

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Análise de séries temporais climáticas para estudo
da recarga do Aquífero Guarani

André Canali Garcia

Monografia - MBA em Ciência de Dados (CEMEAI)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

André Canali Garcia

Análise de séries temporais climáticas para estudo da recarga do Aquífero Guarani

Monografia apresentada ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Ciências de Dados.

Área de concentração: Ciências de Dados

Orientador: Prof. Dr. Afonso Paiva

Versão original

São Carlos
2026

É possível elaborar a ficha catalográfica em LaTeX ou incluir a fornecida pela Biblioteca. Para tanto observe a programação contida nos arquivos USPSC-modelo.tex e fichacatalografica.tex e/ou gere o arquivo fichacatalografica.pdf.

A biblioteca da sua Unidade lhe fornecerá um arquivo PDF com a ficha catalográfica definitiva, que deverá ser salvo como fichacatalografica.pdf no diretório do seu projeto.

André Canali Garcia

Analysis of climate time series for the study of Guarani Aquifer recharge.

Monograph presented to the Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Data Science.

Concentration area: Data Science

Advisor: Prof. Dr. Afonso Paiva

Original version

**São Carlos
2026**

Folha de aprovação em conformidade
com o padrão definido
pela Unidade.

No presente modelo consta como
folhadeaprovacao.pdf

RESUMO

GARCIA, A.C. Análise de séries temporais climáticas para estudo da recarga do Aquífero Guarani. 2026. 60 p. Monografia (MBA em Ciências de Dados) - Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

Este trabalho apresenta a análise e a modelagem de séries temporais climáticas associadas à região de recarga do Sistema Aquífero Guarani, com foco nas variáveis de precipitação e temperatura. Foram utilizados dados observacionais da estação meteorológica do CRHEA, bem como dados de reanálise ERA5, empregados de forma criteriosa para o preenchimento de falhas nas séries temporais. O estudo contemplou etapas de pré-processamento, incluindo análise exploratória, tratamento de inconsistências, dessazonalização e remoção de tendência. Para a modelagem preditiva, foram comparadas abordagens estatísticas clássicas e métodos baseados em aprendizado de máquina, com avaliação de desempenho por meio das métricas MAE e RMSE. Os resultados indicaram melhor desempenho da abordagem baseada em aprendizado de máquina, embora os erros obtidos reflitam a elevada variabilidade inerente aos processos climáticos, sendo compatíveis com análises de caráter regional. Como desdobramento aplicado dos resultados, foi desenvolvida uma plataforma web para organização, visualização e consulta dos dados e previsões geradas, visando ampliar o potencial de aplicação das informações em estudos ambientais e no suporte à gestão de recursos hídricos.

Palavras-chave: Séries temporais; Climatologia; Aprendizado de máquina; Sistema Aquífero Guarani.

ABSTRACT

GARCIA, A.C. **Analysis of climate time series for the study of Guarani Aquifer recharge.**. 2026. 60 p. Monograph (MBA in Data Sciences) - Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2026.

This work presents the analysis and modeling of climatic time series associated with the recharge region of the Guarani Aquifer System, focusing on precipitation and temperature variables. Observational data from the CRHEA meteorological station were used, together with ERA5 reanalysis data, which were applied under objective criteria to fill gaps in the time series. The study included preprocessing steps such as exploratory analysis, inconsistency treatment, deseasonalization, and trend removal. For predictive modeling, classical statistical approaches were compared with machine learning-based methods, and performance was evaluated using the MAE and RMSE metrics. The results showed superior performance of the machine learning-based approach, although the obtained errors reflect the high variability inherent to climatic processes, being suitable for regional-scale analyses. As an applied outcome of the results, a web platform was developed to organize, visualize, and provide access to the data and generated forecasts, enhancing their potential use in environmental studies and water resources management.

Keywords: Time series; Climatology; Machine learning; Guarani Aquifer System.

LISTA DE FIGURAS

Figura 1 – Mapa com localização do SAG na América do Sul	21
Figura 2 – Localização da bacia do Ribeirão da Onça, Brotas - SP	22
Figura 3 – Estações meteorológicas utilizadas no estudo de Calixto (2020)	23
Figura 4 – Vista geral da estação climatológica do CRHEA	24
Figura 5 – Estação automática de coleta de dados climatológicos do CRHEA	24
Figura 6 – Gráfico de série temporal de temperatura	25
Figura 7 – Fluxo metodológico de coleta, tratamento, pré-processamento e mode- lagem dos dados meteorológicos.	33
Figura 8 – Série Temporal de Precipitação e Temperatura com Dados Ausentes . .	35
Figura 9 – Percentual de dados ausentes de precipitação por Ano e Mês	35
Figura 10 – Gráfico de Decomposição Sazonal	36
Figura 11 – Gráfico de Previsão de Precipitação utilizando o Modelo Extra Trees .	42
Figura 12 – Gráfico de Previsão de Precipitação utilizando o Modelo VAR	44
Figura 13 – Interface principal da plataforma web SGDHM	46
Figura 14 – Consulta e visualização de séries temporais no SGDHM	47
Figura 15 – Gerenciamento de metadados operacionais no SGDHM	47

LISTA DE TABELAS

Tabela 1 – Comparação entre trabalhos sobre séries temporais meteorológicas e hidrológicas	31
Tabela 2 – Métricas de desempenho dos modelos de previsão de precipitação	44

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	<i>AutoRegressive Integrated Moving Average</i>
BRO	Bacia do Ribeirão da Onça
CRHEA	Centro de Recursos Hídricos e Ecologia Aplicada
CART	<i>Classification And Regression Trees</i> (Árvore de Classificação e Regressão)
BIC	Critério de Informação Bayesiano
EMAs	Estações Meteorológicas Automáticas
IAB	Instituto Arruda Botelho
INMET	Instituto Nacional de Meteorologia
IDW	<i>Inverse Distance Weighting</i>
KNN	<i>K-Nearest Neighbors</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
LSTM	<i>Long Short-Term Memory</i>
MAPE	<i>Mean Absolute Percentage Error</i> (Erro Absoluto Percentual Médio)
MAE	<i>Mean Absolute Error</i> (Erro Médio Absoluto)
VAR	Modelo de Vetores Autoregressivos
MLP	<i>Multilayer Perceptron</i>
RNAs	Redes Neurais Artificiais
RMSE	<i>Root Mean Squared Error</i> (Raiz do Erro Quadrático Médio)
SARIMA	<i>Seasonal Autoregressive Integrated Moving Average</i>
SAG	Sistema Aquífero Guarani
SNHT	<i>Standard Normal Homogeneity Test</i> (Teste de Homogeneidade Normal Padrão)
SVR	<i>Support Vector Regression</i>

ADF Teste de Dickey-Fuller Aumentado

USP Universidade de São Paulo

WTF *Water Table Fluctuation*

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Objetivos Gerais e Específicos	19
1.2	Estrutura do Trabalho	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Sistema Aquífero Guarani (SAG)	21
2.1.1	Bacia do Ribeirão da Onça (BRO)	22
2.2	Estações Meteorológicas	22
2.3	Séries Temporais	24
2.3.1	Características das Séries Meteorológicas	24
2.3.2	Modelos para Séries Temporais	25
2.3.3	Preparação dos Dados de Séries Temporais	26
2.4	Pycaret	26
2.5	Modelo Extra Trees	26
2.6	Modelo de Vetores Autoregressivos (VAR)	27
3	TRABALHOS RELACIONADOS	29
4	METODOLOGIA	33
4.1	Coleta de Dados	34
4.2	Tratamento dos Dados	34
4.2.1	Integração e Preenchimento de Falhas	34
4.3	Pré-processamento para Modelagem de Séries Temporais	36
4.4	Modelagem e Previsão de Séries Temporais com PyCaret	36
4.4.0.1	Modelo Extra Trees	37
4.5	Modelagem e Previsão com VAR	38
4.5.0.1	Pré-processamento e Estacionariedade	38
4.5.0.2	Especificação e Ajuste do Modelo VAR(p)	38
4.5.0.3	Previsão e Reversão da Diferenciação	38
4.5.0.4	Avaliação e Diagnóstico do Modelo	39
5	RESULTADOS	41
5.1	Métricas de avaliação	41
5.2	Resultados do modelo Extra Trees	41
5.3	Resultados do modelo VAR	43
5.4	Comparação entre os Modelos de Previsão	44
5.5	Plataforma Web dos Resultados	45

6	CONCLUSÃO	49
	REFERÊNCIAS	51
	APÊNDICES	53
	APÊNDICE A – CÓDIGO-FONTE EM PYTHON PARA ANÁLISE DE DADOS METEOROLÓGICOS	55
A.1	Introdução	55
A.2	Leitura e preparação dos dados	55
A.3	Tratamento de erros e valores faltantes	55
A.4	Análise exploratória e visualização	56
A.5	Modelagem com PyCaret	57
A.6	Teste de Estacionariedade	58
A.7	Modelagem com VAR	59
A.8	Considerações finais	60

1 INTRODUÇÃO

Diversas pesquisas têm sido realizadas em relação às mudanças climáticas, obtendo-se um grande volume de dados. Desde 2004, um grupo da Escola de Engenharia de São Carlos (USP) realiza a coleta sistemática de dados em estações meteorológicas e poços na Bacia do Ribeirão da Onça (BRO), em Brotas/SP. A bacia tem um papel fundamental na recarga do Sistema Aquífero Guarani (SAG) — uma das maiores reservas de água doce subterrânea do planeta (Wendland; Barreto; Gomes, 2007).

Essa base de dados é composta por variáveis como temperatura do ar, precipitação, umidade, radiação solar e vazão de poços. No entanto, esses dados estão organizados em arquivos textos e planilhas, o que dificulta análises automatizadas e a integração com modelos de previsão.

Esta pesquisa visa organizar e tratar os dados de três estações climatológicas (CRHEA, IAB e Monte Alegre), promovendo uma análise exploratória e descritiva. A partir disso, pretende-se aplicar modelos de séries temporais, detectar padrões e tendências e comparar com dados globais, tornando o processo analítico mais robusto e acessível. O objetivo final é gerar subsídios científicos para decisões ambientais e políticas públicas.

1.1 Objetivos Gerais e Específicos

Objetivo geral deste trabalho é desenvolver uma análise robusta e automatizada de dados meteorológicos da BRO, aplicando e analisando modelos descritos através de séries temporais para subsidiar estudos de recarga do SAG por meio da análise e previsão de séries climáticas.

Objetivos específicos:

- Realizar revisão bibliográfica sobre análise de séries temporais aplicadas a dados climáticos;
- Organizar, limpar e padronizar os dados das três estações meteorológicas;
- Identificar dados faltantes, outliers e inconsistências;
- Aplicar modelos estatísticos para previsão de precipitação e temperatura;
- Desenvolver uma plataforma web para visualização e disseminação dos dados analisados.

1.2 Estrutura do Trabalho

Este trabalho está dividido em cinco capítulos principais. Após esta introdução, o Capítulo 2 apresenta a fundamentação teórica, abordando os principais conceitos necessários para a compreensão do estudo e das metodologias utilizadas. O Capítulo 3 reúne os trabalhos relacionados, com foco em aplicações de análise de séries temporais em climatologia e hidrologia. Na sequência, o Capítulo 4 descreve a metodologia empregada no desenvolvimento da pesquisa. O Capítulo 5 apresenta e discute os resultados obtidos. Por fim, o Capítulo 6 expõe as conclusões do trabalho, bem como considerações finais e sugestões para trabalhos futuros.

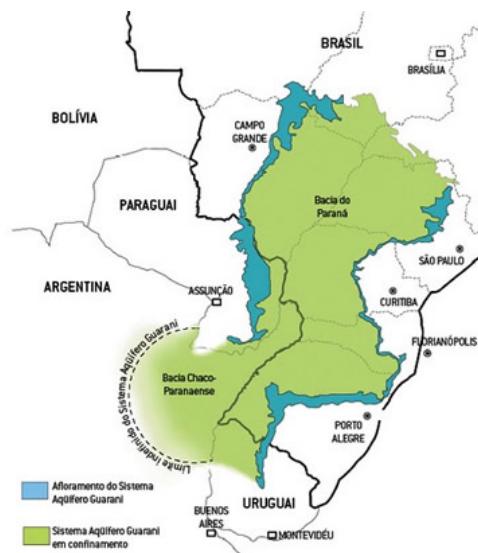
2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica aborda inicialmente os principais conceitos relacionados ao SAG, destacando sua relevância hidrogeológica e a importância da BRO como uma zona estratégica de recarga (Lucas, 2012; Calixto, 2020). Na sequência, discute-se o papel das estações meteorológicas na coleta de dados climatológicos fundamentais para a compreensão e modelagem de processos hidrológicos (Gomes, 2008; Calixto, 2020). Por fim, são apresentados os fundamentos de séries temporais, a biblioteca PyCaret — utilizada para o pré-processamento automático dos dados, treinamento e comparação de múltiplos algoritmos de aprendizado de máquina (Ali, 2020) — o modelo Extra Trees, que introduz um alto grau de aleatoriedade no processo de construção das árvores (Geurts; Ernst; Wehenkel, 2006) e o modelo de Vetores Autorregressivos (VAR), apropriado para a análise conjunta de séries temporais correlacionadas (Sims, 1980).

2.1 Sistema Aquífero Guarani (SAG)

O SAG (Figura 1) é uma das maiores reservas de água subterrânea do planeta, com aproximadamente 1,2 milhão de km² de extensão, abrangendo partes do Brasil, Argentina, Paraguai e Uruguai. No Brasil, cerca de 70% de sua área está localizada nos estados do Centro-Sul, com destaque para as regiões de afloramento das Formações Botucatu e Pirambóia, onde ocorre a principal recarga do aquífero (Lucas, 2012).

Figura 1 – Mapa com localização do SAG na América do Sul

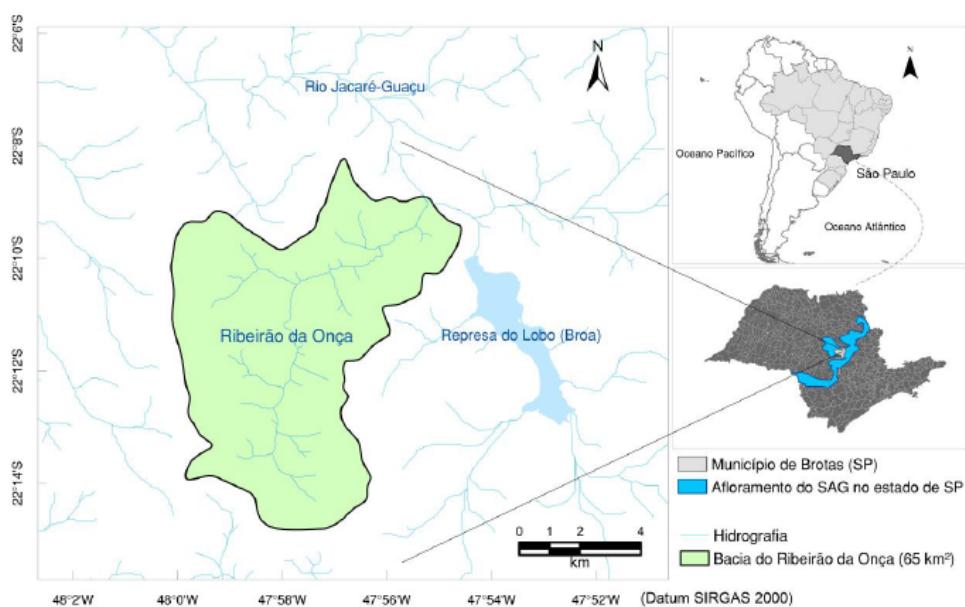


Fonte: ((SGB), 2025)

2.1.1 Bacia do Ribeirão da Onça (BRO)

A BRO (Figura 2) está localizada no município de Brotas, na região central do estado de São Paulo. Com uma área de 65 km², é um afluente da margem esquerda do rio Jacaré-Guaçu e caracteriza-se por ter uso predominantemente agrícola. Uma característica marcante da bacia é sua representatividade hidrogeológica, pois está situada em uma área de afloramento do SAG. O solo da bacia possui textura arenosa e alta capacidade de infiltração, o que faz com que grande parte da precipitação total se infiltrre, com uma contribuição de 10% a 15% para a recarga do aquífero freático (Calixto, 2020).

Figura 2 – Localização da bacia do Ribeirão da Onça, Brotas - SP



Fonte: (Calixto, 2020)

Os estudos sobre o SAG são essenciais diante das ameaças crescentes relacionadas à mudança do uso do solo, ao crescimento populacional e às incertezas climáticas. A gestão eficiente desse aquífero requer uma compreensão aprofundada dos mecanismos de recarga, das interações solo-vegetação-clima e do papel das zonas de afloramento como áreas-chave para a sustentabilidade dos recursos hídricos subterrâneos.

2.2 Estações Meteorológicas

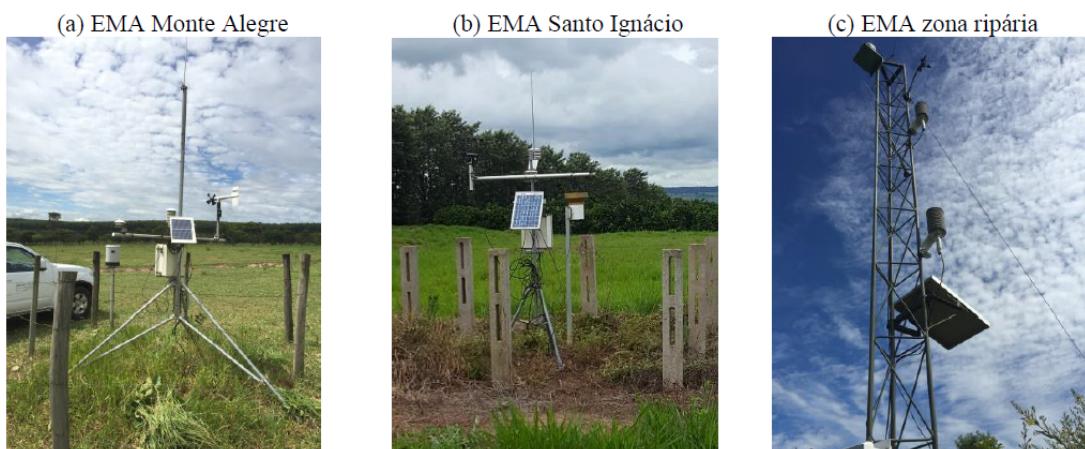
As estações meteorológicas desempenham um papel fundamental em estudos hidrológicos por fornecerem dados essenciais à quantificação dos componentes do ciclo hidrológico, como precipitação, temperatura, umidade, radiação solar e vento. Esses dados são insumos críticos para modelagens de balanço hídrico, estimativas de recarga de aquíferos, calibração de modelos conceituais e análise da variabilidade climática em bacias hidrográficas.

Na BRO, diferentes trabalhos utilizaram estações meteorológicas automáticas (EMAs) e convencionais para coleta de dados de precipitação e evapotranspiração. Gomes (2008) utilizou dados de precipitação diária obtidos por pluviômetros automáticos instalados na bacia e dados de nível de água em poços de monitoramento para aplicar o método *Water Table Fluctuation* (WTF) na estimativa da recarga subterrânea.

Lucas (2012) expandiu essa abordagem ao integrar dados de quatro estações localizadas dentro ou nas proximidades da bacia — entre elas, as do CRHEA-USP, Santo Inácio, Monte Alegre e Santa Maria. O autor realizou análises comparativas e de correlação entre as séries diárias e mensais dessas estações, identificando boa coerência espacial e temporal entre elas. Além disso, ele utilizou as informações meteorológicas para alimentar métodos locais (WTF e *Chloride Mass Balance*) e regionais (*Recession Curve Displacement Method*) de estimativa da recarga.

Calixto (2020) também utilizou extensas séries temporais de estações meteorológicas para estimar a precipitação média espacial sobre a bacia. Sua abordagem envolveu métodos de interpolação espacial como o IDW (*Inverse Distance Weighting*) para gerar superfícies contínuas de precipitação, além de dados de evapotranspiração de referência obtidos segundo a metodologia FAO-56. O estudo destacou a importância de estações confiáveis e bem distribuídas espacialmente para assegurar a acurácia de modelos hidrológicos e das estimativas de disponibilidade hídrica (Figura 3).

Figura 3 – Estações meteorológicas utilizadas no estudo de Calixto (2020)



Fonte: (Calixto, 2020)

A Estação Climatológica do CRHEA, localizada próxima à BRO, teve papel essencial no trabalho de Barreto (2006), fornecendo dados meteorológicos utilizados na análise do balanço hídrico da BRO, em zona de afloramento do SAG. Equipamentos como tanque Classe A, heliógrafo e estação automática registraram variáveis como precipitação e temperatura, fundamentais para estimar a evapotranspiração e a recarga do aquífero.

Esses dados permitiram avaliar métodos de estimativa de evapotranspiração e compreender os processos de recarga, destacando a importância do monitoramento contínuo realizado pela estação do CRHEA.

Figura 5 – Estação automática de coleta de dados climatológicos do CRHEA



Figura 4 – Vista geral da estação climatológica do CRHEA



Fonte: (Barreto, 2006)

Fonte: (Barreto, 2006)

Portanto, as estações meteorológicas são ferramentas indispensáveis para o monitoramento ambiental e a gestão hídrica, pois viabilizam a avaliação contínua dos recursos hídricos subterrâneos e superficiais, especialmente em áreas estratégicas como as zonas de recarga do SAG.

2.3 Séries Temporais

2.3.1 Características das Séries Meteorológicas

A análise de séries temporais é uma área fundamental da estatística e da ciência de dados, voltada ao estudo de fenômenos que evoluem ao longo do tempo. De acordo com Nielsen (2019), séries temporais podem ser definidas como sequências de observações ordenadas temporalmente, geralmente coletadas em intervalos regulares. A principal característica que distingue esse tipo de dado é a *dependência temporal* — isto é, os valores presentes em determinados instantes estão correlacionados com valores passados da própria série. Séries temporais têm ampla aplicação em áreas como meteorologia, finanças e medicina (Zou *et al.*, 2019).

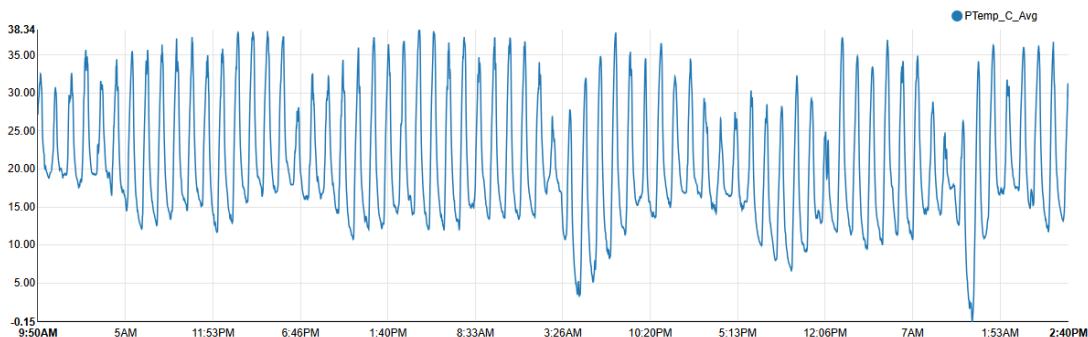
As séries temporais apresentam quatro componentes fundamentais: tendência, sazonalidade, ciclos e ruído. A tendência refere-se a um padrão de crescimento ou declínio persistente ao longo do tempo. A sazonalidade é caracterizada por flutuações regulares e

previsíveis em determinados períodos (por exemplo, variações climáticas entre estações do ano). Os ciclos dizem respeito a oscilações de longo prazo que, diferentemente da sazonalidade, não seguem necessariamente uma periodicidade regular. Por fim, o ruído é a variação aleatória, não explicada por nenhum dos outros componentes, que compõe a parte imprevisível da série (Nielsen, 2019).

No contexto da meteorologia, a análise de séries temporais é uma ferramenta fundamental, devido à natureza sequencial e dependente no tempo desses dados, como temperatura, precipitação, umidade e pressão atmosférica. Séries temporais de dados meteorológicos são essenciais para a compreensão das mudanças climáticas e o desenvolvimento de previsões meteorológicas (Zou *et al.*, 2019).

A Figura 6 mostra um exemplo de uma série temporal de temperatura da estação meteorológica do CRHEA.

Figura 6 – Gráfico de série temporal de temperatura



Fonte: Estação CRHEA

2.3.2 Modelos para Séries Temporais

Em relação à modelagem, a utilização de séries temporais é uma técnica valiosa para detectar tendências, remover sazonalidade e capturar padrões históricos nas variáveis hidrológicas (Caminha *et al.*, 2025). Entre os modelos estatísticos clássicos, destacam-se os modelos AR (Auto Regressivo), MA (Média Móvel), ARIMA (*AutoRegressive Integrated Moving Average*) e SARIMA (*Seasonal ARIMA*), amplamente utilizados para previsão de curto prazo em séries estacionárias ou com estrutura de tendência e sazonalidade comuns em variáveis ambientais (Nielsen, 2019; Caminha *et al.*, 2025).

No entanto, diante da crescente disponibilidade de dados e da complexidade dos fenômenos temporais, técnicas de aprendizado de máquina têm ganhado espaço. Métodos como regressão, redes neurais artificiais e florestas aleatórias são capazes de capturar relações não lineares entre as variáveis.

2.3.3 Preparação dos Dados de Séries Temporais

Outro ponto crítico e complexo para a eficácia da modelagem é a preparação dos dados temporais. Essa preparação envolve a identificação e remoção de *outliers*, imputação de valores ausentes, transformações (como logaritmo ou diferenciação) e normalização dos dados (Nielsen, 2019; Soares, 2022).

Séries que apresentam lacunas (falhas de medição) comprometem a análise, modelagem e a tomada de decisão baseada nos dados (Junior *et al.*, 2024). A ocorrência de falhas em séries temporais meteorológicas, como as de temperatura do ar ou precipitação, é comum devido a problemas técnicos ou ambientais nas estações de medição. Essas lacunas prejudicam a análise estatística e os modelos preditivos e, por isso, o preenchimento confiável dessas falhas é fundamental para manter a integridade dos dados (Junior *et al.*, 2024).

2.4 PyCaret

A PyCaret é uma biblioteca de código aberto desenvolvida em Python com o objetivo de simplificar e automatizar o ciclo de vida de projetos de *machine learning*. A ferramenta adota uma abordagem de alto nível (*low-code*), abstraindo a complexidade das etapas tradicionais do aprendizado de máquina e permitindo a construção, comparação e avaliação de modelos preditivos com poucas linhas de código. Segundo Ali (2020), a PyCaret foi concebida para aumentar a produtividade, a padronização e a reproduzibilidade dos experimentos, sendo aplicável tanto em contextos acadêmicos quanto profissionais.

De acordo com Ali (2020), a PyCaret organiza o fluxo de trabalho em etapas bem definidas, incluindo o pré-processamento automático dos dados, o treinamento e a comparação de múltiplos algoritmos, o ajuste de hiperparâmetros e a avaliação de desempenho. A biblioteca é estruturada em módulos especializados para diferentes tipos de problemas, como classificação, regressão, agrupamento, detecção de anomalias e análise de séries temporais, além de disponibilizar recursos para interpretação e reutilização de modelos treinados em aplicações práticas.

2.5 Modelo Extra Trees

O modelo *Extremely Randomized Trees* (Extra Trees) é um método de aprendizado de máquina baseado em conjuntos de árvores de decisão, proposto por Geurts, Ernst e Wehenkel (2006). Assim como o *Random Forest*, o Extra Trees constrói um conjunto de árvores independentes e combina suas previsões por meio de agregação, geralmente pela média, visando aumentar a robustez e a capacidade de generalização do modelo.

A principal característica do Extra Trees é a introdução de um alto grau de aleatoriedade no processo de construção das árvores, tanto na seleção das variáveis quanto

na definição dos pontos de corte, que são escolhidos de forma aleatória. Essa estratégia reduz a correlação entre as árvores do conjunto e contribui para a diminuição da variância do modelo, sendo particularmente eficaz em problemas com elevada complexidade e não linearidade.

Devido a essas propriedades, o modelo Extra Trees tem sido amplamente utilizado em aplicações envolvendo séries temporais ambientais e climáticas, nas quais a presença de padrões não lineares, sazonalidade e variabilidade natural elevada impõe desafios aos modelos estatísticos tradicionais (Geurts; Ernst; Wehenkel, 2006).

2.6 Modelo de Vetores Autoregressivos (VAR)

O modelo VAR constitui uma das abordagens mais influentes para a análise de séries temporais multivariadas. Proposto originalmente por Sims (1980), o VAR surgiu como uma alternativa aos modelos econométricos estruturais tradicionais, que dependiam fortemente de restrições teóricas e de hipóteses causais muitas vezes arbitrárias. A principal inovação de Sims foi tratar todas as variáveis como endógenas, permitindo que cada uma fosse explicada por defasagens de si mesma e das demais variáveis do sistema, o que possibilita capturar de forma mais realista as interdependências dinâmicas entre os fenômenos analisados.

De acordo com Lutkepohl (2005), o modelo VAR oferece uma estrutura geral e flexível para a modelagem de sistemas de séries temporais, servindo de base para diversas extensões, como os modelos de Correção de Erros Vetoriais e os VAR Estruturais. Já Hamilton (1994) destaca que o VAR é amplamente utilizado em previsões conjuntas e em análises de relações dinâmicas entre variáveis correlacionadas, sendo particularmente útil em campos como economia, finanças, meteorologia e ciências ambientais.

No contexto da Ciência de Dados, o modelo VAR assume papel relevante na análise preditiva de sistemas multivariados, permitindo investigar conjuntamente séries temporais correlacionadas — como temperatura, precipitação, pressão atmosférica e umidade. Essa capacidade torna o VAR uma ferramenta poderosa para compreender padrões dinâmicos e desenvolver modelos preditivos baseados em interações temporais, contribuindo para a tomada de decisão orientada por dados.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta uma revisão da literatura sobre modelos e algoritmos desenvolvidos para a análise de dados de séries temporais, com foco em aplicações em climatologia e hidrologia. São comparadas diferentes abordagens e reunidas contribuições acadêmicas relevantes relacionadas à previsão e análise de tendências em dados meteorológicos e hidrológicos.

No primeiro trabalho analisado, Kratzert *et al.* (2018) aplicaram redes neurais LSTM (*Long Short-Term Memory*) para modelagem chuva–vazão em 241 bacias hidrográficas dos Estados Unidos. Os modelos foram treinados com dados observacionais, sem incorporar conhecimento físico explícito. Os resultados mostraram que as LSTM superaram modelos hidrológicos tradicionais, com bom desempenho em diferentes regiões e condições. O estudo destaca o potencial do aprendizado de máquina como alternativa eficiente para simulações hidrológicas.

Calixto (2020) utilizou dados meteorológicos de estações da região da BRO para calcular a evapotranspiração de referência com base no método de Penman-Monteith FAO-56. O autor aplicou um controle de qualidade criterioso: análise integrada de consistência com regressões lineares entre diferentes estações, exclusão de dados fora de padrões definidos, e preenchimento de falhas com base em dados da estação INMET São Carlos. Essas etapas garantiram a confiabilidade dos dados utilizados e reforçam a importância do pré-processamento em estudos hidrológicos.

Soares (2022) propôs o uso de Redes Neurais Artificiais (RNAs) regressivas para previsão de volume de chuva. Os modelos LSTM e MLP (*Multilayer Perceptron*) alcançaram acurárias de 80% e 81%, respectivamente, demonstrando o potencial das abordagens de inteligência artificial na previsão meteorológica.

Junior *et al.* (2024) realizaram uma análise comparativa de algoritmos de aprendizado de máquina para preenchimento de falhas em séries de temperatura do ar em quatro regiões brasileiras. Foram avaliados seis modelos: Regressão Linear, LASSO (*Least Absolute Shrinkage and Selection Operator*), Elastic Net, K-Nearest Neighbors (KNN), Árvores de Decisão (CART) e *Support Vector Regression* (SVR). O modelo SVR obteve os menores valores de RMSE (*Root Mean Squared Error*), sendo considerado o mais promissor. As Árvores de Decisão apresentaram desempenho estável, enquanto os modelos LASSO e Elastic Net foram os menos eficazes. O KNN obteve desempenho intermediário.

No estudo de Caminha *et al.* (2025), foram analisadas séries temporais de vazões e precipitações mensais na microrregião de Caratinga, pertencente à bacia hidrográfica do Rio Doce. A modelagem baseou-se nos critérios de seleção de Akaike, Hannan-Quinn e Schwarz,

sendo o modelo SARIMA o que apresentou melhor aderência às séries, especialmente na representação das vazões mínimas e dos períodos de estiagem, ainda que com limitações na reprodução dos picos de precipitação.

Por fim, Moro *et al.* (2025) investigaram as tendências em séries temporais meteorológicas e seus impactos na evapotranspiração de referência. A homogeneidade das séries foi verificada por meio dos testes estatísticos de Pettitt, Buishand e SNHT (Teste de Homogeneidade Normal Padrão), enquanto a análise de tendências utilizou os testes de Mann-Kendall e Sen. Os autores identificaram rupturas estatisticamente significativas nas séries diárias de temperatura máxima e mínima, umidade relativa, radiação solar, velocidade do vento e evapotranspiração de referência. Os resultados reforçam a importância de aplicar múltiplos testes de homogeneidade e compreender as interações entre variáveis meteorológicas que influenciam o processo de evapotranspiração.

A análise dos trabalhos demonstra que há uma diversidade de aplicações, desde a detecção de rupturas e tendências em séries meteorológicas até a previsão de eventos hidrológicos e o preenchimento de lacunas em dados ambientais. Observa-se também a ampla utilização de variáveis como temperatura do ar, precipitação, umidade relativa, radiação solar, velocidade do vento e evapotranspiração de referência, evidenciando sua relevância no contexto da modelagem e previsão climática.

A combinação de testes estatísticos tradicionais, como os de homogeneidade e tendência, com modelos de previsão baseados em redes neurais e algoritmos de regressão, reforça a importância de uma abordagem integrada e comparativa na análise de séries temporais. Os resultados obtidos em diferentes estudos indicam que, embora os modelos baseados em aprendizado de máquina tenham demonstrado alto desempenho preditivo, a robustez dos métodos estatísticos clássicos ainda é essencial para garantir a consistência e a interpretação dos dados analisados.

A Tabela 1 apresenta uma síntese comparativa entre os principais estudos revisados, destacando os objetivos, os métodos empregados, as bases de dados utilizadas e os resultados alcançados. Esta sistematização permite observar os enfoques metodológicos adotados na análise de séries temporais meteorológicas e hidrológicas, evidenciando as abordagens estatísticas clássicas e modernas técnicas baseadas em inteligência artificial e aprendizado de máquina.

Tabela 1 – Comparação entre trabalhos sobre séries temporais meteorológicas e hidrológicas

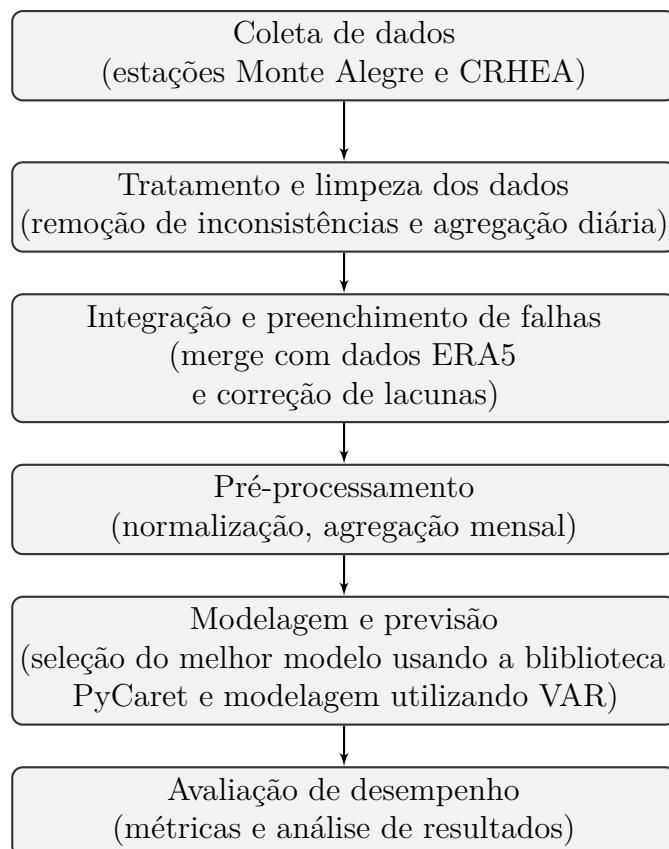
Referência	Objetivo	Dados utilizados	Modelos / Algoritmos	Principais conclusões
Kratzert <i>et al.</i> (2018)	Modelagem chuva–vazão com LSTM	241 bacias nos EUA	LSTM	Desempenho superior a modelos conceituais
Calixto (2020)	Calcular ETo e modelar hidrologia local	Dados meteorológicos (T, UR, Rs, Vento) da região da BRO	Método FAO-56 Penman-Monteith, controle de qualidade, interpolação IDW	ETo confiável após controle rigoroso de dados; destaca importância do pré-processamento
Soares (2022)	Prever volume de chuva com IA	Séries históricas de chuva	Redes neurais LSTM e MLP	LSTM e MLP alcançaram 80% e 81% de acerto, respectivamente; RNAs são eficazes para previsão meteorológica
Junior <i>et al.</i> (2024)	Preencher falhas em dados de temperatura do ar	Séries de temperatura com falhas em 4 estados brasileiros	Regressão linear, LASSO, Elastic Net, KNN, CART, SVR	SVR teve menor RMSE; CART teve desempenho estável; LASSO e EN foram menos eficazes
Caminha <i>et al.</i> (2025)	Modelar séries de precipitação e vazão mensal	Séries de precipitação e vazão (microrregião de Caratinga)	Modelos SARIMA, com seleção via AIC, BIC e HQC	SARIMA apresentou bom desempenho para estagens, mas dificuldade em captar picos de vazão
Moro <i>et al.</i> (2025)	Detectar tendências e rupturas em séries meteorológicas	Séries diárias de temperatura, umidade, radiação, vento e ETo	Testes de homogeneidade (Pettitt, Buishand, SNHT), Mann-Kendall, Sen	Identificadas rupturas em séries meteorológicas; recomendação de múltiplos testes para análise robusta de tendências

Fonte: Elaborado pelo autor (2025).

4 METODOLOGIA

De acordo com Nielsen (2019), Soares (2022), a preparação adequada dos dados temporais é uma etapa essencial para a eficácia da modelagem e para a confiabilidade das análises. O processo metodológico deste trabalho foi estruturado em seis etapas principais: coleta, tratamento, integração, pré-processamento, modelagem e avaliação de desempenho. O fluxo completo é apresentado na Figura 7, que ilustra as fases desde a obtenção dos dados até a análise dos resultados.

Figura 7 – Fluxo metodológico de coleta, tratamento, pré-processamento e modelagem dos dados meteorológicos.



Fonte: Elaborado pelo autor (2025)

O uso dos dados de reanálise ERA5¹, desenvolvido pelo *European Centre for Medium-Range Weather Forecasts* (ECMWF), para o preenchimento de falhas nas séries observadas foi condicionado à verificação de consistência estatística entre as bases. Para esse fim, foi realizada uma análise de correlação entre as séries diárias de precipitação e

¹ Fonte: Copernicus Climate Change Service (C3S) — ERA5 Reanalysis. Disponível em: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels>. Acesso em: 26 out. 2025.

temperatura da estação do CRHEA e as correspondentes séries do ERA5, considerando apenas os períodos com dados simultâneos disponíveis. O coeficiente de correlação de Pearson foi adotado como métrica objetiva de similaridade temporal, sendo estabelecido como critério mínimo aceitável valores superiores a 0,80, indicativos de forte associação linear. Apenas nos intervalos temporais que atenderam a esse critério os dados do ERA5 foram utilizados para o preenchimento das lacunas, de modo a preservar a coerência física e estatística da série consolidada e minimizar a introdução de vieses artificiais na modelagem subsequente.

4.1 Coleta de Dados

Os dados utilizados neste estudo foram obtidos a partir de registros históricos de estações meteorológicas localizadas na BRO, no município de Brotas/SP — área reconhecida como uma das principais zonas de recarga do SAG. As informações foram extraídas tanto de um banco de dados legado quanto de arquivos texto gerados automaticamente pelas estações ao longo dos anos.

A estação climatológica do CRHEA foi selecionada como fonte principal por apresentar a série temporal mais longa e contínua (1980–2025), o que proporciona maior robustez estatística à análise. Para complementar eventuais lacunas, foram empregados dados do conjunto ERA5. Esse repositório global combina observações meteorológicas com modelos de reanálise, fornecendo estimativas horárias de variáveis atmosféricas, terrestres e oceânicas. As variáveis consideradas neste estudo foram precipitação total e temperatura média do ar, ambas em frequência diária.

4.2 Tratamento dos Dados

O tratamento dos dados teve como objetivo eliminar inconsistências e padronizar os registros antes da modelagem estatística. As etapas compreenderam:

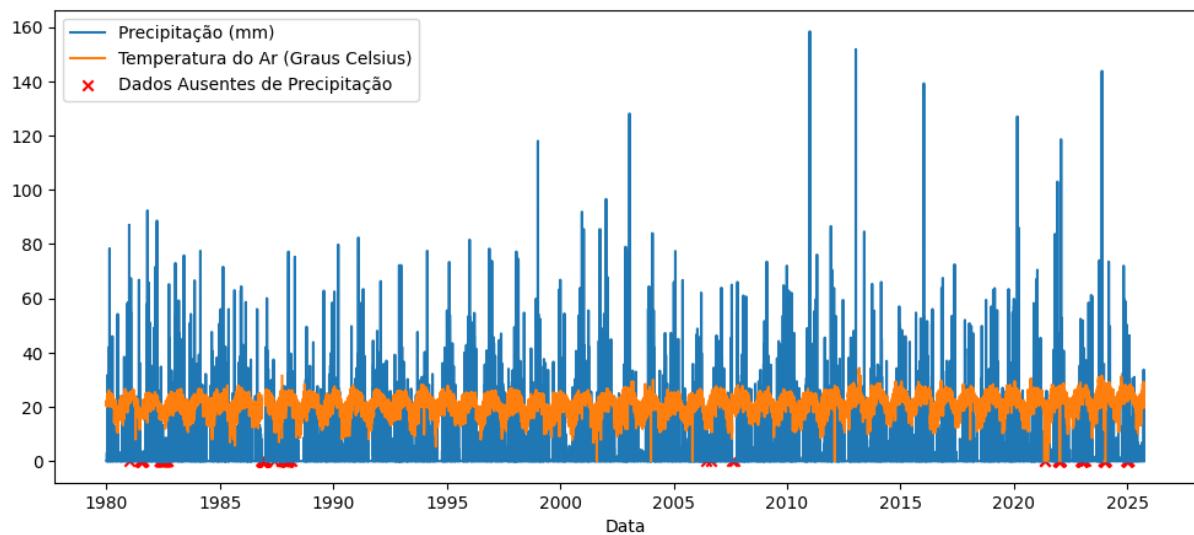
- **Remoção de inconsistências:** exclusão de registros textuais em campos numéricos, correção de formatação e eliminação de valores fora dos limites físicos plausíveis;
- **Conversão temporal:** criação de um índice de data a partir das colunas de ano e dia do ano (DOY), garantindo ordenação cronológica adequada;
- **Seleção de variáveis:** manutenção apenas das variáveis relevantes ao estudo (precipitação e temperatura média).

4.2.1 Integração e Preenchimento de Falhas

Após a limpeza inicial, verificou-se a existência de lacunas nas séries históricas (Figuras 8 e 9). Para reduzir o impacto desses dados ausentes, realizou-se a integração entre

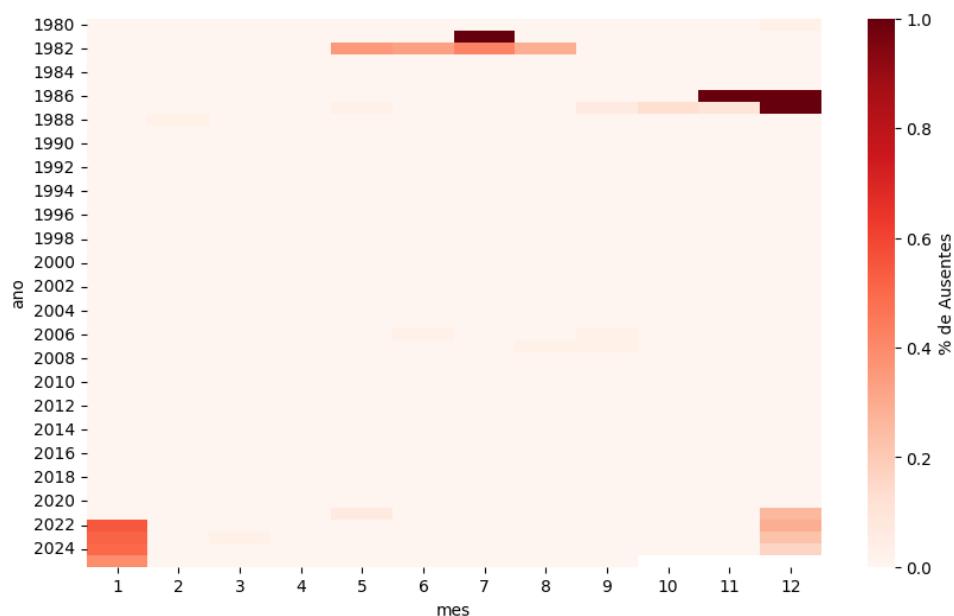
os dados do CRHEA e da ERA5, utilizando as datas como chave de correspondência. O preenchimento foi efetuado apenas quando as séries apresentaram alta correlação temporal, assegurando coerência física entre os conjuntos. Esse procedimento resultou em uma base consolidada, mais completa e representativa do comportamento climático da região.

Figura 8 – Série Temporal de Precipitação e Temperatura com Dados Ausentes



Fonte: Elaborado pelo autor (2025)

Figura 9 – Percentual de dados ausentes de precipitação por Ano e Mês



Fonte: Elaborado pelo autor (2025)

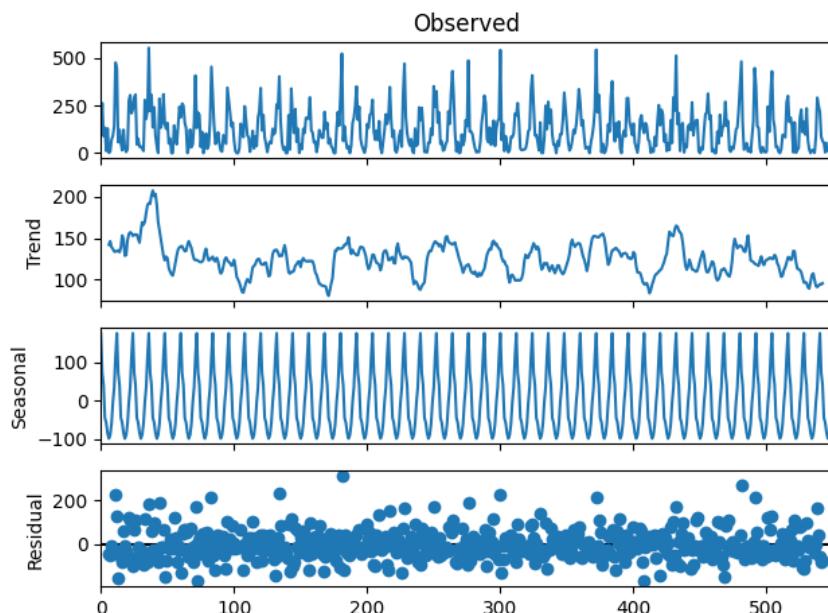
4.3 Pré-processamento para Modelagem de Séries Temporais

Com o conjunto de dados consolidado, aplicaram-se procedimentos de agregação e transformação para adequar as séries à modelagem temporal. Os dados diários foram convertidos para frequência mensal, sendo:

- a **precipitação** representada pela **soma mensal acumulada**;
- a **temperatura média**, pela **média aritmética mensal**.

Essa transformação reduziu a variabilidade diária e destacou padrões sazonais de longo prazo. Em seguida, procedeu-se à decomposição sazonal aditiva da precipitação (Figura 10), que evidenciou uma sazonalidade anual bem definida e uma tendência estável ao longo dos anos, com resíduos aleatórios, indicando que o modelo aditivo foi adequado para representar a estrutura temporal da série.

Figura 10 – Gráfico de Decomposição Sazonal



Fonte: Elaborado pelo autor (2025).

4.4 Modelagem e Previsão de Séries Temporais com PyCaret

A modelagem e a previsão das séries temporais foram realizadas utilizando a biblioteca PyCaret, por meio do módulo específico para análise de séries temporais (*time_series*). Essa biblioteca foi escolhida por automatizar e padronizar as etapas de pré-processamento, treinamento, comparação e seleção de modelos, garantindo reproduzibilidade e consistência metodológica.

A variável-alvo considerada foi a precipitação mensal, obtida após as etapas de tratamento e pré-processamento descritas nas seções anteriores. O ambiente experimental foi configurado por meio da função *setup*, na qual foram definidos a frequência temporal mensal da série e o horizonte de previsão de 60 meses. Nessa etapa, a PyCaret organiza automaticamente o processo de validação temporal e preparação dos dados.

Na sequência, diferentes modelos de previsão de séries temporais foram treinados e comparados com base em métricas estatísticas apropriadas. O modelo que apresentou melhor desempenho foi o Extra Trees com deseasonalização condicional e *detrending*, implementado internamente no PyCaret por meio do *forecaster* BaseCdsDtForecaster. Esse *forecaster* encapsula um *pipeline* que realiza, de forma automática, a remoção condicional da sazonalidade e da tendência antes da etapa de previsão, reintegrando esses componentes ao final do processo.

Após a seleção, o modelo foi ajustado à série completa e utilizado para gerar previsões futuras da precipitação mensal. Os resultados foram analisados conjuntamente com os dados históricos observados, permitindo a avaliação do comportamento esperado da variável ao longo do horizonte de previsão. Os principais trechos de código utilizados encontram-se descritos no Apêndice A.

4.4.0.1 Modelo Extra Trees

O modelo *Extremely Randomized Trees* (Extra Trees) é um método de aprendizado de máquina baseado em conjuntos de árvores de decisão, caracterizado pela introdução de um elevado grau de aleatoriedade tanto na seleção das variáveis quanto na definição dos pontos de corte, conforme proposto por Geurts, Ernst e Wehenkel (2006). Essa estratégia contribui para a redução da variância do modelo e para o aumento de sua capacidade de generalização, especialmente em problemas com elevada complexidade e não linearidade.

No contexto deste trabalho, o modelo Extra Trees foi aplicado à previsão de séries temporais de precipitação por meio do *BaseCdsDtForecaster*, que combina o regressor Extra Trees com etapas de pré-processamento específicas para séries temporais, incluindo a dessazonalização condicional e a remoção de tendência. Essas etapas permitem que o modelo aprenda padrões residuais mais estáveis, favorecendo o desempenho preditivo.

A divisão entre os conjuntos de treinamento e teste foi realizada de forma sequeencial, respeitando a ordem temporal das observações. O conjunto de treinamento foi composto pela porção inicial da série temporal, utilizada para o ajuste dos parâmetros do modelo, enquanto o conjunto de teste correspondeu ao período final da série, reservado exclusivamente para avaliação do desempenho preditivo. Essa abordagem evita vazamento de informação (*data leakage*) e é considerada apropriada para problemas de previsão em séries temporais.

As previsões foram realizadas considerando um horizonte de um passo à frente (*one-step ahead*), sendo os valores previstos comparados diretamente com as observações correspondentes no período de teste para o cálculo das métricas de erro.

4.5 Modelagem e Previsão com VAR

Como abordagem comparativa, foi utilizado o modelo VAR, adequado à análise conjunta de séries temporais correlacionadas, neste caso, precipitação e temperatura média mensal. O procedimento seguiu as etapas de verificação de estacionariedade, diferenciação das séries quando necessário, definição da ordem ótima de defasagem com base no Critério de Informação Bayesiano (BIC), ajuste do modelo e geração de previsões.

4.5.0.1 Pré-processamento e Estacionariedade

Como o modelo VAR requer séries estacionárias, aplicou-se o Teste de Dickey-Fuller Aumentado (ADF) para verificar a estacionariedade das séries. Diante da presença de não-estacionariedade, procedeu-se à diferenciação de primeira ordem, obtendo séries estacionárias adequadas ao ajuste do modelo.

4.5.0.2 Especificação e Ajuste do Modelo VAR(p)

O modelo VAR(p) foi definido conforme a Equação 4.1:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t \quad (4.1)$$

em que Y_t representa o vetor de variáveis endógenas, c o vetor de constantes, Φ_i as matrizes de coeficientes, p a ordem de defasagem e ε_t o vetor de erros.

A ordem de defasagem do modelo VAR foi definida com base no BIC, o qual se caracteriza por uma penalização mais severa da complexidade do modelo, favorecendo especificações mais parcimoniosas (Lutkepohl, 2005). De acordo com esse critério, a ordem ótima indicada foi $p = 17$. Embora esse valor seja relativamente elevado para séries mensais, o modelo VAR foi empregado neste estudo com finalidade essencialmente comparativa, não sendo utilizado como ferramenta principal de inferência ou interpretação das relações dinâmicas. Dessa forma, a escolha de $p = 17$ não compromete os objetivos do trabalho, uma vez que a análise concentrou-se no desempenho preditivo global e na adequação estatística do modelo.

4.5.0.3 Previsão e Reversão da Diferenciação

Com o modelo ajustado, geraram-se previsões para os 60 meses subsequentes (cinco anos). As previsões foram inicialmente obtidas para as séries diferenciadas e, em seguida, revertidas para o nível original por meio da soma cumulativa, conforme a Equação 4.2:

$$\hat{Y}_t = \hat{Y}_{t-1} + \Delta \hat{Y}_t \quad (4.2)$$

4.5.0.4 Avaliação e Diagnóstico do Modelo

A validação do modelo incluiu os seguintes procedimentos:

- **Teste de estabilidade:** verificação das raízes do polinômio característico, assegurando que todas permanecessem dentro do círculo unitário;
- **Teste de Portmanteau:** confirmação de que os resíduos apresentaram comportamento de ruído branco;
- **Avaliação preditiva out-of-sample:** comparação das previsões com dados reais de um conjunto de teste (24 meses) por meio das métricas RMSE e MAE;
- **Análise de resposta ao impulso (IRF):** avaliação dos efeitos dinâmicos de choques em uma variável sobre a outra ao longo do tempo.

Ressalta-se que, embora a temperatura do ar tenha sido analisada ao longo do processo de tratamento e exploração dos dados, a avaliação quantitativa do desempenho preditivo foi conduzida exclusivamente para a variável precipitação. Essa escolha se justifica pelo papel central da precipitação nos processos de recarga aquífera, bem como por sua maior variabilidade temporal e impacto direto na dinâmica hidrológica da área de estudo.

Os principais trechos de código utilizados nas etapas de tratamento, análise exploratória e modelagem apresentados nesta seção encontram-se documentados no Apêndice A, que contém o código-fonte desenvolvido em *Python* para a análise dos dados meteorológicos.

5 RESULTADOS

Este capítulo apresenta os resultados obtidos a partir da modelagem de séries temporais de precipitação, considerando duas abordagens distintas: (i) a abordagem automatizada baseada na biblioteca PyCaret, que selecionou o modelo Extra Trees com melhor desempenho e (ii) o modelo estatístico multivariado VAR. Em ambos os casos, a avaliação foi realizada exclusivamente sobre o período de teste.

5.1 Métricas de avaliação

A avaliação do desempenho preditivo foi conduzida com base nas métricas *Mean Absolute Error* (MAE) e *Root Mean Squared Error* (RMSE), calculadas a partir da comparação entre os valores observados e previstos de precipitação ao longo de todo o período de teste. O MAE expressa o erro médio absoluto entre as previsões e as observações, mantendo a mesma unidade da variável analisada, conforme a Equação 5.1.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.1)$$

Já o RMSE é definido pela Equação 5.2 e penaliza de forma mais severa erros de maior magnitude, sendo particularmente relevante em séries de precipitação, caracterizadas pela ocorrência de eventos extremos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.2)$$

Nessas equações, y_i representa o valor observado, \hat{y}_i o valor previsto pelo modelo e n o número de observações no período avaliado. Em ambas as métricas, valores menores indicam melhor desempenho preditivo.

A métrica MAPE (*Mean Absolute Percentage Error*) não foi utilizada, pois sua aplicação a séries de precipitação apresenta limitações relevantes. A ocorrência frequente de valores nulos ou próximos de zero pode gerar erros percentuais elevados ou indefinidos, comprometendo a interpretação dos resultados. Dessa forma, optou-se pelo uso exclusivo das métricas MAE e RMSE, que se mostram mais adequadas às características da variável analisada.

5.2 Resultados do modelo Extra Trees

Os resultados apresentados nesta seção referem-se ao desempenho do modelo *Extremely Randomized Trees* (Extra Trees), selecionado automaticamente por meio da

biblioteca PyCaret a partir do conjunto de modelos avaliados. Ressalta-se que a PyCaret foi utilizada exclusivamente como ferramenta de apoio ao processo de seleção e ajuste, sendo o Extra Trees o modelo efetivamente empregado na etapa de previsão.

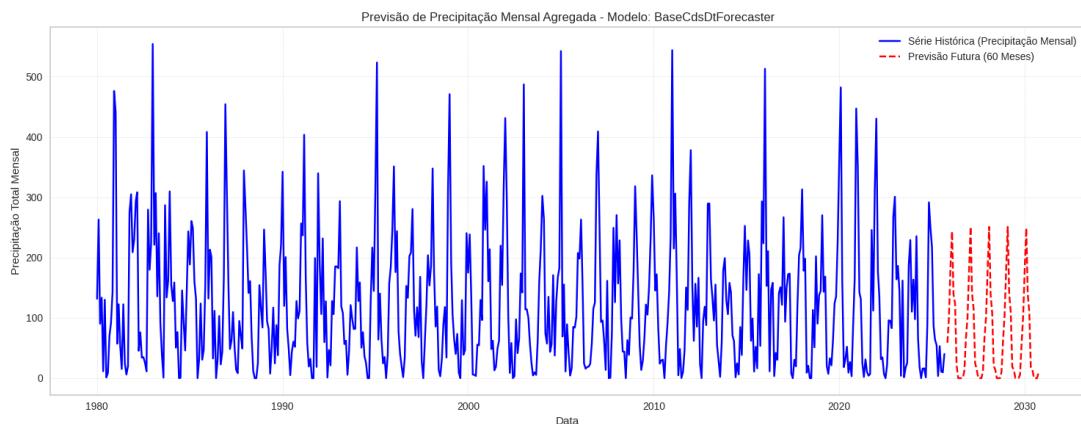
O modelo Extra Trees foi aplicado à série temporal de precipitação após a realização dos procedimentos de pré-processamento descritos no Capítulo 4, incluindo a remoção de tendência e a dessazonalização condicional. Essas etapas visaram reduzir a influência de padrões determinísticos de longo prazo e da sazonalidade anual, permitindo que o modelo aprendesse de forma mais eficiente os componentes residuais da série.

A avaliação do desempenho preditivo foi realizada exclusivamente sobre o conjunto de teste, utilizando uma divisão temporal sequencial entre os dados de treinamento e teste. As previsões foram geradas considerando um horizonte de um passo à frente (*one-step ahead*), sendo posteriormente comparadas com os valores observados para o cálculo das métricas de erro.

Os resultados indicaram que o modelo Extra Trees apresentou bom desempenho na representação da variabilidade temporal da precipitação, conforme evidenciado pelos valores reduzidos de MAE e RMSE obtidos no período de teste. Observa-se que o modelo foi capaz de acompanhar adequadamente as flutuações da série, incluindo variações mais abruptas associadas a eventos de maior intensidade pluviométrica, ainda que com limitações inerentes à elevada variabilidade natural desse tipo de fenômeno climático.

A Figura 11 ilustra a comparação entre os valores observados e os valores previstos pelo modelo Extra Trees ao longo do período avaliado. Nota-se que, de forma geral, as previsões seguem o comportamento da série observada, com discrepâncias mais pronunciadas concentradas em períodos de extremos pluviométricos, os quais representam um desafio comum em modelos de previsão climática.

Figura 11 – Gráfico de Previsão de Precipitação utilizando o Modelo Extra Trees



Fonte: Elaborado pelo autor (2025).

De maneira geral, os resultados obtidos nesta etapa demonstram o potencial do modelo Extra Trees na previsão de séries temporais de precipitação, reforçando sua adequação para aplicações ambientais e hidrológicas em contextos caracterizados por não linearidade e alta variabilidade temporal.

5.3 Resultados do modelo VAR

Nesta seção são apresentados os resultados obtidos com a aplicação do modelo de Vetores Autorregressivos (VAR) à série temporal de precipitação, considerando a abordagem multivariada descrita no Capítulo 4. O modelo VAR foi ajustado utilizando exclusivamente os dados do conjunto de treinamento, respeitando a ordem temporal das observações, e avaliado posteriormente sobre o conjunto de teste, de forma análoga ao procedimento adotado para o modelo Extra Trees.

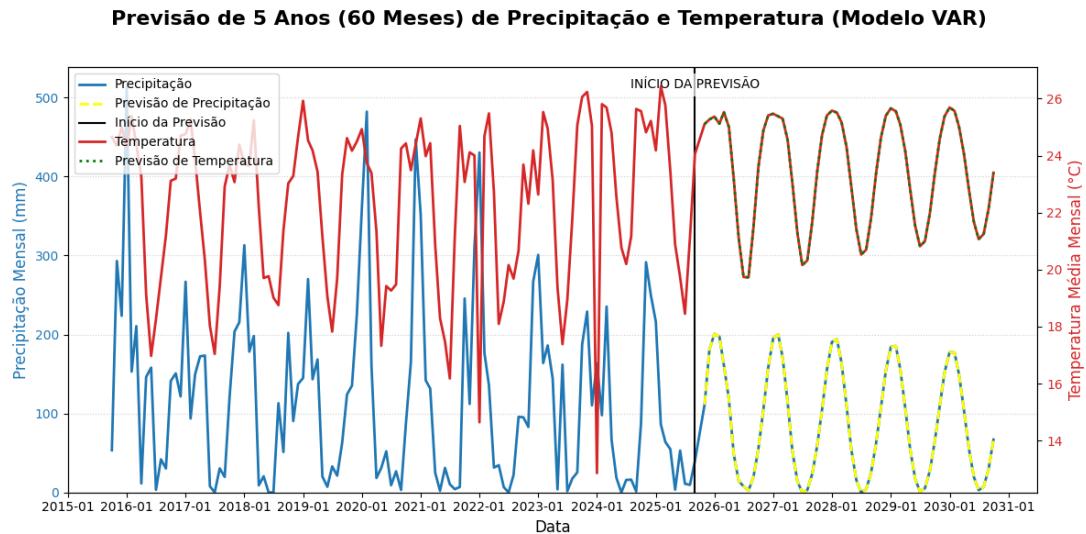
A especificação do modelo VAR exigiu a verificação prévia da estacionariedade das séries envolvidas, bem como a aplicação de procedimentos de diferenciação e transformação, conforme discutido na etapa metodológica. A ordem do modelo foi determinada com base em critérios de informação, visando um equilíbrio entre qualidade de ajuste e complexidade do modelo.

Os resultados obtidos indicaram que o modelo VAR apresentou desempenho inferior ao modelo Extra Trees no período de teste, conforme evidenciado pelos valores mais elevados das métricas MAE e RMSE. Observou-se que, embora o VAR tenha sido capaz de capturar tendências gerais da série, apresentou maior dificuldade em representar variações mais abruptas da precipitação, especialmente aquelas associadas a eventos extremos.

Essa limitação está associada, em grande parte, aos pressupostos de linearidade e estacionariedade inerentes ao modelo VAR, que restringem sua capacidade de modelar comportamentos não lineares e interações complexas típicas de séries climáticas. Como consequência, erros de maior magnitude tendem a ocorrer em períodos caracterizados por alta variabilidade pluviométrica, o que se reflete nos valores mais elevados de RMSE.

A Figura 12 apresenta a previsão de precipitação gerada pelo modelo VAR para o período avaliado, permitindo a comparação visual entre os valores observados e previstos. Nota-se que, embora o modelo apresente desempenho satisfatório em períodos de variação mais suave, sua capacidade preditiva é reduzida em situações de maior complexidade dinâmica, reforçando as limitações observadas na análise quantitativa.

Figura 12 – Gráfico de Previsão de Precipitação utilizando o Modelo VAR



Fonte: Elaborado pelo autor (2025).

5.4 Comparação entre os Modelos de Previsão

A comparação entre os modelos de previsão foi realizada considerando o desempenho do modelo Extra Trees, selecionado automaticamente por meio da biblioteca PyCaret, e do modelo estatístico multivariado VAR, ambos avaliados exclusivamente sobre o conjunto de teste, respeitando a natureza temporal das séries analisadas. A divisão entre os conjuntos de treinamento e teste foi realizada de forma sequencial, utilizando-se a porção inicial da série para o ajuste dos modelos e o período final para a avaliação preditiva, evitando qualquer forma de embaralhamento dos dados, procedimento inadequado para séries temporais.

Cabe destacar que a comparação entre os modelos concentrou-se na variável precipitação, considerada a principal variável de interesse para estudos de recarga hídrica, enquanto a temperatura foi utilizada de forma exploratória e como subsídio à análise climática, não sendo incluída no cálculo das métricas de erro nesta etapa do estudo.

Os valores obtidos para as métricas MAE e RMSE estão apresentados na Tabela 2. Observa-se que o modelo Extra Trees apresentou menores valores de erro quando comparado ao modelo VAR, indicando maior precisão média das previsões e melhor capacidade de representar variações mais intensas da precipitação no período de teste.

Tabela 2 – Métricas de desempenho dos modelos de previsão de precipitação

Modelo	MAE (mm)	RMSE (mm)
Extra Trees	45.45	64.59
VAR	56.92	69.15

Fonte: Elaborado pelo autor (2025).

Esse resultado evidencia a vantagem do modelo Extra Trees na captura de padrões não lineares e de interações complexas presentes nas séries climáticas, mesmo após os procedimentos de pré-processamento, como a remoção de tendência e a dessazonalização. O elevado grau de aleatoriedade introduzido no processo de construção das árvores contribui para a redução da variância do modelo e para uma melhor capacidade de generalização, aspecto fundamental em séries temporais ambientais com alta variabilidade natural.

Por outro lado, embora o modelo VAR seja amplamente utilizado na modelagem de séries temporais multivariadas, sua performance mostrou-se inferior neste estudo, possivelmente em função de seus pressupostos de linearidade e estacionariedade, que limitam sua capacidade de representar variações abruptas e comportamentos não lineares típicos da precipitação pluviométrica. Tal limitação se reflete, sobretudo, nos valores mais elevados de RMSE, indicando maior sensibilidade do modelo a erros associados a eventos extremos.

De forma geral, os resultados demonstram que o modelo Extra Trees apresentou desempenho superior ao modelo VAR no contexto analisado, reforçando o potencial de técnicas de aprendizado de máquina para a previsão de variáveis climáticas quando adequadamente selecionadas, ajustadas e avaliadas em períodos de teste bem definidos.

5.5 Plataforma Web dos Resultados

Esta seção apresenta a plataforma web desenvolvida como um desdobramento aplicado dos resultados obtidos na análise e modelagem das séries temporais climáticas. A implementação da plataforma teve como principal motivação a necessidade de organizar, integrar e disponibilizar, de forma estruturada, os dados observacionais, os dados tratados e os produtos gerados ao longo do estudo, facilitando sua visualização, análise e reutilização em contextos ambientais e hidrológicos.

A plataforma foi concebida como um ambiente integrador, no qual são centralizadas as informações referentes às estações meteorológicas, séries temporais climáticas, metadados operacionais e resultados das previsões geradas pelos modelos aplicados neste trabalho. Dessa forma, a plataforma não executa os modelos de previsão em tempo real, mas disponibiliza os resultados previamente obtidos, permitindo sua consulta, interpretação e análise de forma acessível e sistematizada.

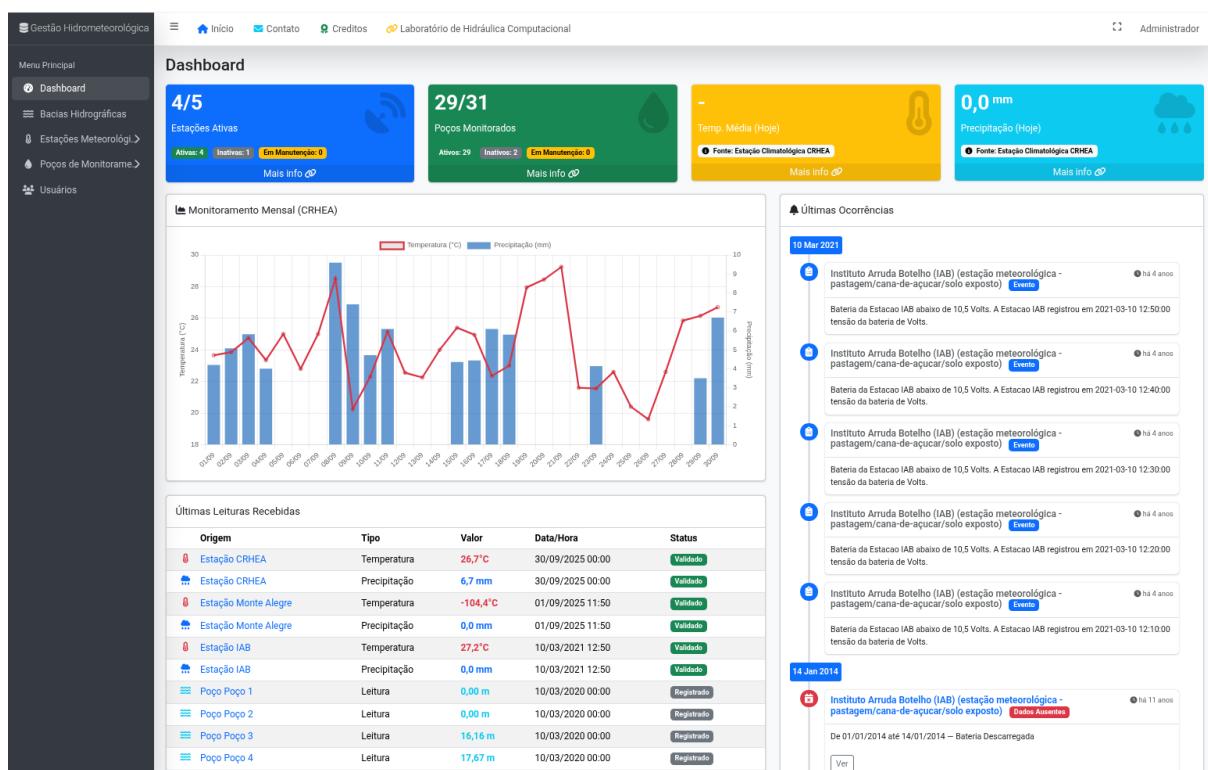
No que se refere à modelagem preditiva, a plataforma possibilita a visualização das séries temporais observadas e das previsões geradas pelo modelo Extra Trees e pelo modelo VAR, incluindo estimativas para períodos futuros dentro do horizonte de previsão considerado. Essa funcionalidade estabelece uma conexão direta entre a etapa de modelagem descrita nos capítulos anteriores e a aplicação prática dos resultados, tornando-os mais acessíveis a usuários não especializados em técnicas de modelagem estatística ou

aprendizado de máquina.

A organização modular da plataforma reflete as etapas metodológicas do estudo, permitindo o acesso aos módulos de gestão de bacias hidrográficas, estações meteorológicas, poços de monitoramento e usuários, bem como a visualização sintética das séries climáticas analisadas. Essa estrutura favorece uma navegação intuitiva e contribui para a coerência entre o processamento dos dados e sua representação no ambiente computacional.

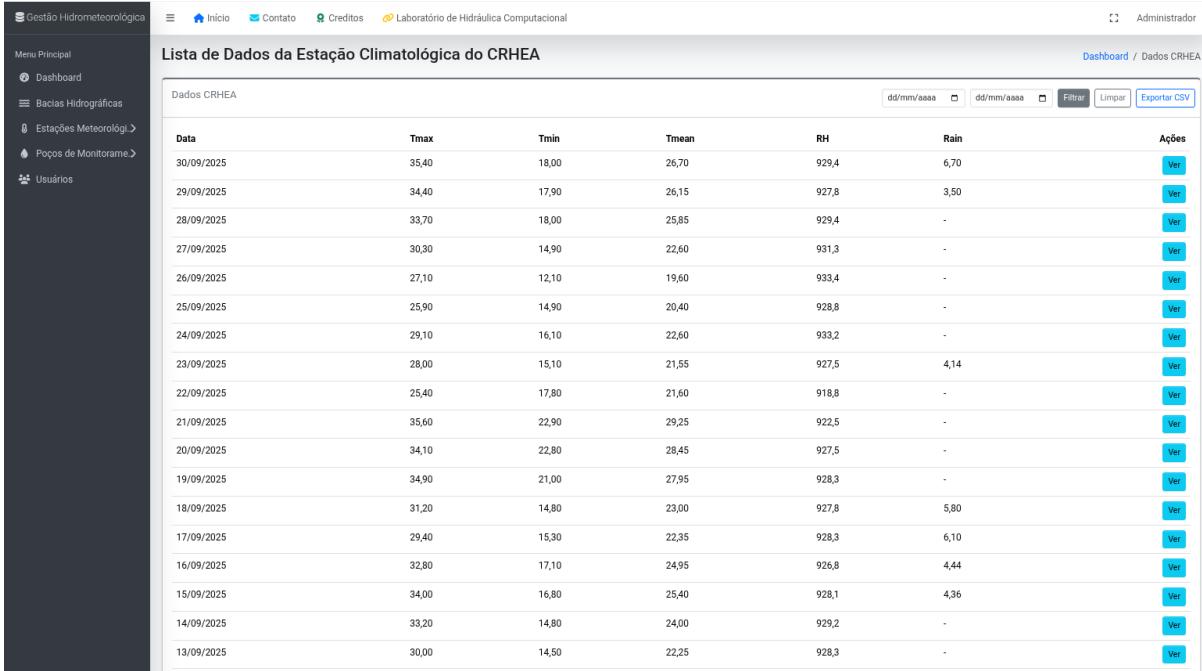
As Figuras 13 a 15 ilustram as principais interfaces da plataforma, incluindo a tela inicial, os módulos de consulta e visualização de séries temporais e o gerenciamento de metadados operacionais. O registro desses metadados é fundamental para a rastreabilidade e contextualização dos dados, contribuindo para a confiabilidade das análises e para a correta interpretação dos resultados gerados.

Figura 13 – Interface principal da plataforma web SGDHM



Fonte: Elaborado pelo autor (2025).

Figura 14 – Consulta e visualização de séries temporais no SGDHM

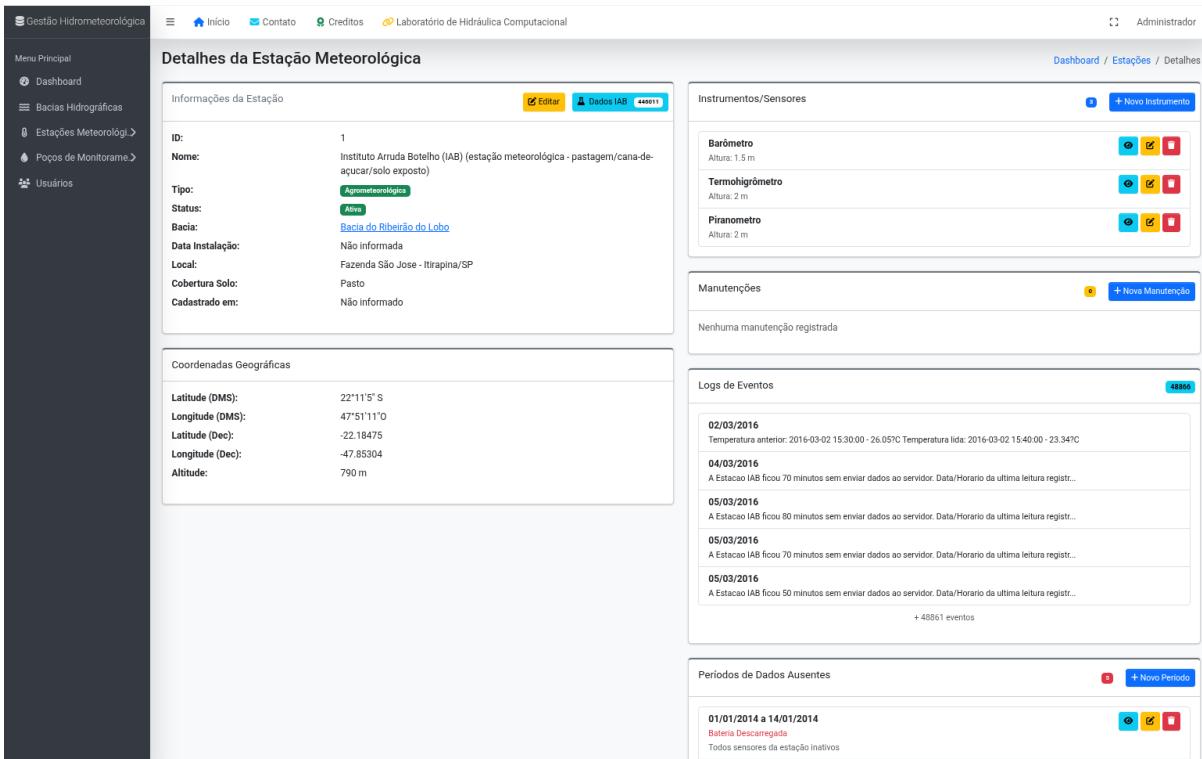


The screenshot shows a web-based application for hydrological management. The left sidebar has a dark theme with a 'Menu Principal' section containing 'Dashboard', 'Bacias Hidrográficas', 'Estações Meteorológicas', 'Poços de Monitoramento', and 'Usuários'. The main content area has a light background. At the top, there are navigation links: Início, Contato, Credito, Laboratório de Hidráulica Computacional, and a user icon labeled 'Administrador'. Below these are links for 'Dashboard / Dados CRHEA' and 'Exportar CSV'. The central part of the screen displays a table titled 'Lista de Dados da Estação Climatológica do CRHEA'. The table has columns for 'Data', 'Tmax', 'Tmin', 'Tmean', 'RH', 'Rain', and 'Ações'. The data rows show various weather parameters for dates from 30/09/2025 to 13/09/2025. Each row includes a blue 'Ver' button in the 'Ações' column.

Data	Tmax	Tmin	Tmean	RH	Rain	Ações
30/09/2025	35,40	18,00	26,70	929,4	6,70	<button>Ver</button>
29/09/2025	34,40	17,90	26,15	927,8	3,50	<button>Ver</button>
28/09/2025	33,70	18,00	25,85	929,4	-	<button>Ver</button>
27/09/2025	30,30	14,90	22,60	931,3	-	<button>Ver</button>
26/09/2025	27,10	12,10	19,60	933,4	-	<button>Ver</button>
25/09/2025	25,90	14,90	20,40	928,8	-	<button>Ver</button>
24/09/2025	29,10	16,10	22,60	933,2	-	<button>Ver</button>
23/09/2025	28,00	15,10	21,55	927,5	4,14	<button>Ver</button>
22/09/2025	25,40	17,80	21,60	918,8	-	<button>Ver</button>
21/09/2025	35,60	22,90	29,25	922,5	-	<button>Ver</button>
20/09/2025	34,10	22,80	28,45	927,5	-	<button>Ver</button>
19/09/2025	34,90	21,00	27,95	928,3	-	<button>Ver</button>
18/09/2025	31,20	14,80	23,00	927,8	5,80	<button>Ver</button>
17/09/2025	29,40	15,30	22,35	928,3	6,10	<button>Ver</button>
16/09/2025	32,80	17,10	24,95	926,8	4,44	<button>Ver</button>
15/09/2025	34,00	16,80	25,40	928,1	4,36	<button>Ver</button>
14/09/2025	33,20	14,80	24,00	929,2	-	<button>Ver</button>
13/09/2025	30,00	14,50	22,25	928,3	-	<button>Ver</button>

Fonte: Elaborado pelo autor (2025).

Figura 15 – Gerenciamento de metadados operacionais no SGDHM



The screenshot shows a detailed view of a meteorological station's metadata. The left sidebar is identical to Figure 14. The main content area has a light background. At the top, there are navigation links: Início, Contato, Credito, Laboratório de Hidráulica Computacional, and a user icon labeled 'Administrador'. Below these are links for 'Dashboard / Estações / Detalhes'. The central part of the screen displays a form titled 'Detalhes da Estação Meteorológica'. It is divided into several sections: 'Informações da Estação' (with fields for ID, Name, Type, Status, Bacia, Data Instalação, Local, Cobertura Solo, and Cadastrado em), 'Instrumentos/Sensores' (listing Barômetro, Termohigrômetro, and Piranômetro with edit and delete icons), 'Manutenções' (listing none), and 'Logs de Eventos' (listing events from 02/03/2016 to 05/03/2016). Below these are sections for 'Coordenadas Geográficas' (Latitude and Longitude) and 'Períodos de Dados Ausentes' (listing a period from 01/01/2014 to 14/01/2014 with a 'Bateria Desacarregada' status). Each section has a 'Novo' button at the top right.

Fonte: Elaborado pelo autor (2025).

Cabe destacar que a plataforma web não tem como objetivo substituir ferramentas especializadas de modelagem hidrológica ou climatológica, nem realizar estimativas diretas de processos físicos, como a recarga aquífera. Seu papel é apoiar a análise e a interpretação dos resultados obtidos neste trabalho, funcionando como um recurso complementar que materializa os produtos da pesquisa e amplia seu potencial de aplicação em estudos ambientais e no suporte à gestão de recursos hídricos.

6 CONCLUSÃO

Este trabalho teve como objetivo analisar e modelar séries temporais climáticas associadas à região de recarga do Sistema Aquífero Guarani, com foco na precipitação, a partir de dados observacionais da estação meteorológica do CRHEA e de dados de reanálise ERA5. A partir das etapas de tratamento, integração e análise dessas séries, buscou-se avaliar o desempenho de diferentes abordagens de modelagem preditiva, bem como disponibilizar os resultados de forma estruturada por meio de uma plataforma web de apoio à análise ambiental.

Os resultados evidenciaram a importância do pré-processamento das séries temporais, incluindo a remoção de tendência, a dessazonalização e o preenchimento de falhas com base em critérios estatísticos objetivos. A utilização dos dados de reanálise ERA5 mostrou-se adequada para complementar as séries observacionais, desde que condicionada à verificação de consistência entre as bases, contribuindo para a redução de lacunas e para a melhoria da continuidade temporal dos dados analisados.

No que se refere à modelagem preditiva, o modelo de aprendizado de máquina *Extremely Randomized Trees* (Extra Trees), selecionado por meio da biblioteca PyCaret, apresentou desempenho superior ao modelo estatístico clássico de Vetores Autorregressivos (VAR), conforme indicado pelas métricas MAE e RMSE calculadas sobre o período de teste. O melhor desempenho do Extra Trees está associado à sua capacidade de capturar padrões não lineares e interações complexas presentes nas séries de precipitação, mesmo em contextos caracterizados por elevada variabilidade climática.

Por outro lado, embora o modelo VAR seja amplamente utilizado na análise de séries temporais multivariadas, seus pressupostos de linearidade e estacionariedade limitaram sua capacidade de representar variações abruptas da precipitação, refletindo-se em valores de erro mais elevados. Esses resultados indicam que, para o contexto analisado, modelos baseados em aprendizado de máquina podem oferecer vantagens significativas em relação a abordagens estatísticas tradicionais, desde que adequadamente pré-processados e avaliados em períodos de teste bem definidos.

Como desdobramento aplicado da pesquisa, foi desenvolvida uma plataforma web destinada à organização, visualização e consulta dos dados observacionais, dos dados tratados e das previsões geradas pelos modelos avaliados. A plataforma atua como um recurso complementar, materializando os resultados obtidos e ampliando seu potencial de aplicação em estudos ambientais e no suporte à gestão de recursos hídricos, sem substituir ferramentas especializadas de modelagem física ou hidrológica.

Ressalta-se que o presente estudo não teve como objetivo estimar diretamente a

recarga do Sistema Aquífero Guarani, mas fornecer subsídios técnicos e metodológicos que possam contribuir para investigações futuras nesse sentido. Como trabalhos futuros, destaca-se a ampliação da abordagem preditiva por meio da inclusão explícita de outras variáveis meteorológicas, como temperatura do ar, evapotranspiração e umidade relativa, bem como a avaliação do desempenho dos modelos em diferentes escalas temporais e espaciais.

Como trabalho futuro, sugere-se a aplicação de modelos baseados em redes neurais recorrentes, em especial arquiteturas do tipo *Long Short-Term Memory* (LSTM), amplamente utilizadas na modelagem de séries temporais não lineares devido à sua capacidade de representar dependências de longo prazo (Hochreiter; Schmidhuber, 1997). A investigação dessas abordagens poderá contribuir para ganhos adicionais de desempenho preditivo em relação aos modelos avaliados neste trabalho.

REFERÊNCIAS

- ALI, M. PyCaret: An open source, low-code machine learning library in Python. [S.l.], 2020. Versão 1.0.0 (Primeira versão estável). Disponível em: <https://pycaret.org/>.
- BARRETO, C. E. A. G. Balanço hídrico em zona de afloramento do Sistema Aquífero Guarani a partir de monitoramento hidrogeológico em bacia representativa. 2006. Dissertação (Dissertação (Mestrado em Engenharia - Área de Concentração: Hidráulica e Saneamento)) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2006. Orientador: Edson Cesar Wendland. Disponível em: <https://teses.usp.br/>.
- CALIXTO, K. G. Relações causais e persistências na dinâmica da disponibilidade hídrica para a formulação de modelos hidrológicos conceituais. 2020. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo, São Carlos, 2020. Disponível em: <https://www.teses.usp.br/teses/disponiveis/18/18138/tde-10062021-112534/>. Acesso em: 25 jun. 2025.
- CAMINHA, A. R. *et al.* Previsão de precipitação e vazão a partir da modelagem de séries temporais. **Revista Brasileira de Climatologia**, v. 36, n. 21, p. 264–290, fev. 2025. Disponível em: <https://ojs.ufgd.edu.br/rbclima/article/view/18885>.
- GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely Randomized Trees. **Machine Learning**, v. 63, n. 1, p. 3–42, 2006.
- GOMES, L. H. Determinação da recarga profunda na bacia-piloto do Ribeirão da Onça em zona de afloramento do Sistema Aquífero Guarani a partir de balanço hídrico em zona saturada. 2008. Dissertação (Dissertação de Mestrado em Hidráulica e Saneamento) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008. Acesso em: 2025-07-06.
- HAMILTON, J. D. **Time Series Analysis**. Princeton: Princeton University Press, 1994.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.
- JUNIOR, A. A. d. S. *et al.* Preenchimento de falhas em séries temporais da temperatura do ar: uma comparação entre modelos de machine learning. **Revista Brasileira de Climatologia**, v. 35, n. 20, p. 362–377, out. 2024. Disponível em: <https://ojs.ufgd.edu.br/rbclima/article/view/17649>.
- KRATZERT, F. *et al.* Rainfall-runoff modelling using long short-term memory (lstm) networks. **Hydrology and Earth System Sciences**, Copernicus Publications, v. 22, n. 11, p. 6005–6022, 2018.
- LUCAS, M. C. Influência da precipitação e do uso do solo sobre a taxa de recarga em zona de afloramento do sistema Aquífero Guarani. 2012. Dissertação (Dissertação (Mestrado)) — Universidade de São Paulo, Escola de Engenharia de São Carlos, São Carlos, SP, 2012. Acesso em: 06 jul. 2025. Disponível em: <http://www.teses.usp.br/teses/disponiveis/18/18138/tde-31102012-092335/>.

LUTKEPOHL, H. **New Introduction to Multiple Time Series Analysis**. Berlin: Springer, 2005.

MORO, I. P. *et al.* Tendências em séries temporais meteorológicas e seus impactos na evapotranspiração de referência diária. **Revista Brasileira de Climatologia**, v. 36, n. 21, p. 202–224, fev. 2025. Disponível em: <https://ojs.ufgd.edu.br/rbclima/article/view/19077>.

NIELSEN, A. **Practical Time Series Analysis: Prediction with Statistics and Machine Learning**. Sebastopol, CA: O'Reilly Media, 2019.

(SGB), S. G. do B. **Aquífero Guarani**. 2025. <https://www.sgb.gov.br/aquifero-guarani>. Acesso em: 06 jul. 2025.

SIMS, C. A. Macroeconomics and reality. **Econometrica**, v. 48, n. 1, p. 1–48, 1980.

SOARES, A. C. **Análise de séries temporais com uso de redes neurais artificiais em dados meteorológicos para previsão de chuva e eventos climáticos severos**. 2022. 112 f. Dissertação (Dissertação de Mestrado em Ciência da Computação) — Universidade Federal de Uberlândia, Uberlândia, 2022.

WENDLAND, E.; BARRETO, C. E. A. G.; GOMES, L. H. Water balance in the guarani aquifer outcrop zone based on hydrogeologic monitoring. **Journal of Hydrology (Amsterdam)**, v. 342, p. 261–269, 2007.

ZOU, Y. *et al.* Complex network approaches to nonlinear time series analysis. **Physics Reports**, v. 787, p. 1–97, 2019. ISSN 0370-1573. Complex network approaches to nonlinear time series analysis. Disponível em: <https://www.sciencedirect.com/science/article/pii/S037015731830276X>.

APÊNDICES

APÊNDICE A – CÓDIGO-FONTE EM PYTHON PARA ANÁLISE DE DADOS METEOROLÓGICOS

A.1 Introdução

Este apêndice apresenta os principais trechos de código utilizados nas análises de precipitação e temperatura do conjunto de dados do CRHEA (1980–2025). Os scripts foram desenvolvidos em Python utilizando bibliotecas como `pandas`, `matplotlib`, `statsmodels`, `pycaret` e `VAR`. O objetivo é documentar as etapas de tratamento, análise exploratória, modelagem e previsão.

A.2 Leitura e preparação dos dados

```

1 import pandas as pd
2
3 # Leitura do arquivo Excel
4 df = pd.read_excel("CRHEA_1980-2025.xlsx")
5
6 # Cria a coluna de data com base no ano e doy (dia
7 # sequencial do ano)
8 df[‘data’] = pd.to_datetime(df[“yr”].astype(str) + df[“
9 doy”].astype(str).str.zfill(3), format=%Y%j)
10
11 # Seta a coluna de data como índice
12 df = df.set_index(‘data’)
13
14 # Ordena por data
15 df = df.sort_values(‘data’)
16
17 # Selecionando apenas as colunas de interesse, precipitaç
18 # ão e temperatura média em um novo dataframe
19 data = df[['rain', 'tmean']].copy()

```

Listing A.1 – Leitura do arquivo e tratamento inicial dos dados

A.3 Tratamento de erros e valores faltantes

```

1 import pandas as pd
2 import numpy as np
3

```

```

4      # Conversão de valores para numérico, transformando
5          '*****' e outros erros em NaN
6      data = data.replace('*****', np.nan)
7      data = data.apply(pd.to_numeric, errors='coerce')
8
9      # Substitui valores vazios ou espaços em branco por NaN
10     data['rain'] = data['rain'].replace(r'^\s*$', np.nan,
11         regex=True)
12     data['tmean'] = data['tmean'].replace(r'^\s*$', np.nan,
13         regex=True)
14
15     # Define a precipitação como NaN se a temperatura também
16         for NaN, ou seja, não houve leitura naquele dia
17     data.loc[data['tmean'].isna(), 'rain'] = np.nan
18
19     # Preenchimento dos dados faltantes de precipitação com o
20         conjunto de dados ERA5
21     df_dados_era5_precipitacao = pd.read_csv(
22         'serie_tp_era_5_area_dados_faltantes_crhea.csv',
23         parse_dates=True, index_col=0)
24
25     # Ordena por data
26     df_dados_era5_precipitacao = df_dados_era5_precipitacao.
27         sort_values("valid_time")
28
29     # Agregação diária com acumulado somado
30     df_dados_era5_precipitacao = df_dados_era5_precipitacao.
31         resample("D").agg({
32             "tp": "sum",
33             "tp_mm": "sum",
34         })
35
36     # Preenchimento dos dados faltantes de precipitação da
37         estação do CRHEA com dados mensais acumulado da ERA5
38     data['rain'] = data['rain'].fillna(
39         df_dados_era5_precipitacao["tp_mm"])

```

Listing A.2 – Correção de erros e preenchimento de dados faltantes

A.4 Análise exploratória e visualização

```
1 import matplotlib.pyplot as plt
```

```

2
3     # Gráfico com leituras diárias de precipitação e com
4     # marcação dos dados ausentes
5     plt.figure(figsize=(12,5))
6     plt.plot(data.index, data['rain'], label='Precipitação')
7     plt.plot(data.index, data['tmean'], label="Temperatura do
8         Ar")
9     plt.scatter(data.index[data['rain'].isna()], [0]*data['
10        rain'].isna().sum(),
11        color='red', label='Ausentes', marker='x')
12     plt.legend()
13     plt.title('Série Temporal com Dados Ausentes Marcados')
14     plt.show()

```

Listing A.3 – Visualização da série temporal de precipitação e temperatura

A.5 Modelagem com PyCaret

```

1     from pycaret.time_series import setup, compare_models,
2         finalize_model, predict_model
3
4     # --- Configuração do PyCaret ---
5     # Inicializando o ambiente com os dados Mensais
6     # fh=60: Horizon of the forecast (60 meses, equivalente a
7     # cinco anos)
8     print("\nIniciando a Configuração do PyCaret (Dados
9         Mensais)...")
10    s_mensal = setup(
11        data = df,
12        target = coluna_precipitacao,
13        fh = 60,
14        session_id = 456,
15        ignore_prep_pipeline = True,
16        verbose = False
17    )
18
19    # --- Comparação e Seleção do Melhor Modelo (Mensal) ---
20    print("Comparando Modelos (Mensal)...")
21    best_model_mensal = compare_models()
22
23    print(f"\n--- O Melhor Modelo Selecionado pelo PyCaret (
24        Mensal) é: {type(best_model_mensal).__name__} ---")

```

```

21
22     # --- Treinamento e Finalização do Melhor Modelo ---
23     final_model_mensal = finalize_model(best_model_mensal)
24
25     # --- Previsão Futura (Forecast) ---
26     # Gerando a previsão para 60 meses
27     predictions_mensal = predict_model(final_model_mensal, X=
28         df[coluna_temperatura_media])
29
29     predictions_mensal['y_pred'] = np.clip(predictions_mensal
30         ['y_pred'], 0, None)

```

Listing A.4 – Modelagem e Previsão com PyCaret

A.6 Teste de Estacionariedade

```

1
2     def check_stationarity(series, name):
3         """Executa o teste ADF e retorna o p-valor e o
4             resultado."""
5         result = adfuller(series, autolag='AIC')
6         is_stationary = result[1] < 0.05
7         print(f" - {name}: p-valor = {result[1]:.4f} ->
8             Estacionário: {is_stationary}")
9         return is_stationary
10
11
12     print("##      Verificação de Estacionariedade (Teste ADF
13         )")
14     is_rain_stationary = check_stationarity(df[
15         coluna_precipitacao], "Precipitação Mensal")
16     is_tmean_stationary = check_stationarity(df[
17         coluna_temperatura_media], "Temperatura Média Mensal")
18
19     # --- Diferenciação se Necessário (Para alcançar a
20         estacionariedade) ---
21     if not is_rain_stationary or not is_tmean_stationary:
22         print("\n      Uma ou ambas as séries NÃO são
23             estacionárias. Aplicando Diferenciação (Ordem
24             1).")
25         model_data = df.diff().dropna()
26         print("\n      Verificação de Estacionariedade após
27             Diferenciação:")

```

```

18         check_stationarity(model_data[coluna_precipitacao]
19                         , "Precipitação Mensal (Diferenciada)")
20         check_stationarity(model_data[
21             coluna_temperatura_media], "Temperatura Média
22             Mensal (Diferenciada)")
23     else:
24         model_data = df.copy()

```

Listing A.5 – Teste de Dickey-Fuller Aumentado - ADF

A.7 Modelagem com VAR

```

1 import pandas as pd
2 from statsmodels.tsa.api import VAR
3
4 # Determinar a ordem ideal do VAR (p).
5 max_lags = 24 # Testar defasagens de até 2 anos
6 var_model = VAR(model_data)
7
8 # Seleciona o número ideal de defasagens 'p' que minimiza
9 # o BIC (Bayesian Information Criterion)
10 lag_order = var_model.select_order(maxlags=max_lags).bic
11 if lag_order == 0:
12     lag_order = 1 # Garante pelo menos uma defasagem
13
14 # Treinar o modelo VAR
15 var_model_fitted = var_model.fit(lag_order)
16
17 # Previsão de 5 anos (60 meses)
18 forecast_steps = 60
19 last_values = model_data.values[-lag_order:]
20 forecast_diff = var_model_fitted.forecast(y=last_values,
21                                             steps=forecast_steps)
22
23 # Organizar o resultado da previsão (Dados Diferenciados)
24 forecast_diff_df = pd.DataFrame(forecast_diff, columns=
25                                 model_data.columns)
26
27 # Reverter a Diferenciação (Necessário para obter valores
28 # reais) ---
29 # O forecast reverte-se somando o último valor real da sé
30 # de origem.

```

```

26     last_original_values = df.iloc[-1]
27     forecast_df = (forecast_diff_df.cumsum() +
28                       last_original_values)
29
29     # Criar o índice de tempo para a previsão
30     last_date = df.index[-1]
31     # O índice começa no mês seguinte ao último dado
32     forecast_index = pd.date_range(start=last_date, periods=
33                                     forecast_steps + 1, freq='M')[1:]
33     forecast_df.index = forecast_index
34
35     # Arredondar e garantir que a precipitação não seja
36     # negativa
36     forecast_df = forecast_df.round(2)
37     forecast_df[coluna_precipitacao] = forecast_df[
37                                         coluna_precipitacao].clip(lower=0)

```

Listing A.6 – Ajuste do modelo VAR e previsão futura

A.8 Considerações finais

Os códigos apresentados neste apêndice servem como base para a análise exploratória, modelagem e previsão dos dados meteorológicos. Pequenas variações podem ser aplicadas conforme a necessidade de atualização dos conjuntos de dados ou ajustes metodológicos.