

Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Engenharia Aeroespacial

**METODOLOGIA PARA CONSTRUÇÃO DE
MODELOS DE PREVISÃO HIDROLÓGICA
USANDO SARIMAX**

Autora: Emilly Meireles Braga Cornelio
Orientador: Prof. Dr. Jorge Cormane
Coorientador: Prof. Dr. Manuel Barcelos Jr

Brasília, DF
2025



EMILLY MEIRELES BRAGA CORNELIO

Metodologia Para Construção De Modelos De Previsão Hidrológica Usando SARIMAX

Monografia submetida ao curso de graduação em Engenharia Aeroespacial da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia Aeroespacial.

Orientador: Prof. Dr. Jorge Cormane

Coorientador: Prof. Dr. Manuel Barcelos Jr

Brasília, DF

2025

A Naum Pereira Braga
Meu avô e primeiro professor

AGRADECIMENTOS

Agradeço à Raquel e ao Flávio pelos esclarecimentos acerca de seu trabalho e por disponibilizar os dados utilizados.

Agradeço, também, ao meu orientador e ao meu coorientador, o professor Jorge e o professor Manuel, por toda a disponibilidade, paciência, dedicação e oportunidades que me proporcionaram.

Agradeço ao professor William, pelas pacientes explicações e pelo apoio.

Agradeço à minha mãe, Gleisimone, meu padrasto, Eudes, e meu pai, Luciano, que, se pudessem, teriam até feito o trabalho por mim.

E ao Rafael, por me salvar das crises de ansiedade e me lembrar que a vida é boa.

O presente estudo foi realizado no âmbito do Termo de Execução Descentralizada Nº 06/2023, celebrado entre a UnB e o CENSIPAM, cujo objeto é "Desenvolvimento de ferramenta computacional baseada em métodos matemáticos para a detecção automática e a adequação supervisionada de inconsistências em séries temporais de dados coletados de postos de medição fluviométrica (nível, vazão, medição de descargas líquida, calibração de curva chave) localizadas na Bacia Amazônica". Processo Nº 60092.000169/2023-12.

RESUMO

A previsão de dados hidrológicos é importante para a gestão de recursos hídricos e para garantir a qualidade de vida da população dependente da dinâmica do rio. Esse processo de previsão considera dados anteriores da série temporal analisada e dados de variáveis correlacionadas, por meio de métodos de regressão como o método dos Mínimos Quadrados. Entretanto, como dados hidrológicos possuem componentes sazonais, modelos que consideram esse comportamento, como o SARIMAX, são mais indicados. Dessa forma, a previsão com Mínimos Quadrados e SARIMAX são comparadas na aplicação em dados de vazão da região hidrográfica do Tocantins-Araguaia. Os resultados mostram que ambos os modelos apresentam bons ajustes. As métricas de erro e a análise dos resíduos entre os valores previstos e observados apontam que a previsão do modelo SARIMAX é mais precisa e o modelo é mais adequado ao problema analisado, mostrando menores resíduos entre os valores observados e previstos. Por fim, o SARIMAX é aplicado a um estudo de caso em São Félix do Araguaia para estimativa de vazão e de cota, apresentando resultados novamente próximos dos valores observados.

Palavras-chave: Previsão de dados hidrológicos; Vazão; SARIMAX; Método dos Mínimos Quadrados.

ABSTRACT

The forecasting of hydrological data is essential for water resource management and for ensuring the quality of life of populations that depend on river dynamics. This forecasting process takes into account historical data from the analyzed time series as well as data from correlated variables, using regression methods such as the Least Squares method. However, since hydrological data often exhibit seasonal components, models that account for this behavior, such as SARIMAX, are more appropriate. Therefore, forecasts using Least Squares and SARIMAX are compared through their application to streamflow data from the Tocantins-Araguaia river basin. The results show that both models provide good fits. The error metrics and the residual analysis between the predicted and observed values indicate that the SARIMAX model provides more accurate forecasts and is better suited to the problem under study, showing smaller residuals between observed and predicted values. Finally, SARIMAX is applied to a case study in São Félix do Araguaia for streamflow and water level estimation, once again producing results that closely match the observed data.

Keywords: Forecasting of hydrological data; Streamflow; SARIMAX; Least Squares Method.

LISTA DE FIGURAS

Figura 1 - a) dados plotados em gráfico de pontos; b) dados plotados em gráfico de linha (Nielsen, 2019)	17
Figura 2 - Coeficiente de mortalidade infantil em São Paulo - SP entre 1900 e 1994 (Antunes, 1997).....	18
Figura 3 - a) correlação entre a distância entre Mercúrio e a Lua e pesquisas do Google por "minecraft"; b) Correlação entre o consumo de iogurte e a produção de energia renovável no Brasil (Vigen, 2025).....	20
Figura 4 – ACF com características não estacionárias (Silva, 2025)	23
Figura 5 - Isoietas de precipitação anual (ANA, 2009).....	27
,	29
Figura 6 – a) Sub Bacia Hidrográfica 22 (Tocantins entre os rios Paranã e Sono); b) Sub Bacia hidrográfica 26 (trecho da ilha do bananal) (ANA, 2016)	29
Figura 7 – Cidade de São Félix do Araguaia - Mato Grosso em estado de emergência devido à cheia no rio Araguaia (Diário Araguaia, 2022)	
Figura 8 - Fluxograma do trabalho	31
Figura 9 - Diagrama de caixas (Trøndheim e Nielsen, 2022).....	33
Figura 10 - Polígonos de Thiessen gerados com os dados de precipitação na região da UHE de Estreito.....	35
Figura 11 - Precipitação na região da UHE de Estreito.....	36
Figura 12 - Distribuição amostral dos dados de vazão do Estreito	36
Figura 13 - resultado do modelo, em que a linha tracejada vermelha indica a previsão com Mínimos Quadrados, a linha pontilhada azul indica o resultado da previsão com SARIMAX, e a linha preta contínua são os resultados observados	38
Figura 14 - Relação entre a vazão observada e a) a vazão prevista pelo Mínimos Quadrados; b) a vazão prevista pelo SARIMAX	39
Figura 15 - Distribuição amostral dos dados observados e previstos.....	39
Figura 16 – a) Resíduos e previsões do modelo MMQ, em que a linha preta contínua preta representa os valores observados, a linha tracejada vermelha representa a previsão do MMQ, e a linha contínua vermelha são os resíduos; b) Resíduos e previsões do modelo SARIMAX, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos; c) Resíduo dos modelos, em que a linha tracejada vermelha indica os resíduos do MMQ, e a linha pontilhada azul indica os resíduos do SARIMAX.	41

Figura 17 - Distribuição dos resíduos entre os dados observados e os previstos	42
Figura 18 - ACF dos resíduos dos modelos, em que os pontos vermelhos indicam os resíduos do modelo do MMQ, e os azuis correspondem aos resíduos do modelo do SARIMAX.....	43
Figura 19 - Gráficos de dispersão dos resíduos pela previsão para o modelo do a) Mínimos Quadrados; b) SARIMAX.....	44
Figura 20 - Polígonos de Thiessen gerados a partir do dados de precipitação na região de São Félix do Araguaia	46
Figura 21 - Dados de precipitação na região de São Félix do Araguaia.....	47
Figura 22 - Vazão prevista em São Felix do Araguaia com o SARIMAX, em que a linha pontilhada azul indica a vazão prevista, e a linha contínua é a vazão observada.....	48
Figura 23 - Relação entre os dados previstos e observados de vazão em São Félix do Araguaia.....	49
Figura 24 - Distribuição amostral dos dados observados e previstos em São Félix do Araguaia.....	49
Figura 25 - Resíduos e previsões do modelo SARIMAX para vazão em São Félix do Araguaia, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos.....	50
Figura 26 - Distribuição dos resíduos.....	51
Figura 27 - Função de autocorrelação dos resíduos entre os valores observados e previstos para a vazão em São Félix do Araguaia.....	51
Figura 28 – Gráfico de dispersão dos resíduos entre a vazão observada e prevista em São Félix do Araguaia	52
Figura 29 – Previsão da cota com o SARIMAX em que os dados previstos são dados pela linha pontilhada verde, os dados observados são representados pela linha contínua preta, e a linha horizontal vermelha destaca a cota de alerta de inundações.....	53
Figura 30 – Relação entre os dados observados e previstos de cota em São Félix do Araguaia.....	54
Figura 31 - Distribuição dos dados observado e previstos de cota em São Félix do Araguaia.....	55
Figura 32 - Resíduos e previsões do modelo SARIMAX para cota em São Félix do Araguaia, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos.....	56

Figura 33 - Distribuição dos resíduos entre os valores previstos e observados de cota em São Félix do Araguaia	56
Figura 34 - ACF dos resíduos entre os valores observados e previstos de cota em São Félix do Araguaia	57
Figura 35 - Dispersão dos resíduos entre os valores previstos e observados de cota em São Félix do Araguaia	58

LISTA DE TABELAS

Tabela 1 – Correlação de Spearman entre a precipitação e a vazão no Estreito.....	37
Tabela 2 - Coeficientes de correlação entre os valores observados e estimados.....	38
Tabela 3 - Resultados do teste de Ljung-Box para os resíduos dos modelos	44
Tabela 4 - Comportamento da variância dos resíduos	45
Tabela 5 - Erros dos modelos	45
Tabela 6 – Correlação de Spearman entre vazão e precipitação em São Félix do Araguaia	47
Tabela 7 - Coeficientes de correlação entre os valores observados e previstos de vazão em São Félix do Araguaia	48
Tabela 8 - Teste de Goldfeld-Quandt para os resíduos entre os valores estimados e observados de vazão em São José do Araguaia.....	52
Tabela 9 – Erros do modelo de vazão.....	52
Tabela 10 - Coeficientes de correlação entre os valores observados e previstos de vazão em São Félix do Araguaia.....	54
Tabela 11 - Teste de Goldfeld-Quandt para os resíduos entre os valores observados e previstos de cota em São Félix do Araguaia.....	57
Tabela 12 – Erros do modelo de cota	58

LISTA DE SIGLAS

ACF – *Autocorrelation Function* (Função de Autocorrelação)

PACF – *Partial Autocorrelation Function* (Função de Autocorrelação Parcial)

AR – *AutoRegressive* (modelo autorregressivo)

MA – *Moving Average* (modelo de médias móveis)

ARIMA – *Autoregressive Integrated Moving Average* (Modelo Autorregressivo Integrado de Médias Móveis)

AIC - *Akaike Information Criterion* (Critério de informação Akaike)

BIC – *Bayesian Information Criterion* (Critério de Informação Bayesiano)

SARIMA – *Seasonal Autorregressive Integrated Moving Average* (Modelo Autorregressivo Integrado de Médias Móveis Sazonal)

SARIMAX – *Seasonal Autoregressive Integrated Moving Average with eXogenous variables* (Modelo Autorregressivo Integrado de Médias Móveis Sazonal com Variáveis Exógenas)

DTW – *Dynamic Time Warping*

MAE – Erro Absoluto Médio

MSE – Erro Quadrático Médio

sMAPE – Erro Percentual Absoluto Médio Simétrico

LISTA DE SÍMBOLOS

Símbolos latinos

- r – coeficiente de correlação de Pearson
- x_i – valor da variável x em um dado momento i
- y_i - valor da variável y em um dado momento i
- \bar{x} - média dos valores de x
- \bar{y} – média entre os valores de y
- n - número de observações
- d_i - diferença entre as classificações dos pares em um momento i
- n_c – número de pares concordantes
- n_d - número de pares discordantes
- n_{total} – número total de pares
- H_0 – hipótese nula
- H_1 – hipótese não nula
- Q – p-valor do teste de Ljung-Box
- m – número de defasagens
- F - p-valor do teste de Goldfeld-Quandt
- $S(x)$ - função de distribuição acumulada empírica
- $F(x)$ - função de distribuição acumulada teórica
- D_{max} - maior diferença encontrada $S(x)$ e $F(x)$
- var_{max} - variância no grupo em que há maior variância
- var_{min} - variância no grupo com menor variância
- y_t – função de previsão
- p – ordem do modelo autorregressivo
- c – constante do modelo autorregressivo
- d – ordem de diferenciação
- q – ordem do modelo de médias móveis
- L – verossimilhança dos dados
- k – constante
- B – termo de defasagem da equação do modelo SARIMA

D – ordem de diferenciação da parte sazonal da série temporal de dados no modelo SARIMA

s – subescrito que representa a parte sazonal da série temporal no modelo SARIMA

Q_{cal} – vazão calculada

P_m – precipitação média

A – área da bacia hidrográfica

res – resíduos entre os valores observados e previstos

y_{prev} – valor previsto

y_{obs} – valor observado

Símbolos gregos

ρ – coeficiente de correlação de Spearman

τ – coeficiente de correlação de Kendall

β_0 – intercepto da função de regressão linear

β_1, β_2 – coeficientes angulares da função de regressão linear

ε – termo de erro da função de regressão linear

φ – parâmetro do modelo autorregressivo

θ – parâmetro do modelo de médias móveis

Φ – parâmetro do modelo autorregressivo da parte sazonal da série temporal de dados no modelo SARIMA

Θ – parâmetro do modelo de médias móveis da parte sazonal da série temporal de dados no modelo SARIMA

SUMÁRIO

1 INTRODUÇÃO	13
1.1 OBJETIVOS	14
1.1.1 Objetivo Geral	14
1.1.2 Objetivos Específicos	15
2 MATERIAIS	16
2.1 ANÁLISE DE SÉRIES TEMPORAIS	16
2.2 SÉRIES TEMPORAIS HIDROLÓGICAS	21
2.3 PREVISÃO DE SÉRIES TEMPORAIS.....	24
3 METODOLOGIA	27
3.1 ÁREA DE ESTUDO	27
3.2 MÉTODO.....	30
4 RESULTADOS	35
4.1 VALIDAÇÃO DO MÉTODO	35
4.2 ESTUDO DE CASO: SÃO FÉLIX DO ARAGUAIA.....	46
5 CONCLUSÃO	59
5.1 Trabalhos Futuros	60
6 REFERÊNCIAS	62

1 INTRODUÇÃO

Nas comunidades ribeirinhas da Bacia Amazônica, o rio representa cultura, subsistência e transporte, de forma que elas são altamente dependentes da dinâmica do rio. Nesse contexto, a navegação, utilizando barcos e canoas, constitui o principal meio de transporte, enquanto a pesca representa a base da alimentação e atividades profissionais. Além disso, o ciclo das águas favorece o plantio, fertilizando naturalmente a terra. Essa relação profunda com o rio também é expressa culturalmente, visível na construção das casas, sempre voltadas para o curso d'água, e no folclore regional, que inclui personagens como a Boiúna, cobra gigante que habita o rio e ataca embarcações (Lira e Chaves, 2016).

Baseada nessa dependência, tem-se a necessidade de prever o comportamento das variáveis hidrológicas para garantir a qualidade de vida da população local, além de definir medidas de mitigação para eventos hidrológicos extremos, sendo imprescindível para a gestão dos recursos hídricos (Pereira, Barbieiro e Quevedo, 2020). Dentre os dados hidrológicos, a previsão de vazão e cota dos rios é importante para minimizar as incertezas do clima, informação da qual dependem sistemas de abastecimentos de água, irrigação, navegação, geração de energia e conservação ambiental (da Silva, Tucci e Collischonn, 2005). Apesar de sua importância, é notável a falta de ferramentas de monitoramento e previsão de dados hidrológicos, colocando em risco a qualidade de vida da população.

Considerando isso, a previsão desses dados hidrológicos é promovida por uma série de métodos estatísticos que se baseiam em medições passadas para estimar o futuro. Monteiro *et al.* (2024) mostram que é possível obter essas previsões de forma simples e eficiente ao usarem o Método dos Mínimos Quadrados (MMQ) para previsão de dados de vazão utilizando valores passados e valores de precipitação como variável auxiliar. Entretanto, dados hidrológicos são caracterizados pela sazonalidade, que não é considerada nesse método. Sendo assim, o uso de modelos que incluam a sazonalidade, como o SARIMAX, podem apresentar um ajuste ainda melhor aos dados estudados.

Dessa forma, para validação do uso do SARIMAX em séries de dados hidrológicas, este trabalho se divide em duas etapas: uma de validação do método e um estudo de caso. A etapa de validação consiste em empregar e comparar o emprego do SARIMAX e do MMQ a dados de vazão na Usina Hidrelétrica (UHE) de Estreito,

conforme o procedimento mostrado em Monteiro *et al.* (2024) para confirmar a superioridade do modelo que inclui sazonalidade. Finalmente, considerando o desempenho na etapa de validação, o SARIMAX é aplicado novamente aos dados de vazão e cota, ou seja, nível do rio em relação ao nível do mar, na região de São Félix do Araguaia.

A região de São Félix do Araguaia, localizada na região hidrográfica do Tocantins-Araguaia, é considerada uma região de risco de alagamento onde a proteção civil e o monitoramento têm papel essencial, se mostrando ainda mais necessária a aplicação de ferramentas de previsão hidrológica aos dados coletados na região (Agência Nacional de Águas, 2024). Embora os dados de vazão possam auxiliar na previsão de inundações, esse risco é medido oficialmente pelos dados de cota. Assim, como forma de se certificar da eficiência da aplicação do método de previsão explorado ao contexto de previsão de alagamentos, o estudo de caso também envolve a previsão das cotas.

Esse risco de alagamentos é monitorado com auxílio do projeto da Universidade de Brasília junto ao Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM) do qual faz parte este trabalho. O projeto visa a identificação de dados hidrológicos inconsistentes por meio de métodos numéricos e estatísticos para auxílio no planejamento de medidas de mitigação de eventos hidrológicos extremos como alagamentos no contexto da Bacia Amazônica.

Dessa forma, este trabalho analisa e compara o emprego do SARIMAX no contexto dos dados hidrológicos da região hidrográfica do Tocantins-Araguaia, buscando promover previsões confiáveis que auxiliem no planejamento de medidas de mitigação de eventos extremos na região estudada, garantindo a segurança e a qualidade de vida da população local.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Desenvolver um modelo estatístico para previsão de dados hidrológicos de vazão e cota, auxiliando no monitoramento de dados hidrológicos e assegurando a segurança da população de regiões dependentes dos rios.

1.1.2 Objetivos Específicos

- Construir um modelo de vazão com auxílio de dados de precipitação;
- Validar os métodos de previsão estudados;
- Comparar as diferentes formas de aplicação dos métodos de previsão.

2 MATERIAIS

2.1 ANÁLISE DE SÉRIES TEMPORAIS

Segundo Antunes e Cardoso (2015), as séries temporais são uma forma de organizar informações quantitativas no tempo, estudando-as conforme sua distribuição no tempo. Box, Jenkins e Reinsel (2008) complementam afirmando que elas são caracterizadas por normalmente apresentarem dependência entre as observações adjacentes, além de repetições ao longo de um intervalo de tempo. Por sua vez, Nielsen (2019) define a análise das séries temporais como uma questão de causalidade, resumindo-se à pergunta “como o passado afeta o futuro?”. Dessa forma, baseando-se nos valores verificados no passado, se torna possível estimar o futuro.

No contexto da medicina o estudo de Antunes e Cardoso (2015) apresenta como as séries temporais são utilizadas para analisar eventos epidemiológicos. Os autores exemplificam como a taxa de mortalidade infantil em São Paulo entre 1900 e 1994, que apresentou uma queda. Considerando esse comportamento, pode-se prever que nos anos seguintes, os valores continuariam em queda. Esse comportamento continuado de declínio é chamado de tendência. Uma série temporal pode ter tendência crescente, decrescente ou estacionária, podendo apresentar mais de uma tendência em diferentes sequências. Segundo Gujarati e Porter (2011), uma série é estacionária se a média e a variância não mudam com o tempo e a covariância entre dois períodos depende apenas da defasagem entre as medições.

Outra característica é a sazonalidade, que os autores comparam ao ritmo de uma música, sendo definida como repetições organizadas no tempo. Um exemplo clássico de sazonalidade são as vendas de cartões de natal, que se dão anualmente na mesma época. Nesse sentido, Davey e Flores (1993) propõem um método para identificação de sazonalidade de séries temporais que se baseia em testes estatísticos de índices sazonais, permitindo a criação de modelos mais precisos para estudo dos dados. Nielsen (2019) ressalta que a forma de visualização dos dados é importante na identificação da sazonalidade. A figura 1a apresenta um gráfico de pontos em que não é possível visualizar o fenômeno, que, por sua vez, fica claro em um gráfico de linha, mostrado na figura 1b.

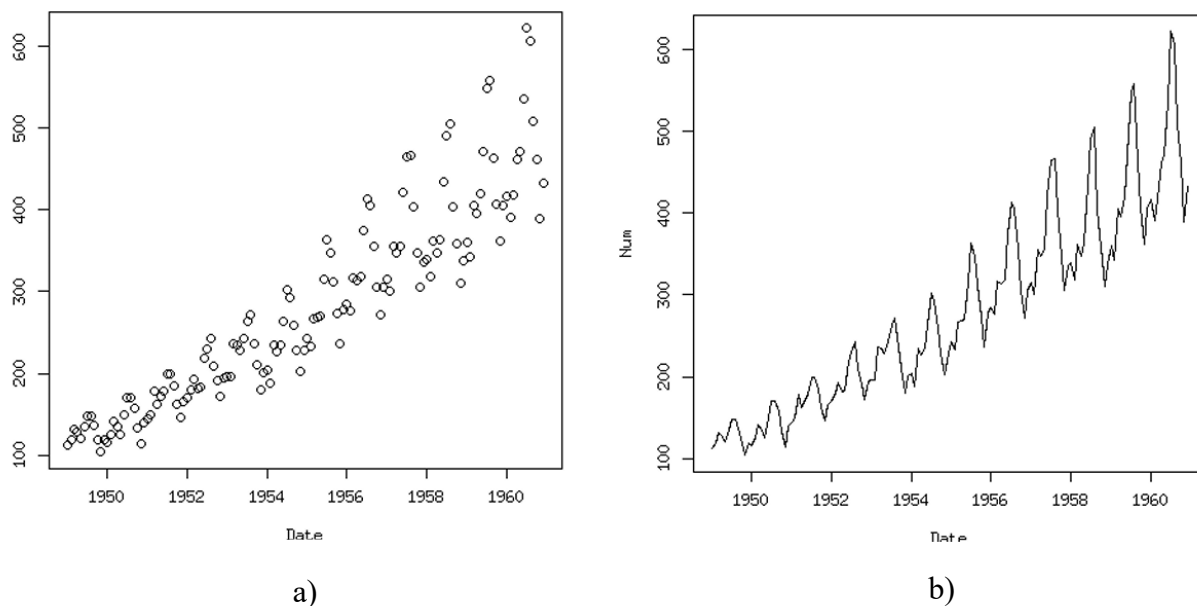


Figura 1 - a) dados plotados em gráfico de pontos; b) dados plotados em gráfico de linha (Nielsen, 2019)

Outro conceito importante destacado por Antunes e Cardoso (2015) é o de associação. É observado que na figura 2, o pico de mortalidade se deu no ano de 1918, indicando que houve um agravante nesse ano. Dessa forma, é possível associar o evento ao surto da gripe espanhola, que se deu no mesmo ano. Nesse sentido, a compreensão do conceito de correlação também se faz necessária. A correlação basicamente é a medida da semelhança entre duas variáveis, indicando a força e a direção da relação entre duas variáveis, e é medida pelo coeficiente de correlação. Esse coeficiente é um valor entre -1 e 1, em que -1 indica uma correlação inversa perfeita, 1 indica uma correlação direta perfeita, e 0 indica que não existe nenhuma correlação. Existem formas diferentes de realizar esse cálculo, considerando o contexto e a natureza dos dados analisados. Os métodos mais comuns são o de Pearson, de Spearman e de Kendall.



Figura 2 - Coeficiente de mortalidade infantil em São Paulo - SP entre 1900 e 1994 (Antunes, 1997)

O coeficiente de Pearson é utilizado para medir a correlação linear existente entre séries temporais contínuas e normalmente distribuídas. Já o coeficiente de Spearman é usado na medição da correlação monotônica, ou seja, relação de crescimento ou decréscimo entre as variáveis, em dados contínuos e que não apresentam distribuição normal. Entretanto, conforme mostrado por Horvatic, Stanley e Podobnik (2011), no caso em que há uma sequência de correlações monotônicas que ocorrem apenas quando um limite é excedido em uma distribuição não normal, o uso do coeficiente de Pearson se mostra mais eficiente que o de Spearman. Por fim, é concluído que quando as distribuições de probabilidade dos dados são muitas, é vantajoso utilizar ambos os métodos. As equações 1 e 2 apresentam as fórmulas utilizadas para os cálculos do coeficiente de Pearson (r) e de Spearman (ρ), respectivamente.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (1)$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2)$$

Em que x_i e y_i e \bar{x} e \bar{y} são os i -ésimos valores e as médias de cada variável x e y , respectivamente. n é o número de observações, e d_i é a diferença entre os ranqueamentos dos pares correspondentes.

O coeficiente de correlação de Kendall, assim como o coeficiente de correlação de Spearman é indicado para uso em distribuições de dados não gaussianas. A diferença reside na maior robustez do coeficiente de Kendall, exceto em casos em que há empates no ranqueamento estatístico dos dados (Puth, Neuhauser e Ruxton, 2015). Esse

ranqueamento é utilizado tanto no coeficiente de Spearman quanto de Kendall, e pode se basear em diferentes critérios. No caso de Spearman, ele consiste na ordenação crescente dos dados. Já no coeficiente de Kendall, ele se baseia no grau de concordância entre pares de valores das variáveis, de forma que seu cálculo pode ser expresso considerando os números de pares concordantes e discordantes, conforme a equação 3. Essa concordância é dada pela correlação monotônica entre as variáveis. Um par é dito concordante se as variáveis aumentam ou diminuem juntas, e é discordante se as variáveis têm comportamentos opostos.

$$\tau = \frac{n_c - n_d}{n_{total}} \quad (3)$$

Em que n_c é o número de pares concordantes, n_d é o número de pares discordantes e n_{total} é o número total de pares.

Além da correlação entre duas variáveis distintas, também é possível calcular a autocorrelação, ou seja, a correlação de uma série temporal consigo mesma com uma defasagem no tempo. Essa análise é importante na formulação de modelos estatísticos, permitindo a avaliação da presença de tendências e sazonalidade. A função de autocorrelação (ACF) indica se os valores atuais dependem dos valores anteriores na série temporal (Feigelson, Babu e Caceres, 2018). Além disso, tem-se a função de autocorrelação parcial (PACF) mede a correlação direta entre uma observação em um dado momento e outra anterior, eliminando a influência dos valores intermediários (Nielsen, 2019). A autocorrelação entre os valores de uma série temporal pode ser verificada numericamente por meio do teste de Ljung-Box, que considera as hipóteses por meio do cálculo da equação 4, se baseando nas autocorrelações amostrais a cada k intervalos (ρ_k) e no número de observações. Esse cálculo gera um p-valor, que, se for maior que 0,05, a hipótese nula é aceita.

H_0 : não há autocorrelação

H_1 : há autocorrelação

$$Q = n(n + 2) \sum_{k=1}^m \frac{\rho_k^2}{n-k} \quad (4)$$

Em que m é o número de defasagens considerada no cálculo.

Partindo desse conceito, surgem os modelos autorregressivos, que são modelos estatísticos usados na predição de séries temporais a partir de medições anteriores.

Feigelson, Babu e Caceres (2018) avaliaram esses modelos na análise da variação do brilho de corpos celestes, concluindo que sua utilização permite a identificação de dados inconsistentes, identificação de pontos de mudança, caracterização de processos de memória longa, classificação de séries temporais e compreensão de processos astrofísicos.

Entretanto, a correlação nem sempre está associada à causalidade. A página *Spurious Correlations* (Vigen, 2025) é conhecida por mostrar correlações entre variáveis aleatórias, sem nenhuma relação de causalidade. A figura 3a, por exemplo, mostra a correlação entre a distância entre o planeta Mercúrio e a Lua e pesquisas por “*Minecraft*” no *Google*, e a 3b correlaciona a produção de energia renovável no Brasil com o consumo de iogurte per capita nos Estados Unidos. O conceito de causalidade em séries temporais indica uma precedência temporal que visivelmente não existe nos exemplos citados.

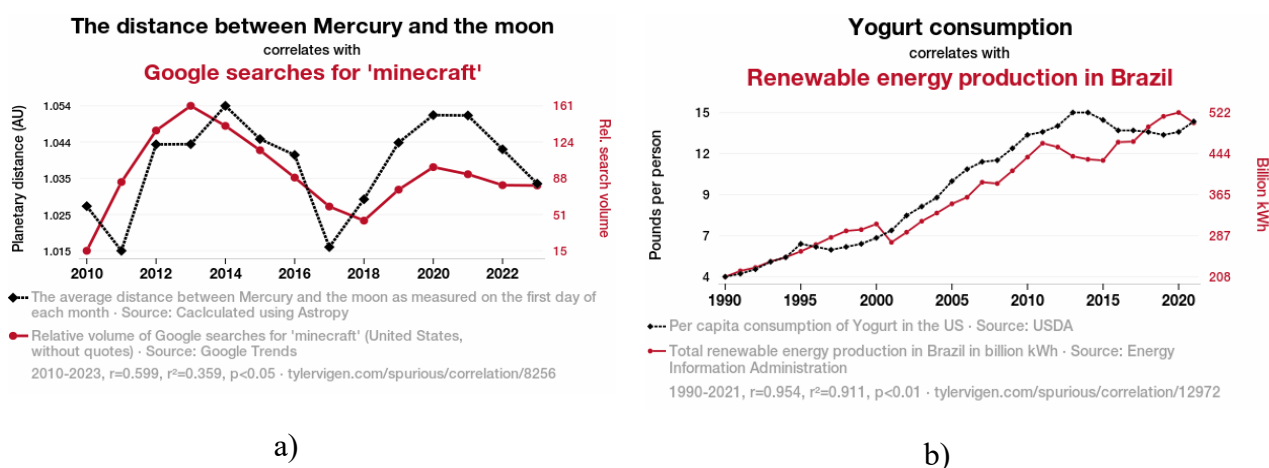


Figura 3 - a) correlação entre a distância entre Mercúrio e a Lua e pesquisas do Google por "minecraft"; b) Correlação entre o consumo de iogurte e a produção de energia renovável no Brasil (Vigen, 2025)

Além disso, é importante verificar a homoscedasticidade das séries temporais, ou seja, se os dados têm variância constante. Isso pode ser feito por meio do teste de Goldfeld-Quandt, dado pela equação 5, que considera que os dados têm variância constante para a hipótese nula. Se o valor obtido for maior que 0,05, aceita-se essa hipótese. Para aplicação desse teste, os dados são divididos em grupo para análise da variância.

$$F = \frac{var_{max}}{var_{min}} \quad (5)$$

Em que var_{max} é a variância no grupo em que há maior variância, e var_{min} é a variância no grupo com menor variância.

2.2 SÉRIES TEMPORAIS HIDROLÓGICAS

O estudo das séries temporais hidrológicas é importante para a gestão de recursos hídricos e na previsão e mitigação de eventos extremos. Dessa forma, para evitar erros de análise se baseando em falsas suposições acerca das características da série de dados estudada, é imprescindível considerar características comuns das séries hidrológicas, de forma a seguir os procedimentos adequados (Machiwal e Jha, 2012).

Nesse sentido, quanto às medidas estatísticas, os autores apontam a média aritmética e a mediana como as medidas de localização mais utilizadas. Entretanto, a média se mostra mais sensível a dados inconsistentes e a valores extremos, exigindo o tratamento dos dados. A mediana, por outro lado, é tida como uma medida mais robusta, sendo pouco afetada pela magnitude de uma única medição, logo, é mais indicada para dados com alta variabilidade, como os dados hidrológicos. Já no caso de medidas de dispersão, entre as medidas mais robustas estão o intervalo interquartil e o desvio absoluto mediano (MAD), sendo mais indicadas do que o desvio padrão para esse caso. Segundo Helsel e Hirsch (2002), nesse tipo de série de dados é comum observar a presença de assimetria, de forma que propriedades estatísticas baseadas somente em média e desvio padrão não são suficientes para os estudos relacionados à gestão de recursos hídricos, já que esses valores podem não corresponder às propriedades da maioria dos dados. Além disso, os autores apontam que testes paramétricos como o teste de correlação de Pearson podem não ser aplicáveis aos dados hidrológicos, uma vez que esses dados comumente não são normalmente distribuídos.

A normalidade das séries de dados hidrológicos pode ser testada graficamente, por meio do gráfico de distribuição, verificando se a curva é simétrica e em forma de sino, e pelo teste de Kolmogorov-Smirnov. Esse teste consiste no cálculo de uma função de distribuição acumulada empírica, $S(x)$, construída a partir dos dados observados, e na sua comparação com uma função de distribuição acumulada teórica $F(x)$. O p-valor desse teste indica a probabilidade de obter o valor correspondente à maior diferença encontrada

entre as duas funções (D_{\max}). Assim, os dados são tidos como normalmente distribuídos se $p\text{-valor} \geq 0,05$.

A assimetria da distribuição das séries temporais está ligada ao conceito de irreversibilidade. Uma série temporal é dita irreversível se suas propriedades estatísticas não permanecem iguais caso o tempo seja invertido. Segundo Müller, Schütze e Bárdossy (2017), essa característica é negligenciada, geralmente sendo desconsiderada em modelos que descrevem fenômenos hidrológicos, o que resulta em modelos enviesados. Adicionalmente, os autores constataram que assimetria apresenta variação sazonal, sendo mais pronunciada em períodos de chuva.

Além desses fatores, é importante verificar a estacionariedade da série temporal. No geral, séries hidrológicas são não-estacionárias, com presença de tendências e sazonalidade, uma vez que há períodos definidos de chuva e de seca que se repetem anualmente, por exemplo. Segundo Barros *et al.* (2021), o primeiro passo para identificação desses padrões é a inspeção visual dos dados, que facilita a escolha de testes de significância estatística adequados, uma vez que essas características por vezes podem ser facilmente identificáveis por meio da observação do gráfico da série temporal. Essa inspeção também se dá de forma mais minuciosa por meio do gráfico ACF, considerando que para um processo não estacionário percebe-se as seguintes características, ilustradas pela figura 4:

- Decaimento lento da curva;
- Comportamento senoidal;
- Valores altos em vários intervalos.

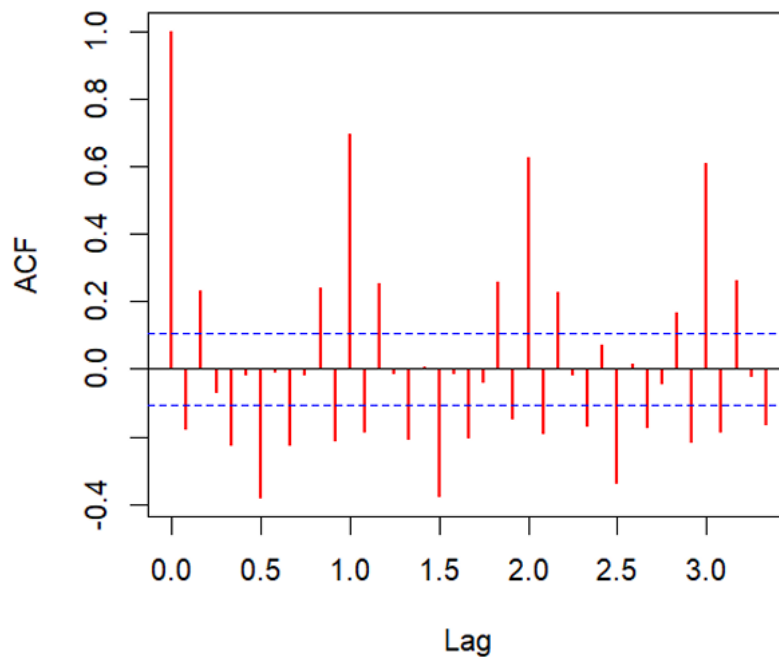


Figura 4 – ACF com características não estacionárias (Silva, 2025)

Nesse tipo de dado é comum o aparecimento de dados inconsistentes, que podem ocorrer devido a erros humanos e anomalias nos sistemas de medição. No caso em que os dados são registrados por um inspetor, a leitura dos dados pode ser influenciada pelo estado de saúde do inspetor, condições meteorológicas e dificuldade de acesso ao local de medição. Quando a leitura é feita por meio de plataformas eletrônicas de coleta de dados, observa-se problemas como perda de sincronização do sistema, alterações na marca zero na régua graduada. Além disso, transformações físicas no sistema podem levar a alterações nos dados, como a presença de galhos obstruindo a vazão de um rio (Clemmens e Whalin, 2006).

Existem inúmeros métodos estatísticos para identificação de inconsistências, dentre os quais os mais utilizados se baseiam no cálculo de limiares além dos quais os valores são tidos como inconsistentes. Esses limiares são calculados a partir de medidas estatísticas como média e desvio padrão.

Em meio à grande variedade de métodos, Gleeson *et al.* (2023) destacam a necessidade de escolher um algoritmo que seja apropriado ao comportamento e à distribuição dos dados. A maioria das técnicas utilizadas assume que os dados são normalmente distribuídos, entretanto, podem apresentar resultados precisos em dados não-gaussianos. Isso é notado em métodos consolidados como o z-score e a regra do intervalo interquartil, que foram comparados por Cornelio *et al.* (2025a) no contexto

da identificação de inconsistências em dados de cota, e obtiveram eficiência próxima de 100%. Entretanto, quando não existe essa possibilidade, é possível adaptá-los por meio de ajustes simples nos cálculos dos limiares, conforme Kim *et al.* (2022) e Jamshidi *et al.* (2021), que propuseram variações do z-score para aplicação em dados hidrológicos.

No caso de estações próximas do mar, também se faz necessário considerar a influência das marés, que impactam diretamente os dados de cota do rio. Entretanto, a depender da vazão, essa influência pode ser notada em diversas estações fluviométricas ao longo do rio. Considerando essa necessidade, Cornelio *et al.* (2025b) analisaram a aplicação de métodos numéricos e estatísticos para a identificação da influência da maré em uma dada estação, concluindo que a Transformada de Fourier e a ACF são eficazes na identificação do fenômeno e que ele é mais facilmente visualizado em rios da alta vazão.

2.3 PREVISÃO DE SÉRIES TEMPORAIS

Conforme descrito por Morettin e Tolo (2006), todos os métodos de previsão se resumem a diferentes procedimentos computacionais de cálculo de previsão de mínimos quadrados de um valor futuro considerando combinações lineares de valores passados.

Um modelo de regressão permite estabelecer uma relação linear entre a variável de previsão y e uma única variável preditora x , tentando minimizar as diferenças entre os valores previstos pela função preditora e os valores medidos. Quando há mais de uma variável preditora, trata-se de uma regressão linear múltipla, que é descrita conforme a equação 6 (Hyndman e Athanasopoulos, 2021).

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t \quad (6)$$

Em que:

β_0 : intercepto

β_1, β_2 : coeficiente angular

ε : termo de erro

Os coeficientes β são estimados a partir dos dados analisados. A aplicação do método dos Mínimos Quadrados escolhe esses valores minimizando a soma dos erros quadrados conforme apresentado na equação 7.

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_n x_{n,t})^2 \quad (7)$$

Conforme Hyndman e Athanasopoulos (2021), no modelo de regressão múltipla, os dados de interesse são previstos por meio de uma combinação linear de variáveis preditoras, já no caso de modelos autorregressivos, esses dados são previstos por uma combinação linear de valores passados do mesmo conjunto de dados. Um modelo autorregressivo de ordem p, chamado AR(p), é descrito conforme a equação 8, em que c é constante.

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon \quad (8)$$

Os modelos de médias móveis, por sua vez, utilizam os erros de previsão passados, seguindo a equação 9. Um modelo de médias móveis de ordem q é chamado MA(q). Cada valor de y_t pode ser considerado uma média móvel ponderada dos últimos erros de previsão.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (9)$$

A combinação de diferenciação com autorregressão e um modelo de médias móveis origina o modelo ARIMA(p,d,q), em que d se refere ao grau da primeira diferenciação envolvida, considerando que pode haver mais de uma diferenciação. Esse modelo é expresso na equação 10 (Morettin e Tolo, 2006).

$$y'_t = c + \varphi_1 y'_{t-1} + \dots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (10)$$

A observação da ACF e da PACF, conforme Barros *et al.* (2021) é importante para se ter uma ideia inicial dos parâmetros que minimizem os critérios de informação Akaike e Bayesiano, que são mostrados nas equações 11 e 12 (Akaike, 1973; Akaike, 1978). O valor de p corresponde ao número de defasagens com picos antes de entrar na faixa de valor crítico no gráfico da PACF; O q é determinado da mesma forma, mas observando o gráfico da ACF. O parâmetro d é definido pela quantidade de vezes em que é necessário diferenciar a série temporal para que ela seja estacionária.

$$AIC = -2 \log L + 2(p + q + k + 1) \quad (11)$$

$$BIC = AIC + [\log(n) - 2](p + q + k + 1) \quad (12)$$

Em que L é a verossimilhança dos dados, e k = 1 se $c \neq 0$, e k = 0, se $c = 0$.

Para o caso de séries sazonais, utiliza-se o modelo SARIMA(p,d,q)(P, D, Q)_s, que inclui o S referente à sazonalidade na sigla. Os valores p, d e q se referem à ordem do polinômio não sazonal, enquanto P, D e Q são os equivalentes na parte sazonal do modelo, o s subscrito representa o tempo de repetição dos eventos sazonais, o B é o termo de defasagem, Θ é o termo de médias móveis referente à parte sazonal e Φ é o parâmetro do modelo autorregressivo da parte sazonal. Esse modelo é expresso na equação 13, considerando $s = 12$ (Morettin e Tolo, 2006; Barros *et al.* 2021).

$$\varphi(B)\Phi(B^{12})(1 - B^{12})^D(1 - B)^d y_t = \theta(B)\Theta(B^{12})\varepsilon_t \quad (13)$$

Segundo Barros *et al.* (2021), o uso de variáveis auxiliares pode aperfeiçoar a modelagem e a previsão, e se baseiam na relação causal entre as séries temporais. O acréscimo de variáveis auxiliares ao modelo SARIMA origina o modelo SARIMAX.

Esse método consiste em três etapas iterativas: identificação, estimação de parâmetros e verificação. A etapa de identificação busca os valores adequados para p, d e q, correspondentes à parte estacionária do modelo. Isso se dá com a análise dos gráficos de ACF e PACF e testes estatísticos. O processo de estimação, por sua vez, estima os parâmetros autorregressivos e de médias móveis por meio de um processo iterativo de MMQ. Por fim, a verificação considera os critérios AIC e BIC e métricas de erro para avaliação dos resíduos entre os valores previstos e observados (Araújo *et al.*, 2023).

Os modelos autorregressivos supõem que os resíduos entre os valores previstos e observados se comportam como ruído branco, ou seja, apresentam média igual a zero, independência entre os valores, variância constante e distribuição normal, sendo necessário verificar essas características na avaliação dos resultados do modelo de previsão.

3 METODOLOGIA

3.1 ÁREA DE ESTUDO

A área de estudo considerada é a da região hidrográfica do Tocantins-Araguaia, considerada a de maior área de drenagem totalmente contida no território brasileiro, com área total de 918.822 km², e passando pelas regiões Centro-Oeste, Norte e Nordeste. A região apresenta pouca variabilidade climática ao longo do ano, e tem precipitação anual média de 1.774 mm. Na figura 5, observa-se que os índices pluviométricos são crescentes no sentido do sul para o norte e decrescentes de oeste para leste, conforme mostrado. O período chuvoso na região se dá nos meses de novembro a março (ANA, 2009).

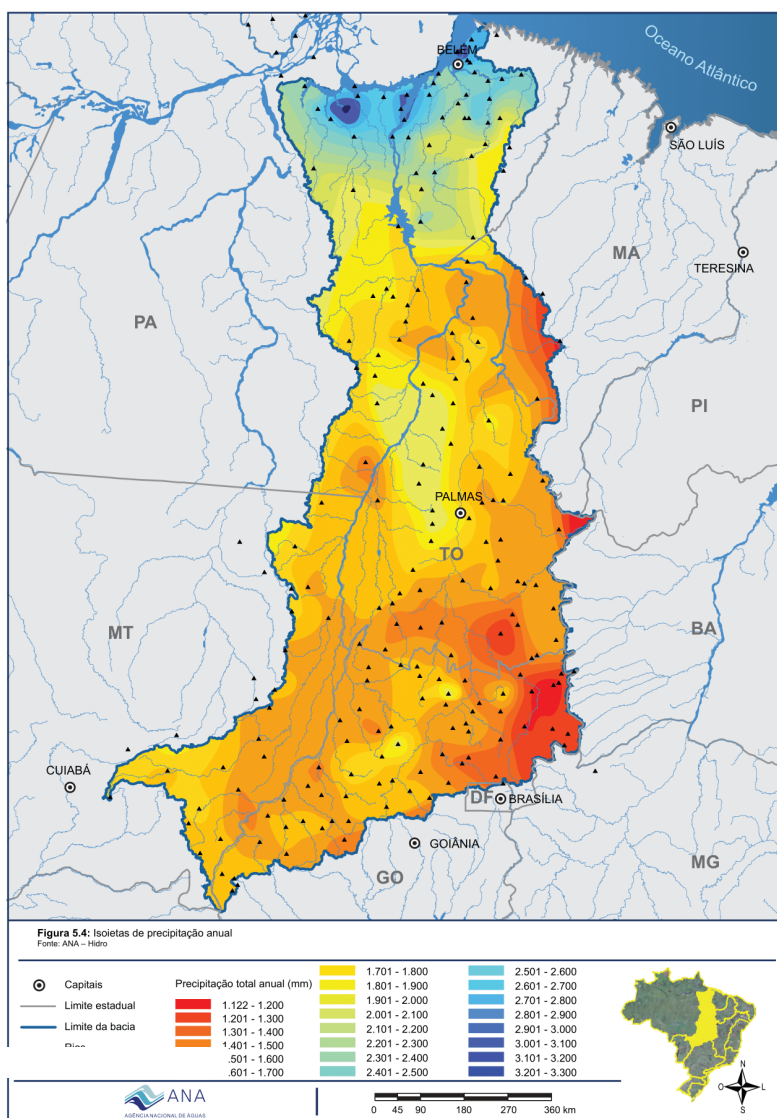
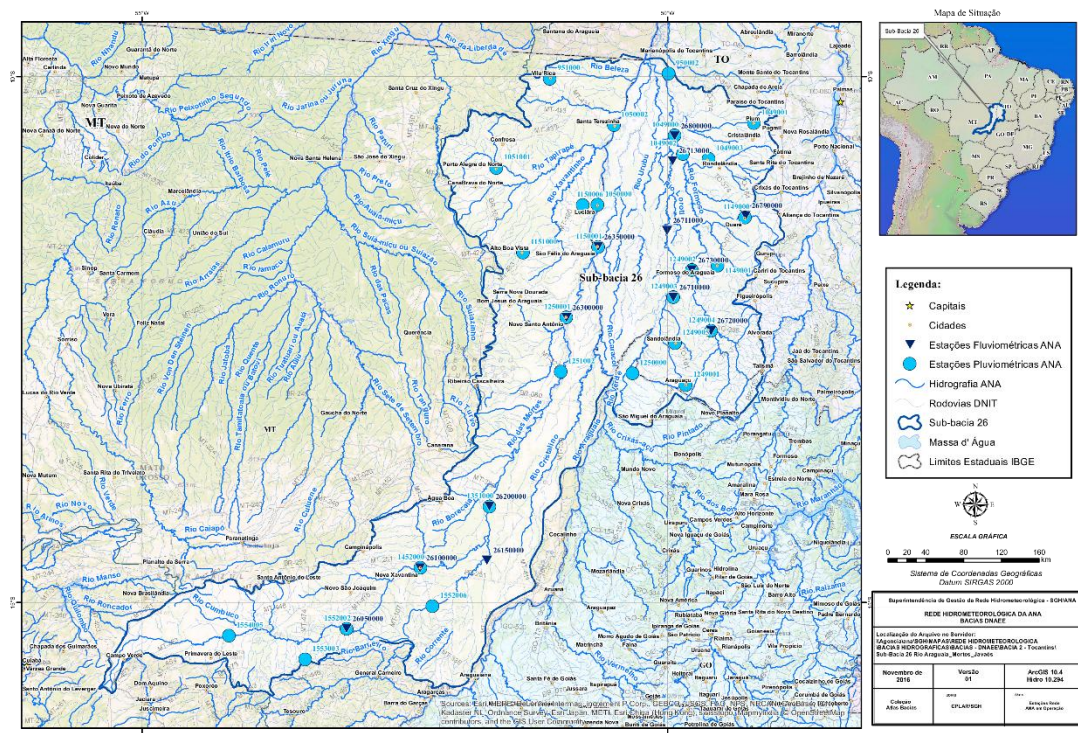


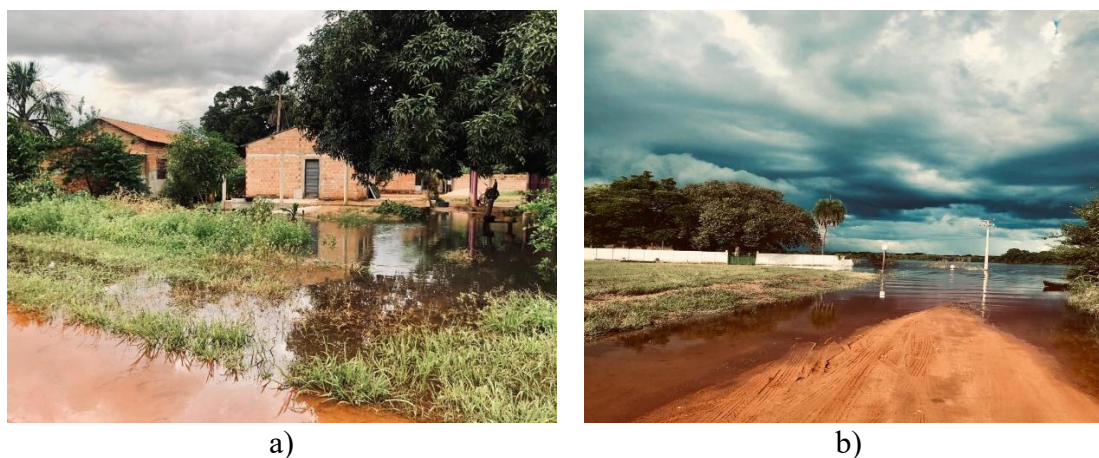
Figura 5 - Isoietas de precipitação anual (ANA, 2009)



b)

Figura 6 – a) Sub Bacia Hidrográfica 22 (Tocantins entre os rios Paranã e Sono); b) Sub Bacia hidrográfica 26 (trecho da ilha do bananal) (ANA, 2016)

São Félix do Araguaia sofre de forma recorrente com enchentes devido às cheias do rio Araguaia na estação chuvosa, chegando a subir quase 9 metros no ano de 2022 com fortes chuvas do mês de janeiro, alagando moradias nas margens, e obrigando a sua evacuação, conforme ilustra a figura 7. Nesse contexto, a Defesa Civil atua em conjunto com a prefeitura para dar suporte à população e buscar medidas de mitigação (Diário Araguaia, 2022).



a)

b)

Figura 7 – Cidade de São Félix do Araguaia - Mato Grosso em estado de emergência devido à cheia no rio Araguaia (Diário Araguaia, 2022)

Evidenciando o papel dessa ação conjunta auxiliada pela supervisão dos dados, em 2025 foi noticiado o aumento do monitoramento, de avaliações meteorológicas e de análise da vazão do rio na cidade, que passaram a ser conduzidas diariamente como forma de auxiliar a população diante do aumento das chuvas observado no mês de fevereiro. Diante da situação, a Defesa Civil também ampliou a sinalização de áreas de risco e reforçou orientações e medidas preventivas junto às comunidades ribeirinhas. Além disso, para o reforço da segurança, equipes especializadas foram colocadas de prontidão para resgates e suporte emergencial a possíveis desalojados. Para enfrentar o período crítico, a prefeitura garantiu a distribuição emergencial de cestas básicas, kits de higiene e colchões (Agência da Notícia, 2025).

3.2 MÉTODO

O procedimento experimental se divide em duas etapas: validação do modelo e aplicação em um estudo de caso. A validação é realizada com a aplicação do MMQ conforme Monteiro *et al.* (2024) na UHE de Estreito, e comparando esses resultados com os obtidos com o uso do SARIMAX no mesmo contexto. Em seguida, utiliza-se o SARIMAX na previsão de vazão e cota do rio na região de São Félix do Araguaia. Os procedimentos para a análise consistem nos seguintes passos, apresentados na figura 8, utilizando a linguagem de programação python no ambiente jupyter notebook:

- 1) Importação dos dados hidrológicos utilizados para o ambiente jupyter notebook e QGIS;
- 2) Cálculo da precipitação mensal média a partir dos dados coletados pelo conjunto de satélites IMERG pelo Método de Thiessen (1911) no QGIS;
- 3) Cálculo da vazão mensal por meio da precipitação para a região estudada, considerando a área total;
- 4) Identificação e correção de inconsistências com base no método do Intervalo Interquartilico, conforme Tukey (1977);
- 5) Análise da correlação entre a precipitação e vazão medida na estação de medição de referência considerando até quatro meses de defasagem;
- 6) Aplicação do método de previsão;
- 7) Análise dos resultados.

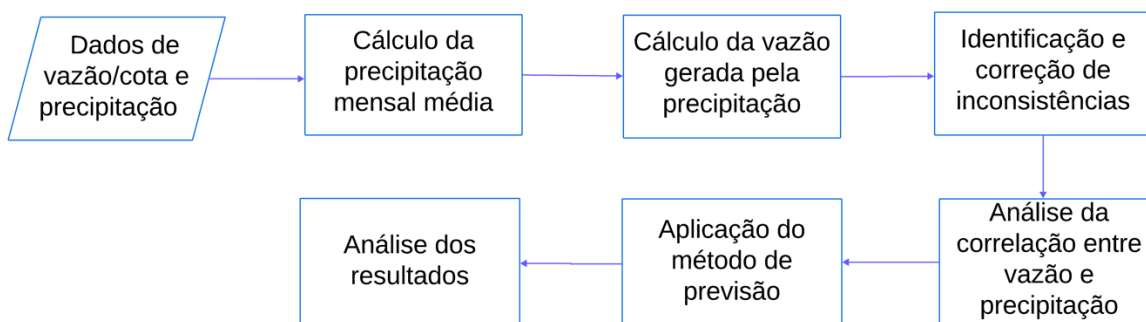


Figura 8 - Fluxograma do trabalho

No passo 1, os dados de vazão e cota são disponibilizados pela Agência Nacional de Águas, e os dados de precipitação são obtidos pelo conjunto de satélites IMERG v.7, da NASA. Utiliza-se dados de vazão afluente na UHE do Estreito e na UHE de Lajeado, a montante, no período entre março de 2014 e dezembro de 2021, tendo início no começo do período da seca na região. Os dados até 2020 são utilizados para calibração dos modelos, e os demais são usados na validação. A fase de calibração do modelo tem o objetivo de encontrar os parâmetros que melhor definem o comportamento dos dados. Realizada em seguida, a validação consiste na aplicação do modelo calibrado para verificar a precisão das previsões.

Os dados são analisados para compreensão das suas características como distribuição e tendências para definir a melhor abordagem e os parâmetros ideais dos modelos. Em dados hidrológicos, é comum que haja tendências e sazonalidade de acordo com o período de chuvas, além de distribuição não-gaussiana.

Por sua vez, os dados de precipitação são disponibilizados pela NASA no endereço <<https://giovanni.gsfc.nasa.gov/giovanni/>>, buscando pelos dados de precipitação acumulada por mês para cada mês no período de análise na região estudada. No passo 2, esses dados são tratados no sistema de informação geográfica QGIS, em que se aplica o método de Thiessen (1911) para cálculo da precipitação média. Esse método consiste na aplicação do diagrama de Voronoi para divisão da região e no cálculo da média ponderada dos valores de precipitação, conforme a equação 14.

$$P_m = \frac{\sum PRP_i A_i}{A_{total}} \quad (14)$$

Em que P_m é a precipitação média na bacia, PRP_i é a precipitação no polígono i , A_i é a área do mesmo polígono e A_{total} é a área da sub bacia.

No passo 3, realiza-se o cálculo da vazão média mensal em m^3/s por meio dos dados de precipitação segue a equação 15. Esse cálculo parte do pressuposto que toda a precipitação se converte em vazão, e dividindo o volume de chuvas pela quantidade de dias do mês, tem-se a precipitação mensal. O numerador da equação $P_m * A_{total}$ corresponde ao volume total de precipitação durante o mês. A constante 2592 combina a média do número de dias de cada mês, 30, e a conversão da unidade de medida para m^3/s .

$$Q_{cal} = \frac{P_m * A_{total}}{2592} \quad (15)$$

Antes de prosseguir para a estimativa, se faz necessário garantir a consistência dos dados no passo 4. Para tanto, realiza-se o processo de identificação de dados inconsistentes utilizando o método do intervalo interquartílico, que consiste no cálculo de limites fora dos quais os dados são tidos como inconsistentes. Esse cálculo se baseia nas diferenças entre os quartis dos dados, conforme as equações 16 e 17 (Tukey, 1977). As inconsistências detectadas são, então, substituídas por valores das médias locais, de forma a não destoarem dos valores vizinhos. O valor da média local é escolhido para correção dos dados, já que a exclusão dos dados inconsistentes poderia comprometer a precisão dos métodos de previsão, que se baseiam em dados anteriores para fazer as estimativas, e a substituição pela média global não é indicada devido ao comportamento sazonal característico das séries de dados fluviométricas, já que esse valor pode destoar dos valores vizinhos ao valor que deve ser substituído, de forma que sua substituição pela média global ainda pode representar uma inconsistência.

$$lim_{inf} = Q_1 - 1,5IQR \quad (16)$$

$$lim_{sup} = Q_3 + 1,5IQR \quad (17)$$

Em que Q_1 e Q_3 correspondem, respectivamente, ao primeiro e ao terceiro quartis, e o intervalo interquartílico é expresso por $IQR = Q_3 - Q_1$.

O valor 1,5 é utilizado com base na distribuição normal dos dados, de forma que cerca de 99% dos dados sejam incluídos no intervalo calculado. Embora seja definido para dados gaussianos, o método do intervalo interquartílico é consolidado e apresenta resultados precisos também em dados não-gaussianos.

O diagrama de caixas mostrado na figura 9 ilustra o funcionamento desse método, que pode ser implementando apenas na forma gráfica também. Como o nome sugere, esse diagrama consiste em uma forma retangular, que representa o intervalo interquartil,

delimitada pelo primeiro e pelo terceiro quartis, e em que os dados discrepantes são representados como círculos além da caixa. Os quartis são medidas que dividem o conjunto ordenado de dados em quatro partes iguais, sendo que:

- O primeiro quartil corresponde ao valor abaixo do qual se encontra 25% dos dados;
- O segundo quartil é a mediana;
- O terceiro quartil delimita o valor abaixo do qual está 75% dos dados;
- Os “bigodes” do gráfico são as linhas que se estendem antes do primeiro quartil e após o terceiro, indicando os limites calculados nas equações 16 e 17;
- Os pontos coloridos além dos valores mínimo e máximo indicam os valores inconsistentes.

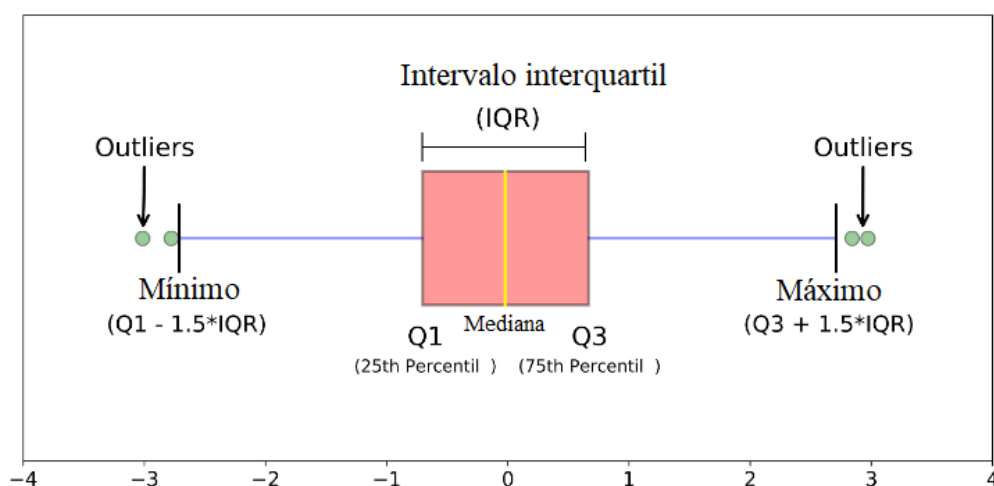


Figura 9 - Diagrama de caixas (Trøndheim e Nielsen, 2022)

Para o passo 5, considera-se que a precipitação não impacta os valores de vazão instantaneamente, uma vez que parte da chuva é infiltrada ou não cai diretamente no rio. Assim, se faz necessário calcular qual o tempo de resposta entre a precipitação e a vazão para que a estimativa seja a melhor possível. Isso ocorre por meio do cálculo do coeficiente de correlação entre a precipitação e a vazão medida no Estreito em diferentes intervalos de tempo para identificar que intervalo apresenta maior correlação. Os dados de vazão calculada são, então, defasados de acordo com esse resultado para sua aplicação aos métodos de previsão.

Finalmente, no passo 6, aplica-se o MMQ, considerando a vazão na estação a montante e a vazão calculada defasada como variáveis exógenas. A aplicação do método

se dá conforme Monteiro *et al.* (2024), para validar os resultados obtidos em seguida pelo SARIMAX.

Para a aplicação do SARIMAX, tem-se o auxílio da função `auto_arima`, da biblioteca `pmdarima python`, que calcula os parâmetros ideais do modelo de acordo com o comportamento dos dados, por meio de uma série de testes estatísticos. Em seguida, no passo 7, os resultados dos modelos são comparados entre si de forma gráfica e numérica, considerando os resíduos entre os valores previstos e observados e diferentes métricas de erro.

Validada a metodologia, o processo de previsão com SARIMAX é feito novamente para a segunda região de estudo: São Félix do Araguaia. Nesse caso, além da vazão, realiza-se a previsão da cota do rio, usando a vazão calculada não defasada e a vazão no Trecho Médio como variáveis exógenas. Nota-se que na estimativa de cota não se utiliza a defasagem, já que, diferente da vazão, essa variável é instantaneamente afetada pela precipitação.

4 RESULTADOS

4.1 VALIDAÇÃO DO MÉTODO

4.1.1 Aplicação dos Métodos de Previsão

Os dados de vazão em Estreito e Lajeado disponibilizados pela ANA têm frequência diária, sendo agrupados em médias mensais para a aplicação aos métodos de estimação. Não houve presença de lacunas ou de valores nulos nas séries temporais.

Os dados de precipitação obtidos pelo IMERG e processados pelo Método de Thiessen (1911) são disponibilizados pelo Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM). Para garantir a autonomia da metodologia proposta, este estudo refaz esse cálculo considerando o mesmo método e as mesmas fontes de dado, utilizando o QGIS.

Os dados do IMERG são importados para cada mês no formato NetCDF, na forma de mapa de precipitação acumulada com frequência de 30 minutos. Por meio do método de Thiessen (1911), obtém-se a média mensal, que é aplicada aos modelos. Os polígonos de Thiessen gerados são mostrados na figura 10. A série de precipitação obtida é apresentada na figura 11, em que é visível o comportamento sazonal a cada ano, com dados variando de valores próximos de 0 mm até 350 mm.

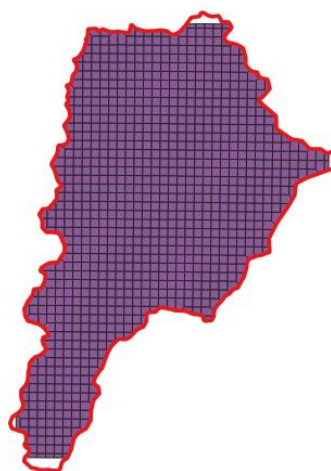


Figura 10 - Polígonos de Thiessen gerados com os dados de precipitação na região da UHE de Estreito

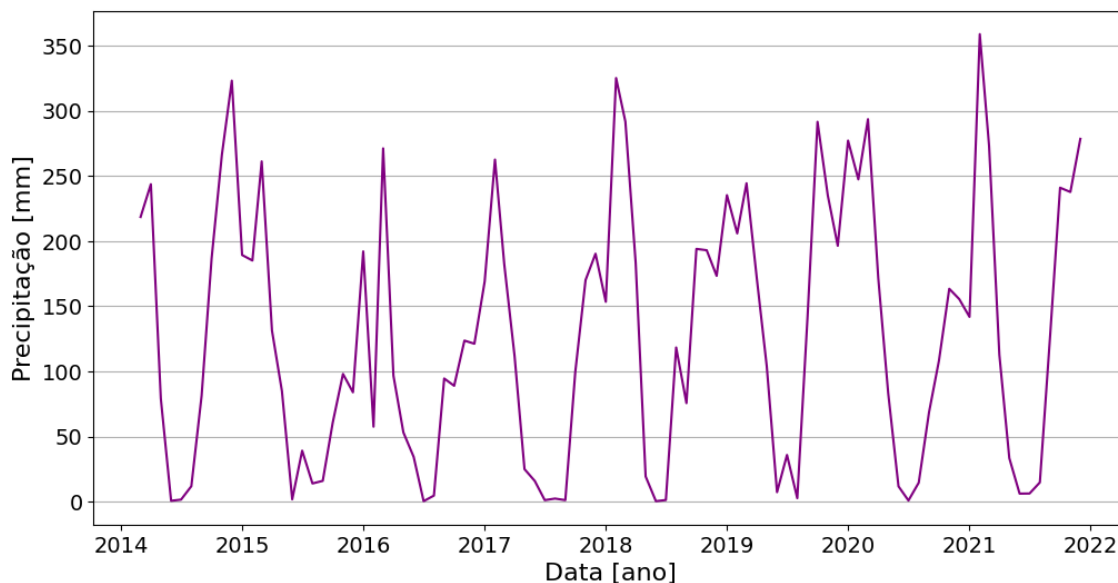


Figura 11 - Precipitação na região da UHE de Estreito

Quanto aos dados de vazão, com o auxílio do gráfico da figura 12, confirma-se que a série temporal apresenta uma distribuição não-gaussiana, com a mediana deslocada para a esquerda, e a distância entre o primeiro e o segundo quartis visivelmente menor do que a distância entre o segundo e o terceiro quartis, com a maioria das medições apresentando valores entre 1000 e 2000 m^3/s .

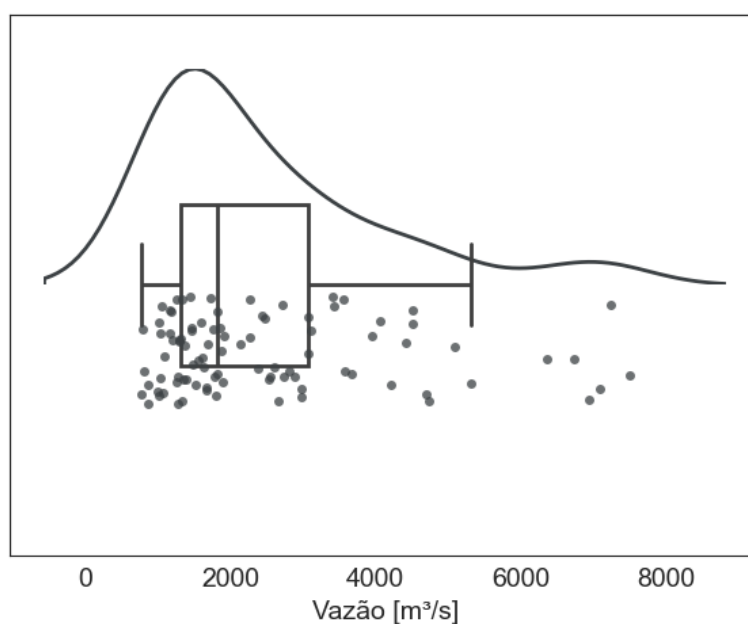


Figura 12 - Distribuição amostral dos dados de vazão do Estreito

Dessa forma, considerando a distribuição dos dados, calcula-se o coeficiente de correlação de Spearman entre os dados de vazão afluente na UHE do Estreito e a precipitação na mesma região considerando uma defasagem de até 4 meses. Pelos resultados dispostos na tabela 1, observa-se que a correlação mais forte se dá com a defasagem de 1 mês, indicando que esse é o tempo que a precipitação leva para impactar a vazão na região. Baseando-se nesse resultado, os métodos de previsão são aplicados antecipando em um mês os dados da vazão calculada.

Tabela 1 – Correlação de Spearman entre a precipitação e a vazão no Estreito

Defasagem	Coefficiente de correlação
T-4	0,111
T-3	0,501
T-2	0,783
T-1	0,824
T-0	0,603

Levando em consideração esses resultados, o Método dos Mínimos Quadrados é aplicado tendo como variáveis auxiliares: a vazão afluente na UHE de Lajeado e a vazão calculada defasada. Com os coeficientes calculados, obtém-se a equação 18.

$$y_t = 0,110x_1 + 1,737x_2 - 83,431 \quad (18)$$

Em que:

x_1 : vazão ligada à precipitação (m^3/s)

x_2 : vazão afluente no Lajeado (m^3/s)

O SARIMAX foi aplicado da mesma forma, considerando os parâmetros ideais encontrados pela função `auto_arima`, resultando no modelo SARIMAX(0, 1, 2)(2, 1, 1)₁₂. Os coeficientes obtidos pelo modelo geram a equação 19.

$$(1 - 0,00154B^{12} + 0,379B^{24})(1 - B)(1 - B^{24})y_t = (1 - 0,457B - 0,399B^2)\varepsilon_t \quad (19)$$

Por fim, os resultados dos modelos de previsão são mostrados no gráfico da figura 13. O gráfico foi plotado com a remoção da primeira e da última amostra da série temporal, anulando os efeitos extremos observados nos limites, que são um problema

comum na previsão de séries temporais, dada a ausência de dados antes ou depois desses pontos.

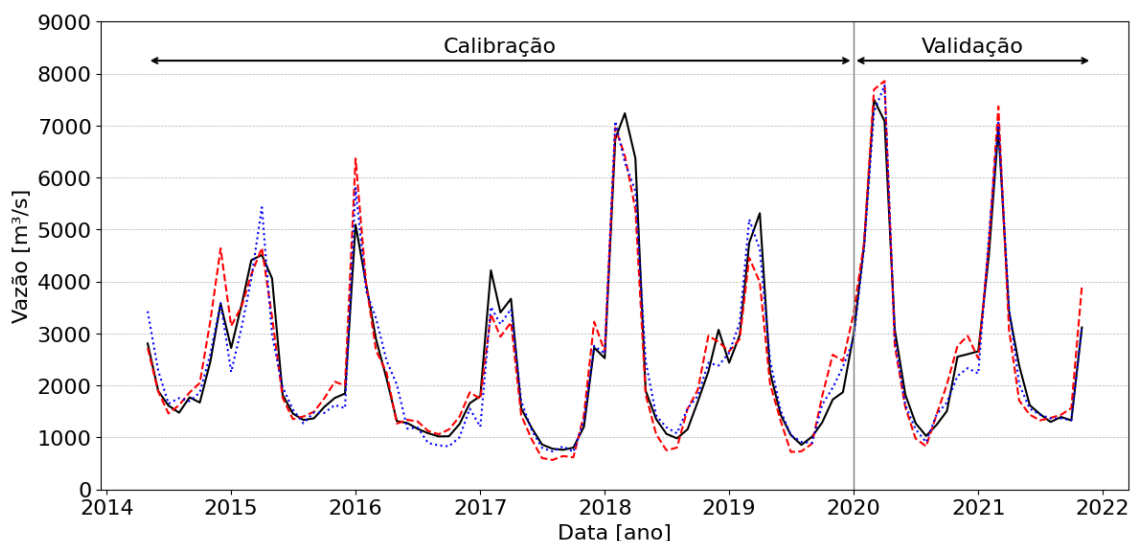


Figura 13 - resultado do modelo, em que a linha tracejada vermelha indica a previsão com Mínimos Quadrados, a linha pontilhada azul indica o resultado da previsão com SARIMAX, e a linha preta contínua são os resultados observados

Pela observação da figura 13, percebe-se que os dois modelos apresentam bons ajustes. A previsão gerada pelo SARIMAX apresenta melhores resultados nos picos de medição, região onde se encontram as maiores diferenças entre os valores previstos pelos Mínimos Quadrados e os valores reais.

4.1.1.1 Avaliação dos Resultados

Os gráficos da figura 14 apresentam a relação entre os valores previstos e os valores observados. Observa-se um comportamento linear nos gráficos, principalmente no caso da previsão gerada pelo SARIMAX, em que os pontos aparecem mais próximos. Para aprofundar essa análise, calculou-se os coeficientes de correlação de Pearson, Spearman e Kendall entre os valores observados e estimados. Os resultados, apresentados na tabela 2, mostram a forte correlação presente nos dois casos, que é ligeiramente maior na previsão do SARIMAX.

Tabela 2 - Coeficientes de correlação entre os valores observados e estimados

	MMQ	SARIMAX
Pearson	0,969	0,977

Spearman	0,964	0,966
Kendall	0,840	0,853

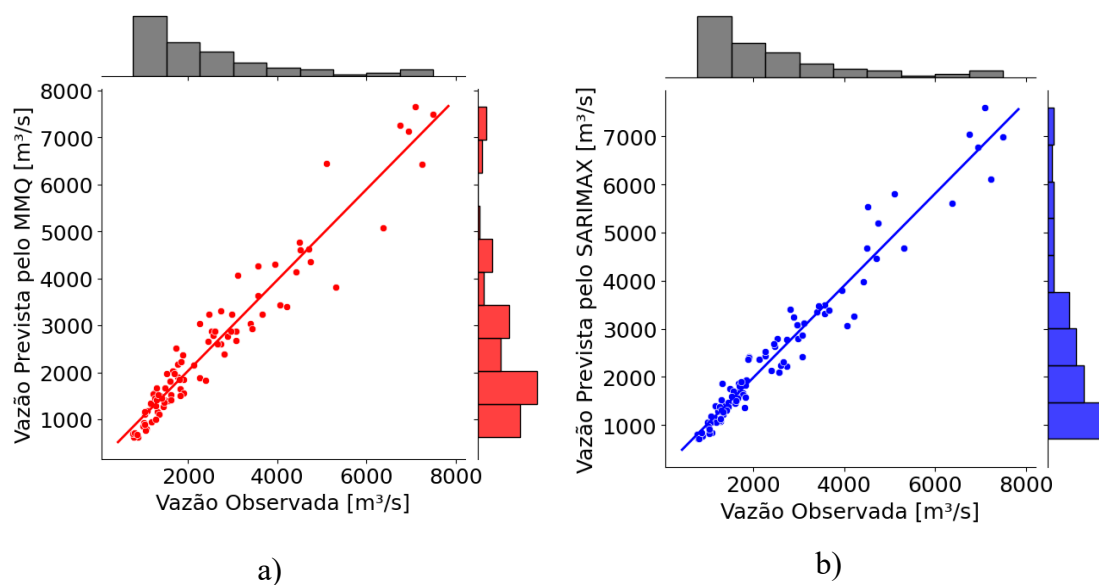


Figura 14 - Relação entre a vazão observada e a) a vazão prevista pelo Mínimos Quadrados; b) a vazão prevista pelo SARIMAX

Na parte superior dos gráficos da figura 14, nota-se, também, que os dados previstos mantêm a distribuição não-gaussiana, da mesma forma que os dados observados. Isso é evidenciado na figura 15. Os dois gráficos têm o mesmo comportamento dos dados observados.

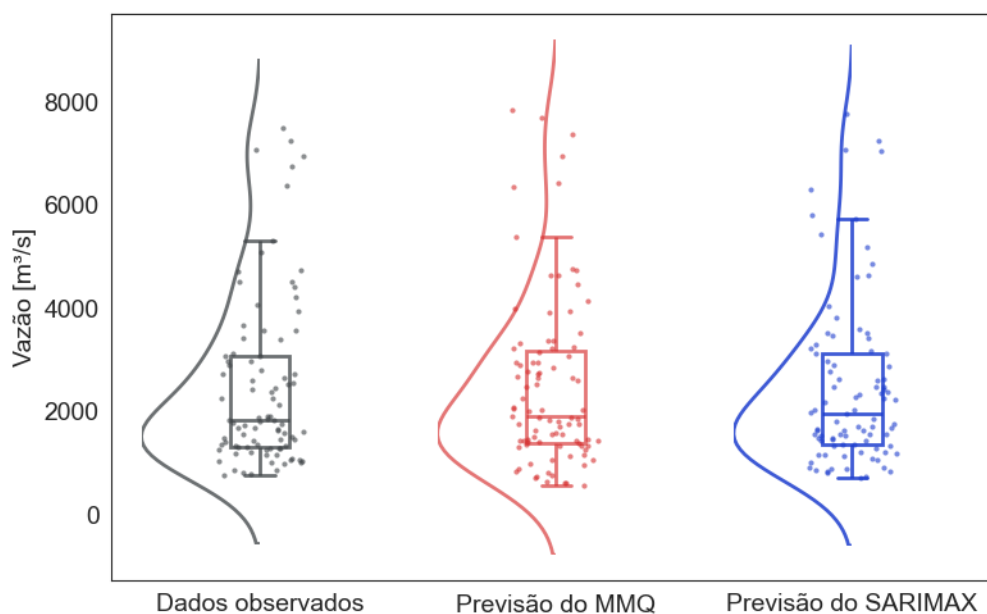


Figura 15 - Distribuição amostral dos dados observados e previstos

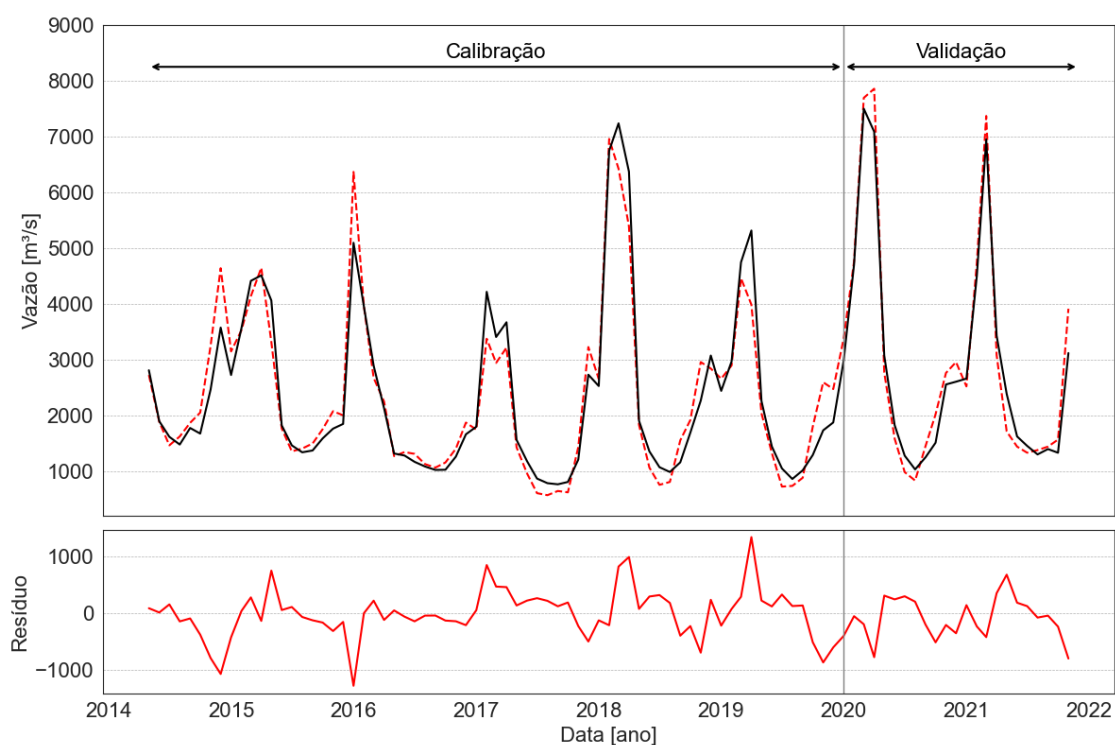
4.1.1.2 Análise dos Resíduos

A observação dos resíduos gerados é importante para avaliar a performance dos modelos. Para realizar essa análise, os resíduos foram calculados como a diferença entre os valores previstos e os valores observados, conforme a equação 20.

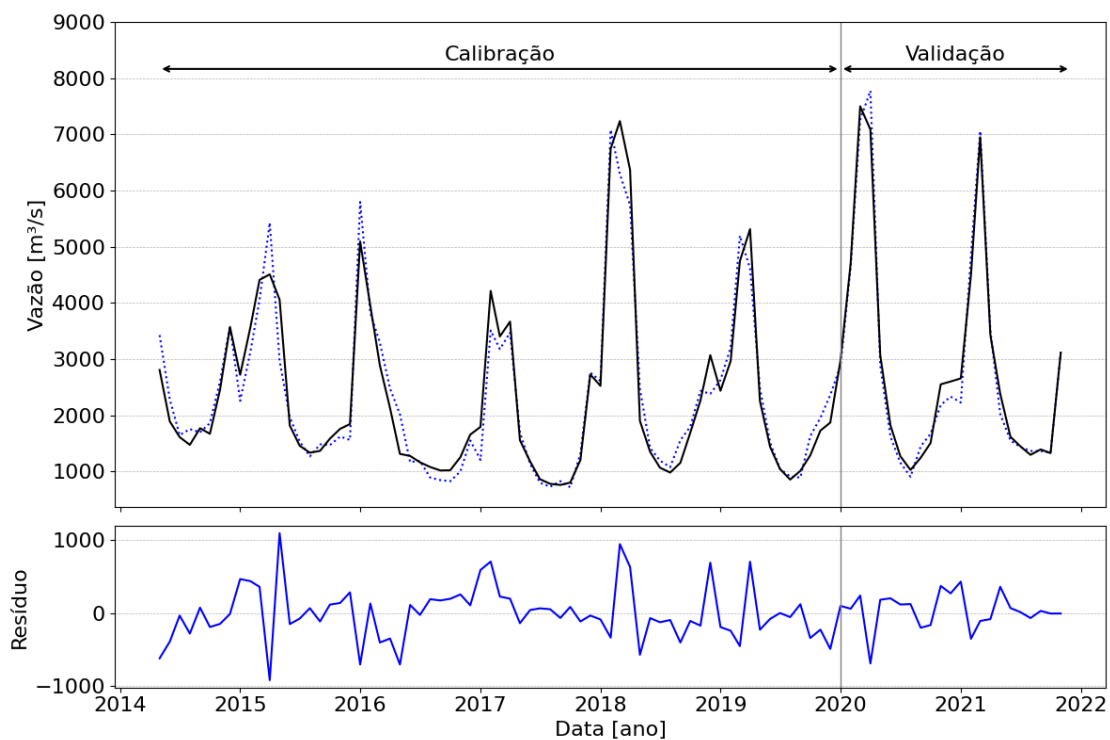
$$res = y_{obs} - y_{prev} \quad (20)$$

Em que res representa os resíduos, y_{prev} são os valores previstos e y_{obs} são os valores observados.

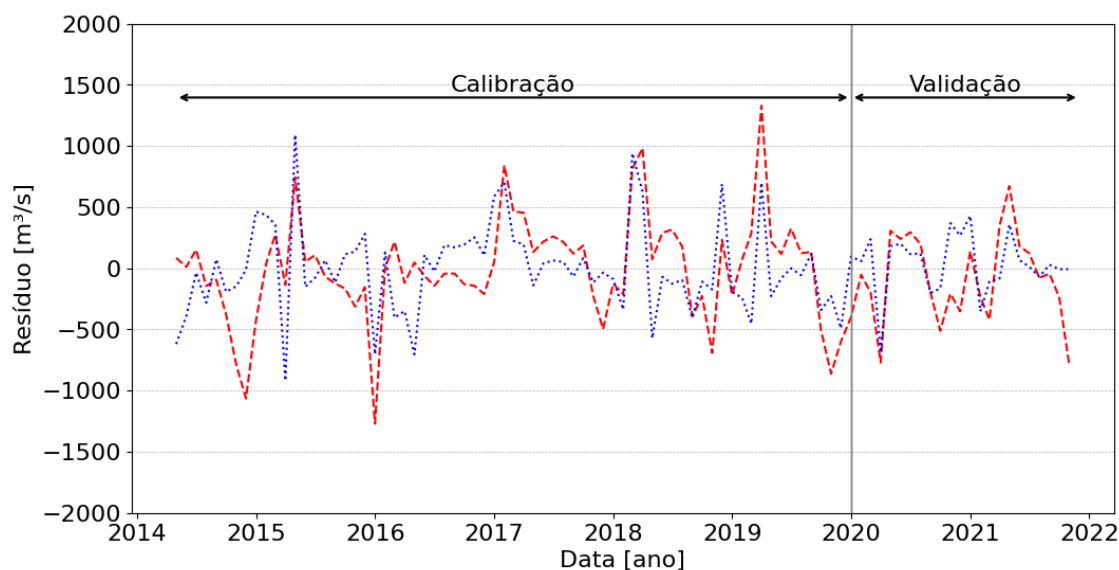
Esses valores foram plotados no gráfico da figura 16. Observa-se resultados similares entre os dois casos, no entanto, os maiores valores são verificados no modelo dos Mínimos Quadrados, representado pela linha tracejada vermelha. As figuras 16a e 16b mostram os resíduos entre os valores observados e previstos comparados com os dados observados e previstos pelo MMQ e pelo SARIMAX, respectivamente. Esses valores são comparados entre si no gráfico da figura 16c. Neste caso, os resíduos são representados pela linha traço-ponto.



a)



b)



c)

Figura 16 – a) Resíduos e previsões do modelo MMQ, em que a linha preta contínua representa os valores observados, a linha tracejada vermelha representa a previsão do MMQ, e a linha contínua vermelha são os resíduos; b) Resíduos e previsões do modelo SARIMAX, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos; c) Resíduo dos modelos, em que a linha tracejada vermelha indica os resíduos do MMQ, e a linha pontilhada azul indica os resíduos do SARIMAX.

Os modelos regressivos assumem que os resíduos se comportam como ruídos brancos, o que significa que eles têm as seguintes propriedades:

- Distribuição normal;
- São independentes, ou seja, não possuem autocorrelação;
- Média zero;
- Homocedasticidade, ou seja, variância constante.

Nesse sentido, para garantir a qualidade dos modelos utilizados, se faz necessário verificar essas características. No que diz respeito à distribuição dos dados, a figura 17 evidencia que, nos dois modelos, os resíduos se afastam ligeiramente da distribuição normal, o que é confirmado pelo teste de Komolgorov-Smirnov.

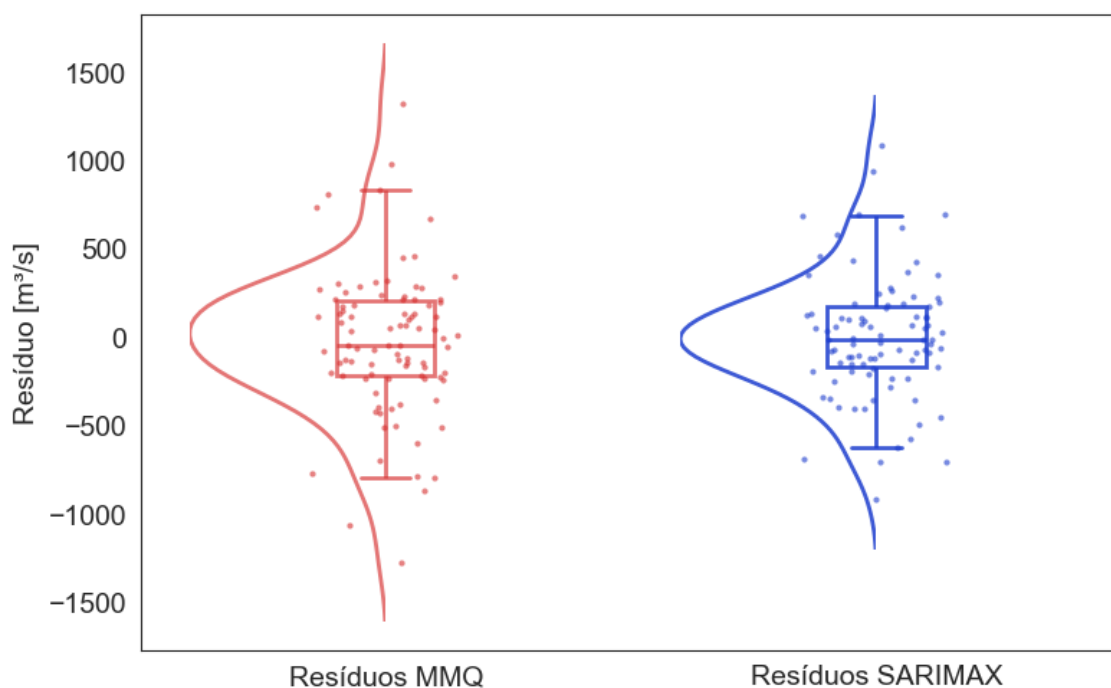


Figura 17 - Distribuição dos resíduos entre os dados observados e os previstos

A autocorrelação é verificada por meio do gráfico da ACF apresentado na figura 18. Para fazer essa análise, verifica-se se os pontos se localizam dentro do intervalo de confiança, indicado pela área colorida na figura. Caso isso aconteça, significa que os valores são independentes. Na figura, os intervalos de confiança dos dois resíduos se sobrepõem, o que se verifica na mistura das cores azul e vermelho. Isso acontece porque o tamanho dos intervalos de confiança é determinado pelo tamanho da série temporal, que é o mesmo nos dois casos. No gráfico, os valores do eixo y do gráfico indicam a defasagem dos dados em que se verifica a correlação dos dados, enquanto o eixo x mostra o coeficiente de correlação calculado. São consideradas defasagens de até 24 meses, correspondente ao período necessário para verificar a repetição do ciclo de chuvas, que

dura 12 meses. Quanto mais o valor de x se aproxima de 0, menor a autocorrelação na defasagem avaliada. Dessa forma, em $m = 0$, o conjunto de dados é comparado com ele mesmo sem nenhuma defasagem, gerando a correlação é máxima, com coeficiente igual a 1. No caso analisado, nota-se que a maioria dos pontos se encontra dentro do intervalo de confiança no caso dos resíduos do modelo gerado pelo SARIMAX, indicando que os resíduos são independentes. Isso não ocorre, entretanto, no caso dos resíduos do MMQ.

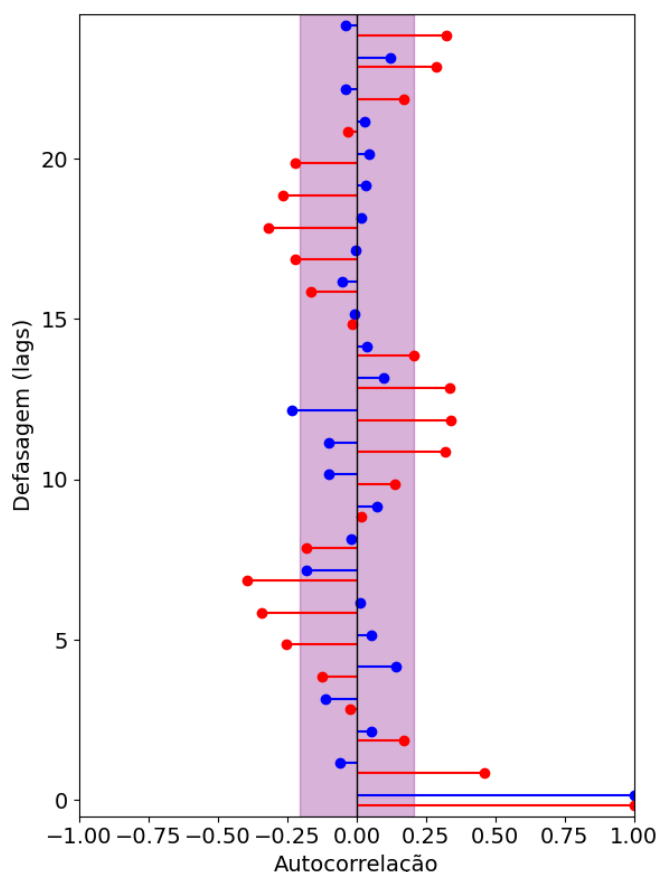


Figura 18 - ACF dos resíduos dos modelos, em que os pontos vermelhos indicam os resíduos do modelo do MMQ, e os azuis correspondem aos resíduos do modelo do SARIMAX

A independência dos resíduos também pode ser verificada por meio do teste de Ljung-Box. Se o p-valor resultante para esse teste for superior a 0,05, considera-se que não há correlação entre os resíduos. Para os casos analisados, foram obtidos os resultados apresentados na tabela 3. Nota-se, que há autocorrelação nos resíduos do modelo do MMQ em relação aos valores observados, ou seja, eles são dependentes. O contrário se observa no caso do SARIMAX, que tem resíduos independentes de acordo com o teste de Ljung-Box.

Tabela 3 - Resultados do teste de Ljung-Box para os resíduos dos modelos

	p-valor	Autocorrelação
Resíduo SARIMAX	0,1993 > 0,05	Ausente
Resíduo MMQ	$2,576 \times 10^{-13} < 0,05$	Presente

Quanto à média dos resíduos, os resultados foram próximos de zero, correspondendo a 0,76% da média das medições, no caso do Mínimos Quadrados e a 1,22% no caso do SARIMAX. A média diferente de zero indica presença de viés sistemático, que pode ser corrigido com ajustes do modelo e pode estar ligada à presença de erros de mensuração sistemáticos da variável dependente do modelo, conforme apontado por Kennedy (2002).

Já a variância dos resíduos é analisada por meio dos gráficos residuais da figura 19, que mostram os resíduos plotados em relação aos valores previstos. Com a observação da figura, nota-se que os modelos estão dispersos de forma relativamente aleatória em torno de zero, que é destacado pela linha tracejada verde. Entretanto, alguns pontos apresentam maior distância em relação ao zero, apontando para a possibilidade da existência de dados inconsistentes.

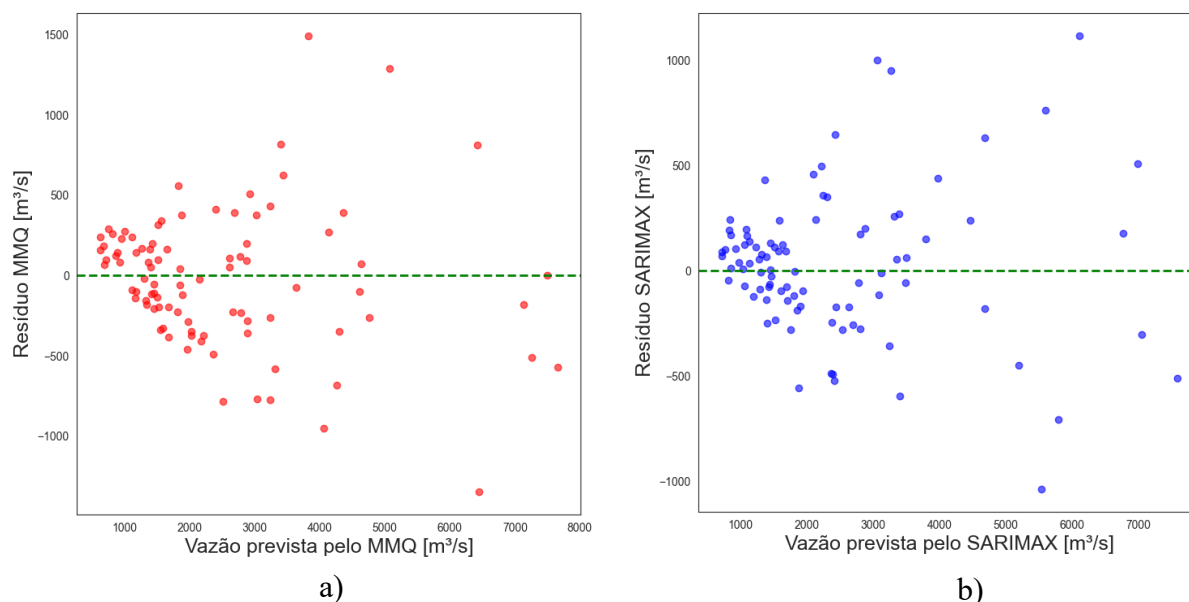


Figura 19 - Gráficos de dispersão dos resíduos pela previsão para o modelo do a) Mínimos Quadrados; b) SARIMAX

Essa característica é verificada numericamente por meio do teste de Goldfeld-Quandt, em que, se o p-valor calculado é maior que 0,05, a variância é tida como

constante. Os resultados apresentados na tabela 4 mostra que essa hipótese se confirma para os dois modelos.

Tabela 4 - Comportamento da variância dos resíduos

	p-valor	Variância
Resíduo SARIMAX	0,666 > 0,05	Constante
Resíduo MMQ	0,061 > 0,05	Constante

Além disso, os modelos também podem ser comparados utilizando outras métricas de erro. Já que os dados não são normalmente distribuídos, os métodos escolhidos para análise dos modelos não incluem medidas sensíveis a valores extremos como média e desvio padrão dos dados de vazão, de forma a evitar vieses. Considerando isso, os erros calculados foram o Erro Absoluto Médio (MAE), o Erro Quadrático Médio (MSE) e o Erro Percentual Absoluto Médio Simétrico (sMAPE), conforme as equações de 21 a 23.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{obs,i} - y_{prev,i}| \quad (21)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{obs,i} - y_{prev,i})^2 \quad (22)$$

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_{obs,i} - y_{prev,i}|}{y_{obs,i}} 100\% \quad (23)$$

Os resultados obtidos por meio desses cálculos são apresentados na tabela 5. Em todos os métodos, os menores valores são verificados no caso do SARIMAX, indicando que esse modelo teve resultados mais próximos dos valores observados.

Tabela 5 - Erros dos modelos

Erro	Mínimos Quadrados	SARIMAX
MAE	316,322	251,877
MSE	426,0342	349,907
sMAPE	13,952%	10,420%

4.2 ESTUDO DE CASO: SÃO FÉLIX DO ARAGUAIA

4.2.1 Aplicação do SARIMAX aos Dados de Vazão

Para o estudo de caso, repete-se a modelagem com o SARIMAX, dessa vez considerando a região de São Félix do Araguaia, usando a vazão calculada e a vazão no Trecho Médio como variáveis exógenas.

Os dados de vazão de São Félix do Araguaia e Trecho Médio são disponibilizados apenas em frequência mensal. Em São Félix do Araguaia, nota-se um dado faltante, que é substituído pelo valor da média local. No caso do Trecho Médio, não há lacunas nem valores nulos, de forma que o único tratamento necessário nesse conjunto de dados é a identificação de inconsistências por meio da regra do intervalo interquartilico, com a substituição dos valores inconsistentes pelos valores de médias locais. A figura 20 apresenta os polígonos de Thiessen gerados no QGIS a partir dos dados de precipitação na sub bacia hidrográfica. Os dados de precipitação são mostrados no gráfico figura 21.

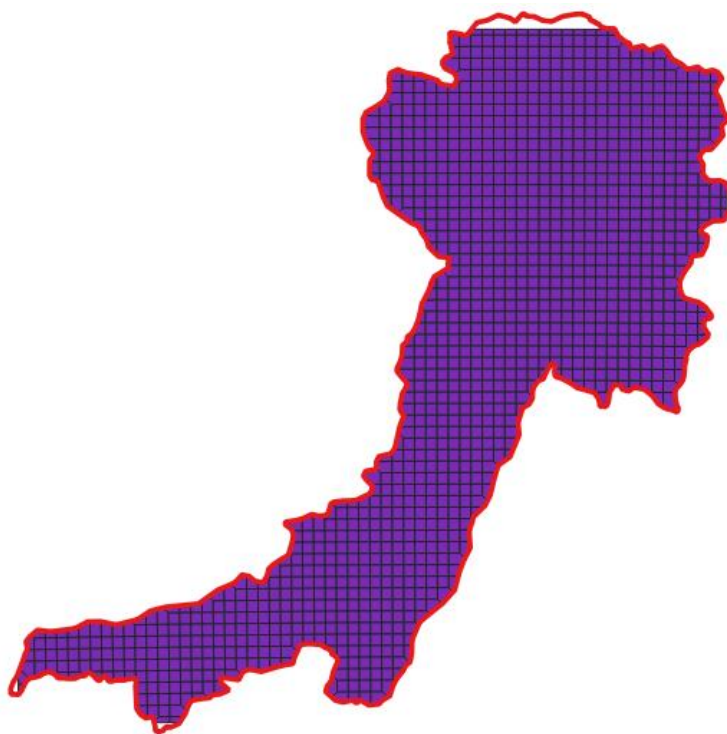


Figura 20 - Polígonos de Thiessen gerados a partir do dados de precipitação na região de São Félix do Araguaia

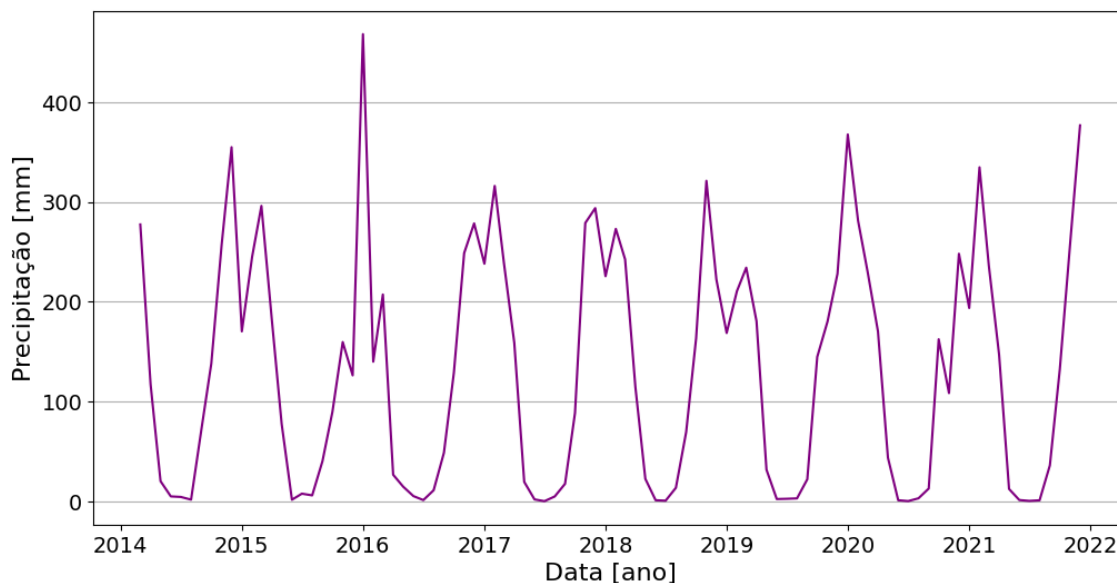


Figura 21 - Dados de precipitação na região de São Félix do Araguaia

Da mesma forma que no caso anterior, faz-se a análise da correlação entre os dados de vazão e da precipitação. Nesse caso, o maior coeficiente de correlação se dá na defasagem de dois meses, conforme a tabela 6.

Tabela 6 – Correlação de Spearman entre vazão e precipitação em São Félix do Araguaia

Defasagem	Coefficiente de correlação
T-4	0,410
T-3	0,747
T-2	0,883
T-1	0,824
T-0	0,467

Aplicando as variáveis à função `auto_arima`, tem-se que o modelo ideal para os dados utilizados é o SARIMAX(3,1,0)(2,1,0)₁₂, obtendo a equação 24:

$$(1 - 0,343B^1 - 0,462B^2 - 0,327B^3 - 0,699B^{12} - 0,297B^{24})y_t = \varepsilon_t \quad (24)$$

Removendo a primeira e a última medição, obtém-se as previsões apresentadas na figura 22. Nota-se um ajuste irregular na etapa de calibração do modelo até o ano de 2017. Já na etapa de validação, os valores previstos se mostram muito próximos dos observados.

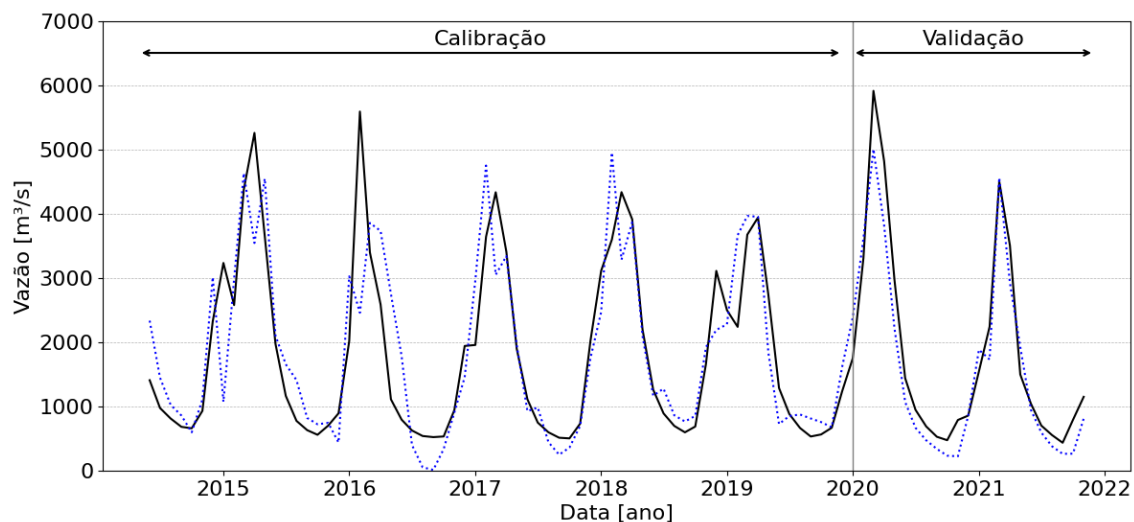


Figura 22 - Vazão prevista em São Félix do Araguaia com o SARIMAX, em que a linha pontilhada azul indica a vazão prevista, e a linha contínua é a vazão observada

4.2.1.1 Avaliação dos Resultados

A relação entre os valores observados e previstos é evidenciada no gráfico da figura 23. Observa-se que há uma relação linear entre ambos, mas alguns pontos aparecem mais distantes, conforme também se verifica nos gráficos de distribuição mostrados nos cantos superior e direito. Esse gráfico também mostra que tanto os dados observados quanto os previstos não são normalmente distribuídos. Isso é evidenciado na figura 24.

Tabela 7 - Coeficientes de correlação entre os valores observados e previstos de vazão em São Félix do Araguaia

Pearson	0,870
Spearman	0,911
Kendall	0,737

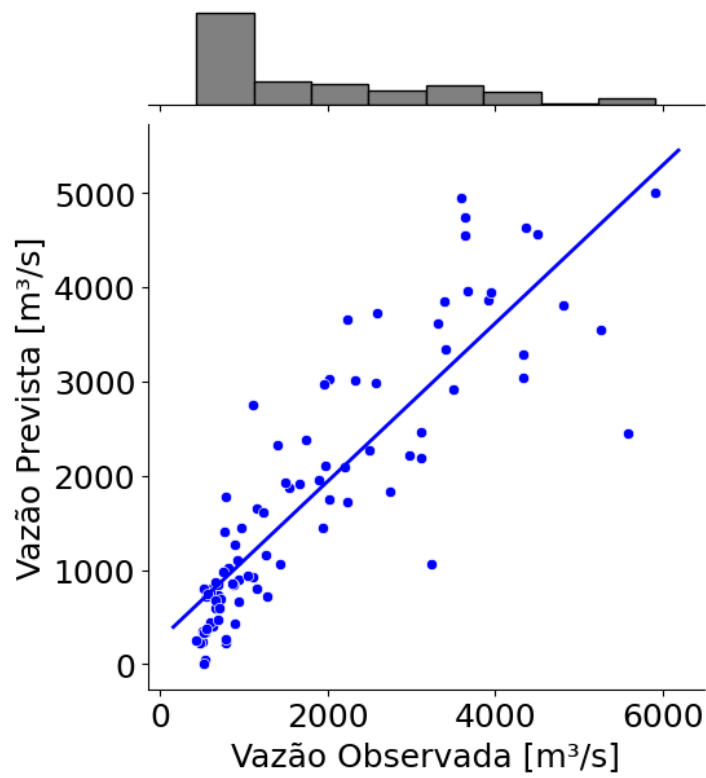


Figura 23 - Relação entre os dados previstos e observados de vazão em São Félix do Araguaia

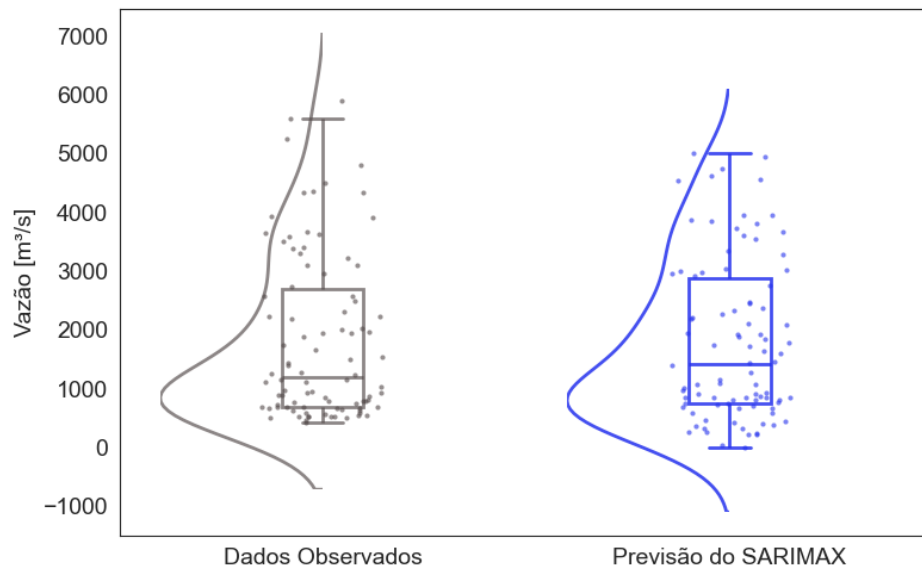


Figura 24 - Distribuição amostral dos dados observados e previstos em São Félix do Araguaia

4.2.1.2 Análise dos Resíduos

A observação dos resíduos entre os valores previstos e observados para a vazão em São Félix do Araguaia na figura 25 evidencia os maiores valores encontrados nos picos e vales de medição, da mesma forma que no caso da UHE de Estreito, além da divergência entre a previsão e as medições no início da calibração, até o ano de 2017.

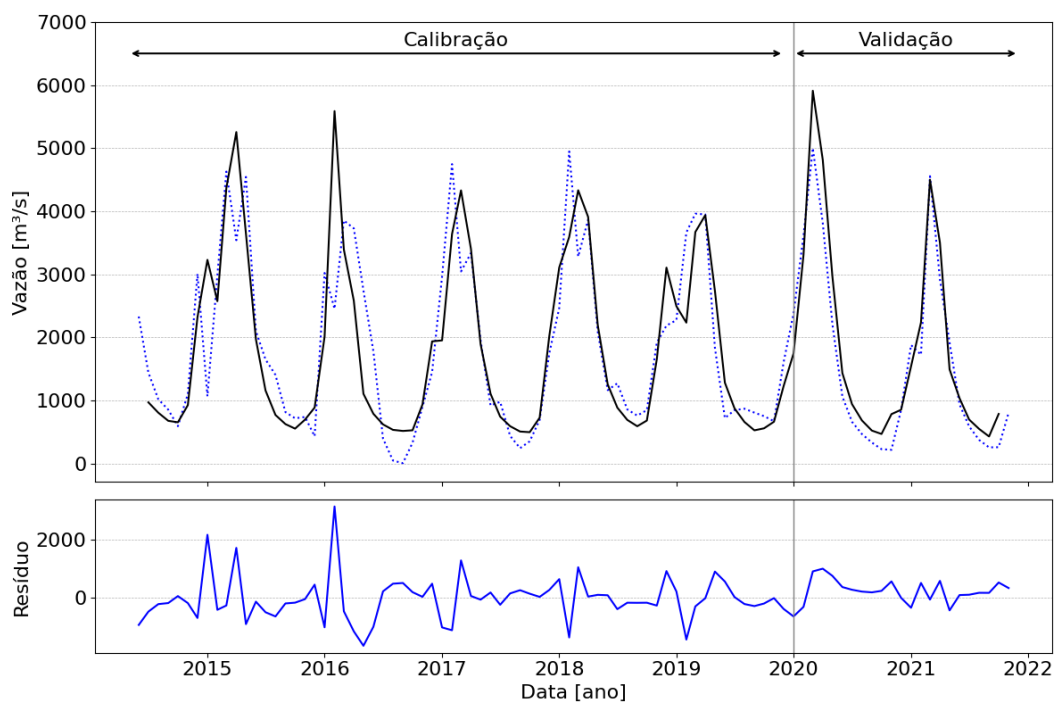


Figura 25 - Resíduos e previsões do modelo SARIMAX para vazão em São Félix do Araguaia, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos.

Na figura 26, nota-se que os resíduos se afastam levemente da distribuição normal, com uma cauda alongada, o que é confirmado pelo teste de Kolmogorv-Smirnov.

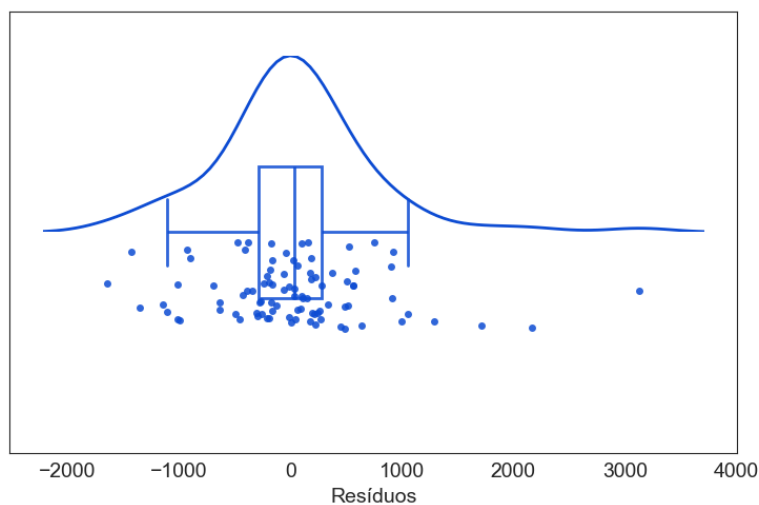


Figura 26 - Distribuição dos resíduos

Quanto à correlação dos resíduos, o gráfico da ACF, mostrado na figura 27 mostra que eles são independentes, já que os valores da autocorrelação estão inseridos dentro do intervalo de confiança representado pela faixa azul em todas as defasagens consideradas. Além disso, a média dos resíduos, bem como no caso da estimativa de vazão na UHE do Estreito, foi próxima de zero. Essa hipótese também é confirmada pelo teste de Ljung-Box.

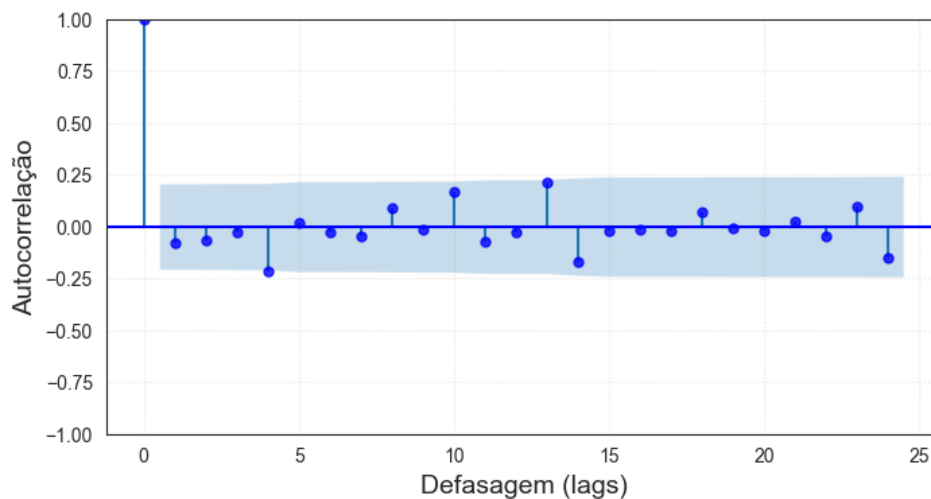


Figura 27 - Função de autocorrelação dos resíduos entre os valores observados e previstos para a vazão em São Félix do Araguaia

Por fim, verifica-se a variância por meio do teste de Goldfeld-Quandt, que indica homocedasticidade, tendo o p-valor de 0,999, o que também é ilustrado no gráfico da figura 27.

Tabela 8 - Teste de Goldfeld-Quandt para os resíduos entre os valores estimados e observados de vazão em São José do Araguaia

	p-valor	Variância
Resíduo da Vazão	0,999 > 0,05	Constante

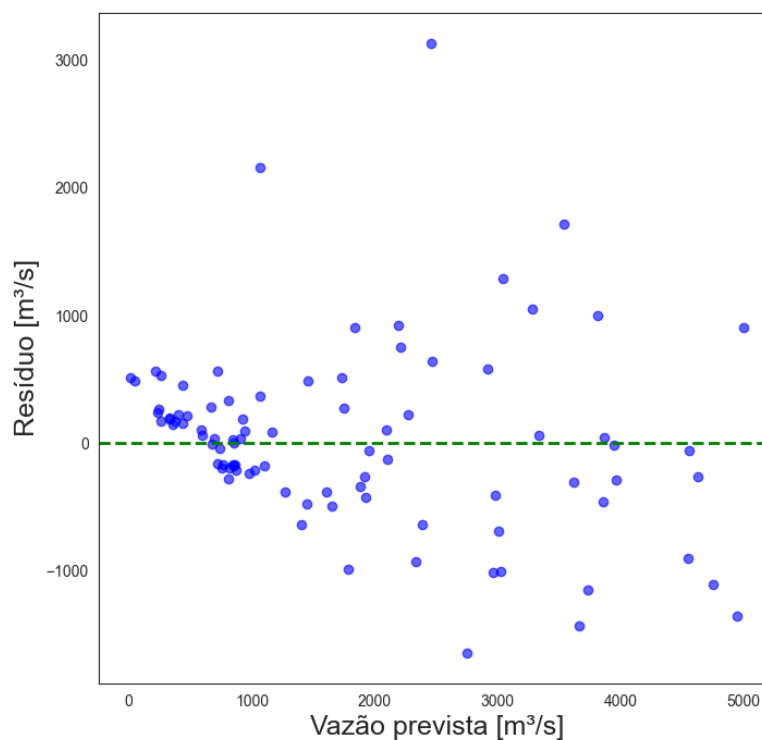


Figura 28 – Gráfico de dispersão dos resíduos entre a vazão observada e prevista em São Félix do Araguaia

Verificando os erros do modelo, apresentados na tabela 9, observa-se valores baixos considerando a escala das medições, indicando proximidade entre os valores observados e previstos.

Tabela 9 – Erros do modelo de vazão

MAE	489,857
MSE	710,162
sMAPE	33,646 %

4.2.2 Aplicação do SARIMAX aos Dados de Cota

Os dados de cota em São Félix do Araguaia são disponibilizados pela ANA com duas medições por dia, às 7h00 e às 17h00. Além dessas medições, os dados importados

possuem uma cota das 12h00, que consiste na média entre as duas medições. Esses dados das 12h00 são utilizados para o cálculo das cotas mensais que entram do modelo. Na série temporal, observa-se presença de uma lacuna no mesmo período em que houve nos dados de vazão na mesma região. Da mesma maneira, esse valor é preenchido pela média local.

Pelo resultado da função `auto_arima`, tem-se que o modelo ideal para os dados de cota analisados é o SARIMAX(2,1,1)(2,1,0)₁₂, obtendo a equação 25:

$$(1 - 0,421B^1 + 0,373B^2 - 0,736B^{12} - 0,259B^{24})y_t = (1 - 0,784B^1)\varepsilon_t \quad (25)$$

Os resultados do modelo de previsão dos dados de cota em São Félix do Araguaia são mostrados sem a primeira e a última medições na figura 29. Nesse caso, os valores na calibração foram mais próximos do que na previsão da vazão. A cota prevista é mostrada pela linha pontilhada em verde, de forma a diferir da indicação da previsão de vazão, e a cota de alerta de enchente (850 cm) é destacada pela linha horizontal vermelha.

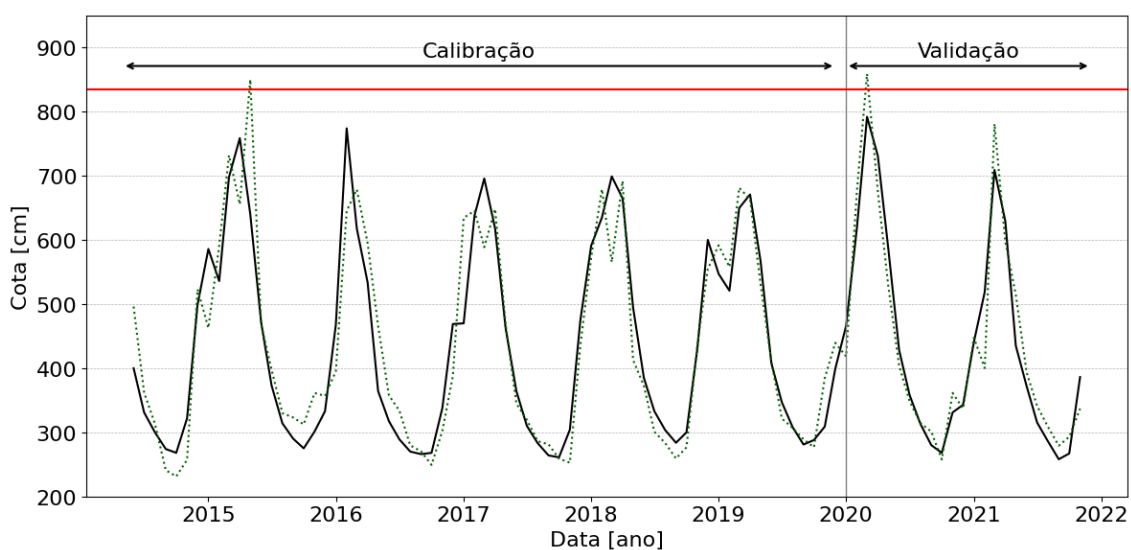


Figura 29 – Previsão da cota com o SARIMAX em que os dados previstos são dados pela linha pontilhada verde, os dados observados são representados pela linha contínua preta, e a linha horizontal vermelha destaca a cota de alerta de inundações

4.2.2. Avaliação dos Resultados

A relação entre os dados previstos e observados são mostrados na figura 30, confirmando que ambos apresentam comportamentos semelhantes, conforme também se

observa nos gráficos de distribuição dos dados mostrados na figura 31 e pelos coeficientes de correlação de Pearson, Spearman e Kendal, na tabela 10.

Tabela 10 - Coeficientes de correlação entre os valores observados e previstos de vazão em São Félix do Araguaia

Pearson	0,917
Spearman	0,926
Kendall	0,758

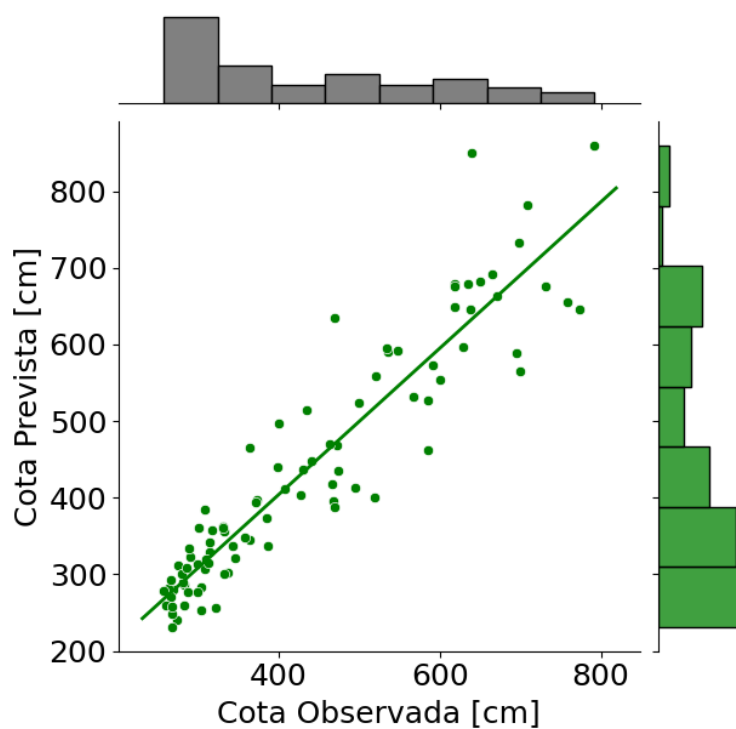


Figura 30 – Relação entre os dados observados e previstos de cota em São Félix do Araguaia

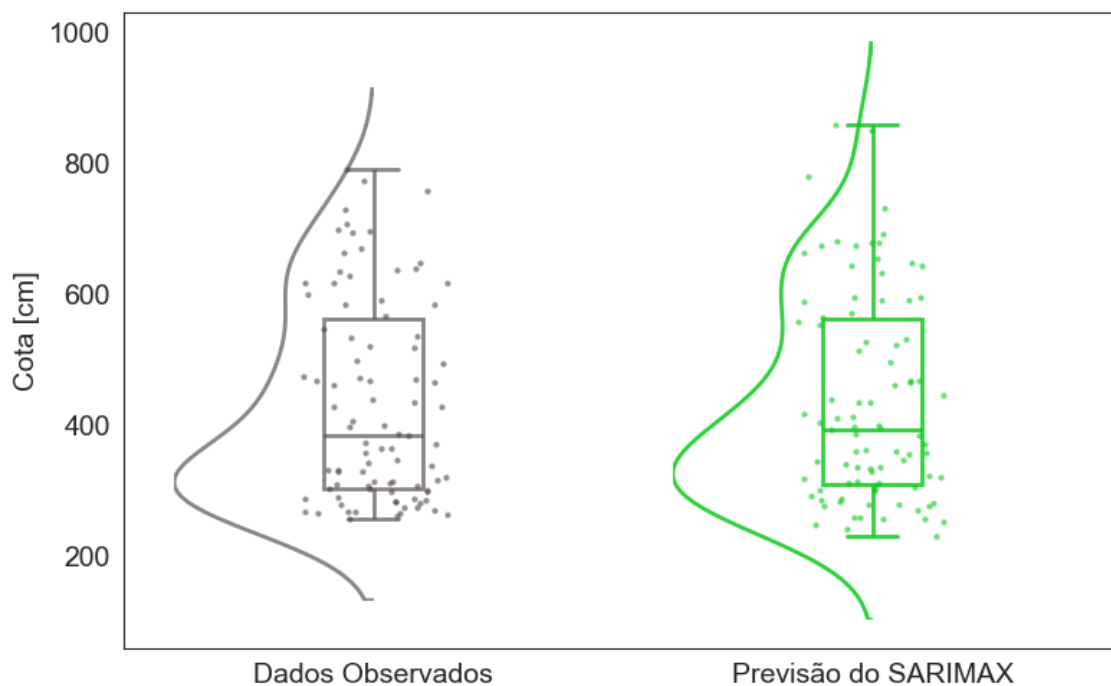


Figura 31 - Distribuição dos dados observado e previstos de cota em São Félix do Araguaia

4.2.2. Análise dos Resíduos

Os resíduos entre os valores observados e previstos de cota em São Félix do Araguaia são destacados na figura 32. Embora também se observe uma pequena divergência no início da etapa de calibração, o modelo passa a convergir antes do modelo de vazão, em 2016. Os maiores resíduos ainda se localizam nos picos, mas há um comportamento mais próximo nos vales, em relação à previsão de vazão na mesma região.

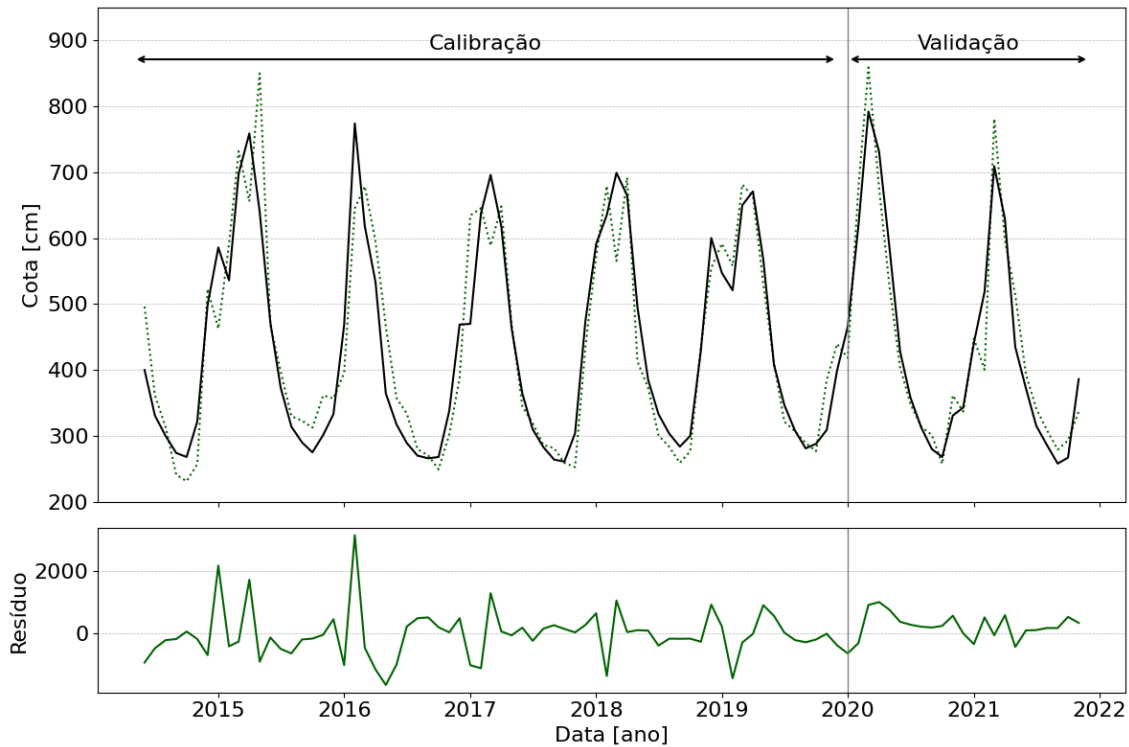


Figura 32 - Resíduos e previsões do modelo SARIMAX para cota em São Félix do Araguaia, em que a linha pontilhada azul são os valores previstos e a linha contínua azul são os resíduos.

Esses resíduos esboçam o mesmo comportamento evidenciado no caso da vazão na região. A figura 33 apresenta a distribuição desses resíduos, com a mesma característica da cauda alongada, o que afasta os resíduos do comportamento de ruído branco esperado.

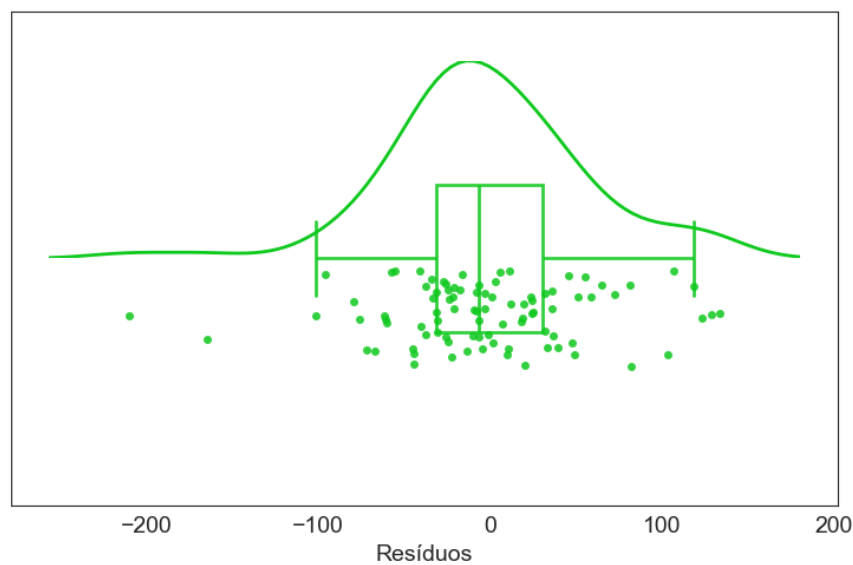


Figura 33 - Distribuição dos resíduos entre os valores previstos e observados de cota em São Félix do Araguaia

Já no que diz respeito à autocorrelação dos resíduos, conforme o gráfico da ACF na figura 34, tem-se que os resíduos são independentes, uma vez que todos os pontos são localizados dentro da faixa de segurança, também conforme o teste de Ljung-Box. Além disso, o cálculo da média dos resíduos foi bastante próximo de zero, conforme esperado.

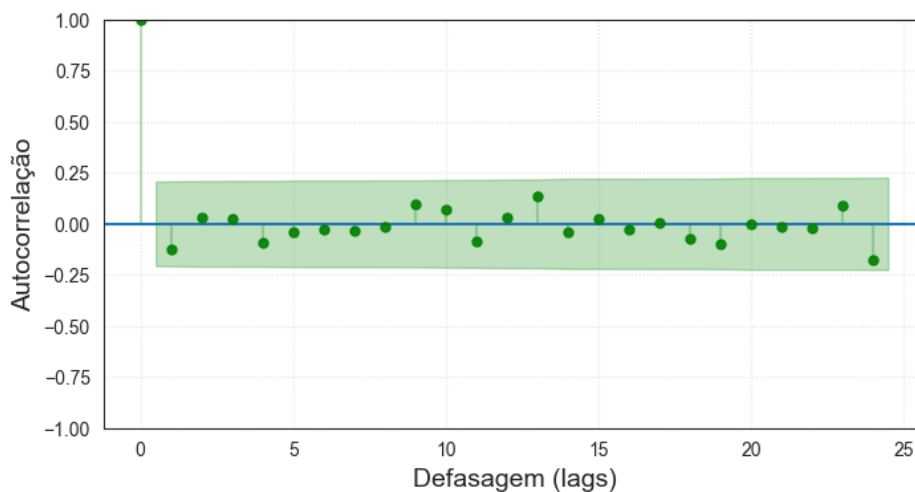


Figura 34 - ACF dos resíduos entre os valores observados e previstos de cota em São Félix do Araguaia

Por fim, a hipótese da variância constante é confirmada pelo teste de Goldfeld-Quandt, que teve o p-valor de 0,999, conforme a tabela 11. Essa informação é reiterada no gráfico de dispersão dos resíduos da figura 35.

Tabela 11 - Teste de Goldfeld-Quandt para os resíduos entre os valores observados e previstos de cota em São Félix do Araguaia

	p-valor	Variância
Resíduo da Cota	0,999 > 0,05	Constante

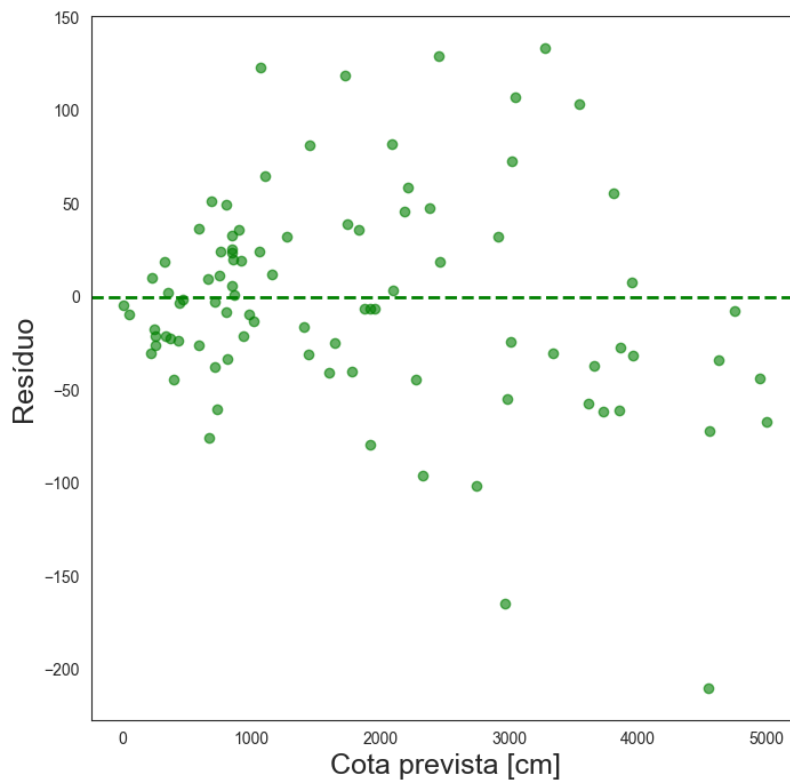


Figura 35 - Dispersão dos resíduos entre os valores previstos e observados de cota em São Félix do Araguaia

Os erros calculados são mostrados na tabela 12. Nota-se que os valores são pequenos em relação às medições. Deve-se recordar que os erros da previsão de vazão não podem ser comparados com os erros da previsão da cota, uma vez que as duas variáveis têm escalas diferentes.

Tabela 12 – Erros do modelo de cota

MAE	42,197
MSE	57,007
sMAPE	9,321%

5 CONCLUSÃO

O trabalho analisou e comparou a performance de dois métodos de previsão de séries temporais hidrológicas no contexto de dados de vazão da UHE do Estreito. Tanto o modelo gerado pelo o MMQ quanto o gerado pelo SARIMAX apresentam resultados próximos aos valores observados, se mostrando eficientes para previsão da vazão. Uma análise comparativa mostra que a previsão do SARIMAX é mais precisa, uma vez que apresenta menores resíduos entre os valores observados e previstos. Para se certificar da eficiência deste método, ele foi aplicado novamente ao estudo de caso em São Félix do Araguaia, obtendo, também, resultados precisos.

No caso dos dados de precipitação mensal obtidos pelo conjunto de satélites IMERG, os valores de precipitação calculados se mostraram diferentes dos disponibilizados pelo CENSIPAM. Isso se deve, possivelmente ao uso de versões distintas dos dados do IMERG, uma vez que o CENSIPAM utiliza a versão 6, já não disponível, e este trabalho emprega a versão 7. Contudo, apesar das diferenças nos valores calculados, ao aplicar ambas as séries aos métodos de previsão propostos, obteve-se resultados muito similares e consistentes, provando que tal diferença entre as versões dos dados não tem grande impacto no desempenho dos modelos.

No que diz respeito à estimativa de vazão, o MMQ apresenta resultados bem ajustados aos valores observados, com as maiores diferenças aparecendo nos picos de medição. A análise dos resíduos entre os valores previstos e os observados prova esse resultado por meio da sua baixa amplitude. Observando o comportamento dos resíduos, nota-se que eles não possuem distribuição normal e média zero, embora próxima de zero, há dependência entre os valores e homocedasticidade. Ou seja, não se aproximam totalmente do comportamento de ruído branco que é considerado ideal para os modelos autorregressivos, embora sejam bastante próximos.

No SARIMAX, por sua vez, também se observa que os maiores resíduos entre os valores previstos e os observados são nos picos de medição. Entretanto, as previsões são mais próximas dos valores observados, o que é evidenciado pelos menores valores dos erros MAE, MSE e sMAPE. O comportamento dos resíduos nesse caso também se aproxima mais dos ruídos brancos do que no modelo do MMQ, uma vez que apresentam independência entre si.

Dessa forma, ficam evidentes a eficiência dos dois modelos na previsão de séries de dados hidrológicas e a melhor performance do SARIMAX em relação à proximidade com os dados observados, por considerar a sazonalidade dos dados, mostrando, assim, a importância de analisar o comportamento dos dados para a escolha do modelo de previsão.

Considerando a maior eficiência do SARIMAX no caso da UHE do Estreito, o método foi aplicado a um estudo de caso na região de São Félix do Araguaia, para estimativa de dados de vazão e de cota, visando auxiliar na previsão de enchentes. Dada a similaridade entre os dados de vazão e de cota, as duas previsões tiveram resultados também bastante próximos entre si.

Na previsão tanto da vazão quanto da cota em São Félix do Araguaia, notou-se que o modelo não se ajustou tão bem no início da etapa de calibração, passando a convergir a partir do ano de 2017. Os valores previstos, entretanto, se mostraram bastante próximos na etapa de validação do modelo.

A análise dos resíduos entre os valores observados e previstos também revelou resultados próximos. Nos dois casos, os resíduos foram independentes, com variância constante, média próxima de zero e distribuição não gaussiana. Ou seja, da mesma forma que na previsão de vazão na UHE de Estreito, os resíduos se aproximam do comportamento de ruído branco, mas não correspondem totalmente a ele.

Analisando os resultados obtidos pelo SARIMAX nas duas regiões de estudo, notou-se que o método resulta em previsões confiáveis para as séries de dados hidrológicas estudadas, com previsões próximas dos valores medidos, se mostrando uma ferramenta eficaz para a previsão de alagamentos, conseqüentemente, contribuindo para a gestão de recursos hídricos e para a segurança da população da região.

5.1 Trabalhos Futuros

Como indicação para trabalhos futuros, recomenda-se:

- Avaliação do uso de modelos combinados para melhor ajuste dos picos de medição;

- Comparação entre os resultados do SARIMAX e de métodos de previsão com *machine learning*;

6 REFERÊNCIAS

AKAIKE, H. **A Bayesian analysis of the minimum AIC procedure**. Annals of the Institute of Statistical Mathematics, v. 30, n. 1, p. 9-14, 1978.

AKAIKE, H. **Information theory and an extension of the maximum likelihood principle**. In: PETROV, B.N.; CSAKI, F. (Eds.). Proceedings of the Second International Symposium on Information Theory. Budapeste: Akademiai Kiado. 1973, p. 267-281.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Plano Estratégico de Recursos Hídricos** - Rios Tocantins e Araguaia, 2009.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. Vulnerabilidade a inundações - Trechos de Curso D'água Inundáveis, 2024. Disponível em: https://dadosabertos.ana.gov.br/datasets/62a3924c1da34f73bf5b7132677213ea_0/explorate?location=-15.377091%2C-52.749507%2C4.65. Acesso em: 23 out. 2025.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Mapas das Bacias Hidrográficas DNAEE**. 2016. Disponível em: <https://metadados.snirh.gov.br/geonetwork/srv/por/catalog.search#/metadata/92c8c8e2-a82e-4753-9bda-f3a6818e961e>. Acesso em: 10 nov. 2025.

ANTUNES, José Leopoldo Ferreira. **"Grow and multiply": social development, birth rates and demographic transition in the Municipality of São Paulo, Brazil, time-series for 1901-94**. Revista Brasileira de Epidemiologia, v. 1, p. 61-78, 1998. Disponível em: <https://www.scielo.org/pdf/rbepid/1998.v1n1/61-78/en>

ANTUNES, José Leopoldo Ferreira; CARDOSO, Maria Regina Alves. **Uso da análise de séries temporais em estudos epidemiológicos**. Epidemiologia e Serviços de Saúde, v. 24, n. 3, p. 565-576, 2015. Disponível em: <https://doi.org/10.5123/S1679-49742015000300024>

ARAÚJO, Eduardo; JÚNIOR, Xavier; BARBOSA, Nyedja; DE OLIVEIRA, Tiago. **Modelagem e previsão de modelos de séries temporais do consumo de energia elétrica na Região Nordeste do Brasil**. 2023.

BARROS, Anna Carolina; MATTOS, Daiane M. de; OLIVEIRA, Ingrid C. L. de; FERREIRA, Pedro Guilherme C.; DUCA, Victor Eduardo L. A. **Análise de Séries**

Temporais em R: curso introdutório. 1. ed. Editora Atlas, 2021. ISBN 978-85-352-9087-5.

BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Series Analysis: Forecasting and Control.** 4. ed. [S. l.]: John Wiley & Sons, 2008. ISBN 978-0-470-27284-8.

CABRAL JÚNIOR, Jório Bezerra; LUCENA, Rebecca Luna. **Analysis of precipitation using Mann-Kendall and Kruskal-Wallis non-parametric tests.** Mercator (Fortaleza), v. 19, p. e19001, 2020. Disponível em: <https://doi.org/10.4215/rm2020.e19001>

CATOLICO, Ana Carolina Chaves; STRAUCH, Julia. **Bacia Hidrográfica Dos Rios Tocantins E Araguaia: Índice Sintético De Diferenças Socioespaciais.** Revista Contemporânea, v. 4, n. 9, p. e5675-e5675, 2024. Disponível em: <https://ojs.revistacontemporanea.com/ojs/index.php/home/article/view/5675/4199>

CLEMMENS, Albert J.; WAHLIN, Brian T. **Accuracy of annual volume from current-meter-based stage discharges.** Journal of hydrologic engineering, v. 11, n. 5, p. 489-501, 2006.

CORNELIO, Emilly; PROCÓPIO, Hugo; DOS SANTOS, Flávio; DA FONSECA, Tiago; BARCELOS JÚNIOR, Manuel; ANGARITA, Jorge. **Identificação da influência das marés sobre o nível dos rios.** In: GOMES, Marília Miranda Forte et al. (Orgs.). Anais do I Encontro de Pós-Graduação na UnB Gama. Brasília: Universidade de Brasília, FCTE - Faculdade de Ciências e Tecnologias em Engenharia, Universidade de Brasília - UnB, Campus Gama, 2025. p. 74-80. Disponível em: <https://www.avaeventosetecnologia.com/iposgama>.

CORNELIO, Emilly; PROCÓPIO, Hugo; DOS SANTOS, Flávio; DA FONSECA, Tiago; BARCELOS JÚNIOR, Manuel; ANGARITA, Jorge. **Influência da aplicação de janelamento a métodos estatísticos de identificação de inconsistências em séries hidrológicas.** In: GOMES, Marília Miranda Forte et al. (Orgs.). Anais do I Encontro de Pós-Graduação na UnB Gama. Brasília: Universidade de Brasília, FCTE - Faculdade de Ciências e Tecnologias em Engenharia, Universidade de Brasília - UnB, Campus Gama, 2025. p. 81-90. Disponível em: <https://www.avaeventosetecnologia.com/iposgama>.

DAVEY, A. M.; FLORES, B. E. **Identification of seasonality in time series: A note. Mathematical and computer modelling**, v. 18, n. 6, p. 73-81, 1993. Disponível em: [https://doi.org/10.1016/0895-7177\(93\)90126-J](https://doi.org/10.1016/0895-7177(93)90126-J)

DE MEDEIROS, Raimundo Mainar et al. **Variáveis Climáticas Sobre A Bacia Hidrográfica Do Rio Uruçuí Preto-Piauí, Brasil**. Subsídios Para Gestão De Recursos Hídricos E Agropecuários. RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218, v. 3, n. 1, p. e311017-e311017, 2022. Disponível em: <https://doi.org/10.47820/recima21.v3i1.1017>.

DEFESA Civil reforça monitoramento do Rio Araguaia e intensifica medidas de prevenção em São Félix do Araguaia. **Agência da Notícia**, 5 fev. 2025. Disponível em: <https://www.agenciadanoticia.com.br/mato-grosso/noticia/137787/defesa-civil-reforca-monitoramento-do-rio-araguaia-e-intensifica-medidas-de-prevencao-em-sao-felix-do-araguaia>. Acesso em: 8 dez. 2025.

FEIGELSON, Eric D.; BABU, G. Jogesh; CACERES, Gabriel A. **Autoregressive times series methods for time domain astronomy. Frontiers in Physics**, v. 6, p. 80, 2018. Disponível em: <https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2018.00080/full>

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria básica**. ed. Porto Alegre: AMGH, 2011.

GLEESON, K.; HUSBAND, S.; GAFFNEY, J.; BOXALL, J. **A data quality assessment framework for drinking water distribution system water quality time series datasets**. AQUA—Water Infrastructure, Ecosystems and Society, v. 72, n. 3, p. 329-347, 2023. Disponível em: <https://doi.org/10.2166/aqua.2023.228>

HELSEL, D. R.; HIRSCH, R. M. **Statistical methods in water resources**: US Geological Survey Techniques of Water Resources Investigations, book 4, chap. 2002.

HORVATIC, Davor; STANLEY, H. Eugene; PODOBNIK, Boris. **Detrended cross-correlation analysis for non-stationary time series with periodic trends**. Europhysics Letters, v. 94, n. 1, p. 18007, 2011.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Previsão: Princípios e Prática**. 3. ed. Otexts, 2021. ISBN 0987507133. Disponível em: <https://otexts.com/fpppg/>

JAMSHIDI, E.; YUSUP, Y.; KAYODE, J.; KAMARUDDIN, M. **Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature.** *Ecological Informatics*, v. 69, p. 101672, 2022. Disponível em: <https://doi.org/10.1016/j.ecoinf.2022.101672>.

KENNEDY, Peter (2002). **Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics**, *Journal of Economic Surveys*, Wiley Blackwell, vol. 16: 569-589. (2009), “A Guide to Econometrics”. Boston: MIT Press.

KIM, Y.; JEONG, J; PARK, H; KWON, M; CHO, C; JEONG, J. **Development of a data-driven ensemble regressor and its applicability for identifying contextual and collective outliers in groundwater level time-series data.** *Journal of Hydrology*, v. 612, p. 128127, 2022. Disponível em: <https://doi.org/10.1016/j.jhydrol.2022.128127>

KVAM, Paul; VIDAKOVIC, Brani; KIM, Seong-joon. **Nonparametric statistics with applications to science and engineering with R.** John Wiley & Sons, 2023.

LIRA, Talita de Melo; CHAVES, Maria do Perpétuo Socorro Rodrigues. **Comunidades ribeirinhas na Amazônia: organização sociocultural e política.** *Interações (Campo Grande)*, v. 17, p. 66-76, 2016. Disponível em: <https://doi.org/10.20435/1518-70122016107>.

MACHIWAL, Deepesh; JHA, Madan Kumar. **Hydrologic time series analysis: theory and practice.** Springer Science & Business Media, 2012.

MONTEIRO, Raquel Macêdo Lopes et al. **Modelagem Estatística Hidrológica na Bacia Hidrográfica do Rio Tocantins.** *Revista Brasileira de Geografia Física*, v. 17, n. 2, p. 1158-1173, 2024.

MORETTIN, Pedro A.; TOLOI, Clélia. **Análise de Séries Temporais.** 2006.

MÜLLER, Thomas; SCHÜTZE, Manfred; BÁRDOSSY, András. **Temporal asymmetry in precipitation time series and its influence on flow simulations in combined sewer systems.** *Advances in Water Resources*, v. 107, p. 56-64, 2017.

NIELSEN, Aileen. **Practical time series analysis: Prediction with statistics and machine learning.** O'Reilly Media, 2019.

PEREIRA, M.; BARBIEIRO, B.; QUEVEDO, D. (2020). **Importance of river basin monitoring and hydrological data availability for the integrated management of**

water resources. Sociedade & Natureza, v. 32, p. 292-303. Disponível em: <https://doi.org/10.14393/SN-v32-2020-43458>

PUTH, Marie-Therese; NEUHÄUSER, Markus; RUXTON, Graeme D. **Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits.** Animal Behaviour, v. 102, p. 77-84, 2015.

RAO, A. Ramachandra; HAMED, Khaled H.; CHEN, Huey-Long. **Nonstationarities in hydrologic and environmental time series.** Springer Science & Business Media, 2003.

SILVA, Benedito Cláudio da; TUCCI, Carlos Eduardo Morelli; COLLISCHONN, Walter. **Previsão de vazão com modelos hidroclimáticos.** Rbrh: revista brasileira de recursos hídricos. Porto Alegre, RS: ABRH. Vol. 11, n. 3, p. 15-29, 2006.

SÃO Félix do Araguaia está em estado de emergência devido a cheia no rio Araguaia. **Diário Araguaia**, 26 jan. 2022. Disponível em: <https://www.diarioaraguaia.com.br/noticia/25/410/Sao-Felix-do-Araguaia-esta-em-estado-de-emergencia-devido-a-cheia-no-rio-Araguaia/>. Acesso em: 8 dez. 2025.

SILVA, Letícia Anchieta. **Modelagem da geração de energia hidrelétrica no Brasil com séries temporais.** 2025.

SILVEIRA, Anderson Garcia; DE MATTOS, Viviane Leite Dias; KONRATH, Andrea Cristina. **Avaliação Da Estacionariedade E Teste De Cointegração Em Séries Temporais O Caso Da Demanda De Energia Elétrica Residencial No Brasil.** RETEC-Revista de Tecnologias, v. 9, n. 3, 2016. Disponível em: <https://www.fatecourinhos.edu.br/retec/index.php/retec/article/view/267>

THIESSEN, Alfred H. **Precipitation averages for large areas.** Monthly weather review, v. 39, n. 7, p. 1082-1089, 1911.

TRÓNDHEIM, Helma; NIELSEN, Terji. **A wind power plant's impact on the grid frequency analysis of measurements in an electrically isolated island. with high penetration of inverter-based wind generation.** 2022.

TUKEY, John W. **Exploratory data analysis.** Reading/Addison-Wesley, 1977.

VIGEN, Tyler. **Spurious correlations.** 2025. Disponível em: <https://tylervigen.com/spurious-correlations>. Acesso em: 10 jun. 2025.

VOGEL, Richard; LALL, Upmanu; CAI, Ximing; RAJAGOPALAN, Balaji; WEISKEL, Peter; HOOPER, Richard; MATALAS, Nicholas. **Hydrology**: The interdisciplinary science of water. *Water Resources Research*, v. 51, n. 6, p. 4409-4430, 2015. Disponível em: <https://doi.org/10.1002/2015WR017049>