, 2019.

HART, W. et al. **Pyomo: Modeling and Solving Mathematical Programs in Python**. 2020. <<https://pyomo.readthedocs.io/>>. Disponível em: <<https://pyomo.readthedocs.io/>>.

HU, S. et al. An innovative hourly water demand forecasting preprocessing framework with local outlier correction and adaptive decomposition techniques. **Water**, MDPI, v. 13, n. 5, p. 582, 2021.

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

_____. **Robust Statistics: Theory and Methods**. [S.l.]: Wiley, 1981.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [S.l.]: OTexts, 2021.

KOO, K.-M. et al. Performance assessment for short-term water demand forecasting models on distinctive water uses in korea. **Sustainability**, MDPI, v. 13, n. 11, p. 6056, 2021.

LEE, J.-Y. et al. Future global climate: Scenario-based projections and near-term information. In: _____. **Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change**. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021. p. 553–672.

MOREIRA, P. H. V. et al. Aplicação de séries temporais na previsão da média mensal da taxa de câmbio do real para o dólar comercial de compra usando o modelo de holt. **XXX Encontro Nacional de Engenharia de Produção**. São Carlos-SP, 2010.

MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais: modelos lineares univariados**. [S.l.]: Editora Blucher, 2018.

MUHAMMAD, A. U.; LI, X.; FENG, J. Artificial intelligence approaches for urban water demand forecasting: A review. In: SPRINGER. **International Conference on Machine Learning and Intelligent Communications**. [S.l.], 2019. p. 595–622.

NIKNAM, A. et al. A critical review of short-term water demand forecasting tools—what method should i use? **Sustainability**, MDPI, v. 14, n. 9, p. 5412, 2022.

PACCHIN, E. et al. A comparison of short-term water demand forecasting models. **Water resources management**, Springer, v. 33, n. 4, p. 1481–1497, 2019.

PANDEY, P. et al. Hybrid models for water demand forecasting. **Journal of Water Resources Planning and Management**, American Society of Civil Engineers, v. 147, n. 2, p. 04020106, 2021.

PONTE, B. et al. Intelligent decision support system for real-time water demand management. **International Journal of Computational Intelligence Systems**, Taylor & Francis, v. 9, n. 1, p. 168–183, 2016.

RANGEL, H. R. et al. Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks. **Journal of Hydroinformatics**, IWA Publishing, v. 19, n. 1, p. 1–16, 2017.

SALLOOM, T.; KAYNAK, O.; HE, W. A novel deep neural network architecture for real-time water demand forecasting. **Journal of Hydrology**, Elsevier, v. 599, p. 126353, 2021.

SALOMONS, E.; HOUSH, M. Practical real-time optimization for energy efficient water distribution systems operation. **Journal of Cleaner Production**, Elsevier, v. 275, p. 124148, 2020.

SARDINHA-LOURENÇO, A. et al. Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy. **Journal of Hydrology**, Elsevier, v. 558, p. 392–404, 2018.

SCUBIC. **SCUBIC - Estudo de Caso: Plataforma Inteligente para Otimização e Eficiência Energética de Sistemas de Abastecimento de Água - SANEPAR - SISTEMA PASSAÚNA**. 2023. <https://scubic.tech/wp-content/uploads/2023/01/Case_Study_Sanepar_pt.pdf>. Acesso em: 18 de fevereiro de 2023.

SHABANI, S. et al. Gene expression programming coupled with unsupervised learning: a two-stage learning process in multi-scale, short-term water demand forecasts. **Water**, MDPI, v. 10, n. 2, p. 142, 2018.

SHIRKOOHI, M. G.; DOGHRI, M.; DUCHESNE, S. Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. **Water Supply**, IWA Publishing, v. 21, n. 5, p. 2374–2386, 2021.

SUEZ. **Real-time energy management**. 2023. <<https://www.suez.com/en/uk/water-network-management/real-time-management-and-optimisation/real-time-energy-management>>. Acesso em: 18 de fevereiro de 2023.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **The American Statistician**, Taylor & Francis, v. 72, n. 1, p. 37–45, 2018.

YOUSEFI, P. et al. Nonlinear dynamic modeling of urban water consumption using chaotic approach (case study: City of kelowna). **Water**, MDPI, v. 12, n. 3, p. 753, 2020.