



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**MODELAGEM DA GERAÇÃO DE
ENERGIA HIDRELÉTRICA NO BRASIL
COM SÉRIES TEMPORAIS**

Letícia Anchieta Silva

Uberlândia-MG

2025

Letícia Anchieta Silva

**MODELAGEM DA GERAÇÃO DE
ENERGIA HIDRELÉTRICA NO BRASIL
COM SÉRIES TEMPORAIS**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Marcelo Tavares

**Uberlândia-MG
2025**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Prof. Dr. Marcelo Tavares

Prof^a Dra. Nádia Giaretta Biase

Prof^a Dra. Tatiane Bento da Costa

**Uberlândia-MG
2025**

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram de maneira significativa para a conclusão deste trabalho, tornando possível a realização deste sonho acadêmico.

Agradeço primeiramente a Deus, por minha vida e por iluminar o meu caminho.

Agradeço aos meus pais, Célia e Valterson, por todo apoio, amor e carinho durante esse processo. Minha admiração por vocês dois é imensa e minha gratidão maior ainda. Sem vocês não seria possível fazer esta graduação.

Agradeço ao meu esposo, Gabriel, por sua constante compreensão, paciência e incentivo que foram fundamentais ao longo deste processo.

Agradeço ao meu professor e orientador Marcelo Tavares, por toda sua paciência, atenção, auxílio e incentivo na conclusão deste trabalho de conclusão de curso.

Agradeço à minha querida professora Nádia, que também esteve presente neste trabalho, contribuindo com atenção, conhecimento e incentivo constante.

Por fim, agradeço aos amigos e colegas que fizeram parte da minha trajetória acadêmica, em especial à minha parceira de todas as horas nesta graduação, Bárbara, que esteve ao meu lado desde o início até a conclusão do curso, mantendo sempre sua amizade fiel e companheirismo incondicional.

Este trabalho é fruto do apoio, amor e dedicação de cada pessoa mencionada. A todos, o meu mais profundo agradecimento.

RESUMO

O setor de energia elétrica no Brasil representa um dos pilares fundamentais da economia e do cotidiano nacional, sendo a geração hidrelétrica sua principal fonte. Este trabalho teve como objetivo analisar, modelar e prever a geração mensal de energia hidrelétrica no Brasil, utilizando técnicas de séries temporais, com foco em modelos da classe SARIMA, a fim de contribuir com o planejamento energético nacional. Foram analisados dados mensais no período de janeiro de 1996 a dezembro de 2024, obtidos do Operador Nacional do Sistema Elétrico (ONS), via plataforma Ipeadata. Após a verificação de tendência e sazonalidade por meio dos testes de Cox-Stuart e Fisher, aplicou-se a diferenciação regular de primeira ordem, tornando a série estacionária. A partir da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), foram propostos modelos candidatos e, por meio da avaliação de critérios de informação (AIC e BIC) e métricas preditivas (MAPE e RMSE), o modelo SARIMA(2,1,1)(0,1,1)₁₂ foi selecionado. Os resíduos do modelo apresentaram comportamento de ruído branco, e as previsões para o ano de 2024 mostraram boa aderência aos valores observados, confirmando a capacidade do modelo em capturar a estrutura temporal da série. Os resultados evidenciam o potencial da modelagem estatística para apoiar o planejamento energético e sugerem que abordagens futuras considerem também técnicas de aprendizado de máquina para ganhos adicionais de precisão.

Palavras-chave: Previsão, Modelos Box & Jenkins, Tendência, Sazonalidade, SARIMA.

ABSTRACT

The electric power sector in Brazil represents one of the fundamental pillars of the national economy and daily life, with hydroelectric generation being its main source. This study aimed to analyze, model, and forecast the monthly hydroelectric energy generation in Brazil using time series techniques, with a focus on SARIMA-class models, in order to contribute to national energy planning. Monthly data from January 1996 to December 2024 were analyzed, obtained from the Brazilian National System Operator (ONS) via the Ipeadata platform. After verifying the presence of trend and seasonality using the Cox-Stuart and Fisher tests, first-order regular differencing was applied, rendering the series stationary. Based on the analysis of autocorrelation (ACF) and partial autocorrelation (PACF) functions, candidate models were proposed. Using information criteria (AIC and BIC) and predictive metrics (MAPE and RMSE), the SARIMA(2,1,1)(0,1,1)₁₂ model was selected. The model residuals behaved as white noise, and the forecasts for the year 2024 showed good adherence to the observed values, confirming the model's ability to capture the temporal structure of the series. The results highlight the potential of statistical modeling to support energy planning and suggest that future approaches also consider machine learning techniques for additional accuracy gains.

Keywords: Forecast, Box & Jenkins Models, Trend, Seasonality, SARIMA.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	II
Lista de Abreviações e Símbolos	III
1 Introdução	1
2 Fundamentação Teórica	4
2.1 Energia Hidrelétrica	4
2.1.1 Eventos Hidrológicos Críticos e Impactos no Setor Elétrico	4
2.2 Séries Temporais	6
2.3 Processos Estocásticos	6
2.4 Componentes de uma Série Temporal	7
2.4.1 Tendência	8
2.4.2 Sazonalidade	9
2.5 Estacionariedade	10
2.6 Diferenciação	11
2.7 Modelos ARIMA e SARIMA	12
2.8 Funções de autocorrelação	13
2.9 Metodologia Box-Jenkins	14
2.9.1 Identificação	14
2.9.2 Estimação	14
2.9.3 Diagnóstico	14
2.9.4 Previsão	14
2.10 Critérios de Seleção de Modelos	14
2.11 Métodos de Comparação de Previsão	15
2.12 Teste de Adequação do Modelo	16
3 Metodologia	18
3.1 Fonte dos Dados	18
3.2 Ferramentas Computacionais	18
3.3 Procedimentos Metodológicos	18
4 Resultados e discussões	21
5 Conclusões	31
Referências Bibliográficas	32
Apêndice A Apêndice	35

LISTA DE FIGURAS

4.1	Série temporal da geração de energia hidrelétrica no Brasil de jan/1996 a dez/2023.	21
4.2	Relação das médias com as amplitudes de cada par de observações.	22
4.3	Função de Autocorrelação da série temporal (ACF).	22
4.4	Série diferenciada em grau 1.	23
4.5	Autocorrelação (ACF) e autocorrelação parcial (PACF) da série diferenciada. . .	24
4.6	Autocorrelação (ACF) e autocorrelação parcial (PACF) residual.	26
4.7	Periodograma acumulado residual.	26
4.8	Valores observados e previsto do modelo SARIMA(2,1,1)(0,1,1) ₁₂	27

LISTA DE TABELAS

4.1	Modelos ajustados com parâmetros significativos e valores de critérios de seleção.	25
4.2	Parâmetros estimados para o modelo SARIMA(2,1,1)(0,1,1) ₁₂	25
4.3	Comparação dos valores realizados e previstos em (GWh), com seus erros absolutos.	28

LISTA DE ABREVIACÕES E SÍMBOLOS

LISTA DE ABREVIACÕES

ACF	Função de Autocorrelação (do inglês, Autocorrelation Function)
AIC	Critério de Informação de Akaike (do inglês, Akaike Information Criterion)
AR	Componente Autorregressivo (do inglês, AutoRegressive)
ARIMA	Modelo ARIMA (do inglês, AutoRegressive Integrated Moving Average)
BEN	Balanco Energético Nacional
BIC	Critério de Informação Bayesiano (do inglês, Bayesian Information Criterion)
GWh	Gigawatt-hora
I	Componente de Integração (do inglês, Integrated)
IPEA	Instituto de Pesquisa Econômica Aplicada
LSTM	Memória de Longo e Curto Prazo (do inglês, Long Short-Term Memory)
MA	Componente de Média Móvel (do inglês, Moving Average)
MAPE	Erro Percentual Absoluto Médio (do inglês, Mean Absolute Percentage Error)
MME	Ministério de Minas e Energia
ONS	Operador Nacional do Sistema Elétrico
PACF	Função de Autocorrelação Parcial (do inglês, Partial Autocorrelation Function)
RMSE	Raiz do Erro Quadrático Médio (do inglês, Root Mean Squared Error)
SAR	Componente Autorregressivo Sazonal (do inglês, Seasonal AutoRegressive)
SARIMA	Modelo SARIMA (do inglês, Seasonal AutoRegressive Integrated Moving Average)
SI	Componente de Integração Sazonal (do inglês, Seasonal Integration)
SMA	Componente de Média Móvel Sazonal (do inglês, Seasonal Moving Average)
SVR	Regressão com Vetores de Suporte (do inglês, Support Vector Regression)

1. INTRODUÇÃO

A eletricidade desempenha um papel central no funcionamento das sociedades modernas. É essencial não apenas para prover iluminação, aquecimento e refrigeração, mas também para impulsionar os sistemas de comunicação, transporte e processos industriais. O acesso confiável à eletricidade está intimamente ligado ao desenvolvimento econômico e à melhoria dos padrões de vida [21].

No entanto, apesar da importância crucial da energia elétrica, o Brasil enfrenta desafios significativos no que diz respeito ao seu abastecimento. A crescente demanda por energia, impulsionada pelo crescimento populacional, pelo desenvolvimento industrial e pelo avanço tecnológico, tem colocado pressão sobre o sistema elétrico nacional. Nos últimos anos, o Brasil vivenciou episódios de escassez de energia, como apagões e racionamentos em diversas regiões, o que evidencia a vulnerabilidade do sistema elétrico nacional e a necessidade de medidas para assegurar a segurança no fornecimento de energia [30].

Esses episódios evidenciam a necessidade de constante expansão e modernização da infraestrutura elétrica, a fim de evitar sobrecargas no sistema, riscos de desabastecimento e oscilações no fornecimento de energia. Embora o Brasil possua uma matriz elétrica predominantemente renovável, sua alta dependência da energia hidrelétrica o torna vulnerável a variações climáticas, exigindo estratégias de diversificação e previsão para garantir a segurança energética.

O Brasil dispõe de uma das matrizes elétricas mais diversificadas do mundo, composta predominantemente por fontes renováveis, com destaque para a fonte hídrica. Conforme divulgado no relatório final do Balanço Energético Nacional (BEN): Ano base 2022, publicado em 2023 [10], as fontes renováveis corresponderam a 88% da oferta interna de eletricidade, isto é, da soma entre a produção nacional e as importações, que são majoritariamente de origem renovável.

Uma parcela relevante dessa importação provém da Usina Hidrelétrica de Itaipu, empreendimento binacional administrado por Brasil e Paraguai. Embora metade da energia gerada pertença ao Paraguai, grande parte dessa cota é revendida ao Brasil, o que, sob a ótica estatística, é registrado como importação de energia, mesmo tratando-se de uma usina de copropriedade brasileira. Dentro desse total, a fonte hídrica, considerando tanto a energia gerada internamente quanto a proveniente de Itaipu, foi responsável por 64% da oferta interna de eletricidade [10]. O restante da matriz foi composto por outras fontes renováveis, como eólica, solar e biomassa, além de fontes não renováveis.

A energia hidrelétrica é gerada a partir do aproveitamento do potencial hidráulico de rios, por meio da conversão da energia cinética da água em energia elétrica, utilizando turbinas e

geradores. No Brasil, essa fonte é particularmente relevante devido à ampla disponibilidade de recursos hídricos e à predominância de grandes bacias hidrográficas, como a do Rio Paraná, do Rio São Francisco e do Rio Amazonas.

A geração de energia hidrelétrica oferece vantagens como baixo custo operacional e menores emissões de gases de efeito estufa em comparação às fontes fósseis. No entanto, sua produção está diretamente condicionada ao regime hidrológico, tornando-a vulnerável a oscilações climáticas, períodos de seca e alterações no uso da água. Diante de sua centralidade no sistema elétrico nacional, compreender o comportamento da geração hidrelétrica ao longo do tempo é essencial para subsidiar o planejamento energético do país e prevenir crises de abastecimento.

O uso de previsões do consumo de energia, ao possibilitar a análise do comportamento histórico e a projeção de tendências futuras, pode trazer benefícios significativos para o setor elétrico. Essas projeções fornecem subsídios para a tomada de decisões que promovem maior economia e confiabilidade no fornecimento de energia elétrica, contribuindo para o planejamento adequado da operação e da expansão do sistema. Dessa forma, busca-se evitar problemas como falhas operacionais, degradação da infraestrutura e até mesmo a paralisação de atividades produtivas e serviços essenciais. O planejamento de investimentos em redes e na capacidade produtiva depende diretamente das estimativas de demanda, além de considerar fatores como disponibilidade de recursos financeiros e prioridades estratégicas [7].

Supor um consumo acima do real faz com que os recursos financeiros sejam utilizados em investimentos supérfluos, o que representa desperdício de capital. Por outro lado, subestimar esse valor pode implicar em custos operacionais significativamente mais altos, uma vez que o sistema pode ser obrigado a acionar fontes de energia de emergência, como termelétricas, que possuem custos mais elevados de geração. Além disso, essa subestimação aumenta o risco de falhas no atendimento à demanda, o que pode acarretar penalidades regulatórias e prejuízos à confiabilidade do fornecimento. Assim, além de contribuir para uma operação mais estável e segura, as previsões representam uma vantagem estratégica tanto do ponto de vista operacional quanto financeiro [22].

Sendo assim, é fundamental utilizar procedimentos e métodos estatísticos como ferramentas de análise para compreender e transmitir os conhecimentos adquiridos sobre o consumo e a geração de energia elétrica no Brasil. Isso envolve descrever e identificar padrões de comportamento ao longo dos meses e anos, além de analisar as relações presentes nas observações. Na análise de dados históricos, é recomendável empregar ferramentas estatísticas adequadas que explorem as informações contidas nos dados, permitindo a elaboração de relatórios alinhados aos objetivos estabelecidos. Isso proporciona conclusões sobre o passado e possíveis previsões dos valores futuros com base nas variáveis analisadas.

Visto que os dados utilizados neste estudo possuem estrutura temporal, é possível aplicar diferentes abordagens estatísticas para explorá-los. Além da modelagem de séries temporais, informações temporais podem ser usadas para gerar indicadores, identificar padrões sazonais, calcular taxas de variação e avaliar tendências de longo prazo. Contudo, quando há indícios de dependência entre as observações ao longo do tempo, torna-se apropriado empregar técnicas

específicas de modelagem de séries temporais.

Segundo Morettin [24], séries temporais consistem em observações ordenadas no tempo, em que a correlação entre os valores pode ser significativa e, portanto, deve ser modelada e analisada. Compreender o comportamento dessas séries por meio de modelos probabilísticos que permitam simulações e previsões com certo grau de confiabilidade é crucial para profissionais e instituições que lidam com planejamento energético.

Neste contexto, este estudo tem como objetivo geral analisar, modelar e prever a geração mensal de energia hidrelétrica no Brasil, utilizando técnicas de séries temporais, com foco em modelos da classe SARIMA, a fim de contribuir com o planejamento energético nacional.

Os objetivos específicos são:

- Coletar e organizar a série histórica da geração de energia hidrelétrica no Brasil entre 1996 e 2024;
- Analisar a presença de tendência e sazonalidade na série por meio de testes estatísticos;
- Ajustar modelos SARIMA adequados à estrutura da série temporal;
- Avaliar a performance preditiva dos modelos com base em métricas como RMSE e MAPE;
- Gerar previsões para o ano de 2024 e interpretar seus resultados.

A escolha pela análise da geração de energia hidrelétrica justifica-se pelo fato de essa fonte representar a principal base da matriz elétrica brasileira. No entanto, sua produção é sensível a fatores naturais, como variações sazonais e condições climáticas. Eventos como o apagão de 2001–2002, a crise hídrica de 2021 e a instabilidade registrada em 2023 evidenciam a vulnerabilidade do sistema elétrico nacional diante de oscilações hidrológicas e operacionais. Nesse cenário, o uso de modelos estatísticos de previsão torna-se uma ferramenta essencial para antecipar variações na geração e apoiar um planejamento mais seguro, eficiente e sustentável do setor energético.

Cabe destacar que este estudo se limita à análise da geração hidrelétrica por ser a principal fonte de eletricidade do país, além de apresentar maior disponibilidade de dados históricos de qualidade e consistência. Embora outras fontes, como eólica, solar e termelétrica, também compoñham a matriz, a escolha por focar exclusivamente na geração hidrelétrica visa aprofundar a compreensão sobre a fonte que mais impacta o sistema. Os resultados obtidos podem ser úteis para subsidiar decisões estratégicas de órgãos governamentais, operadores do sistema elétrico e demais instituições envolvidas no planejamento energético nacional.

2. FUNDAMENTAÇÃO TEÓRICA

A previsão da geração de energia hidrelétrica no Brasil requer a compreensão de conceitos fundamentais relacionados tanto ao setor energético quanto às metodologias estatísticas aplicadas. Esta seção apresenta uma breve revisão teórica sobre energia hidrelétrica, crise do apagão de 2001-2002, a importância da previsão no setor elétrico, séries temporais, metodologia Box-Jenkins e os modelos de previsão SARIMA.

2.1 ENERGIA HIDRELÉTRICA

A energia hidrelétrica é a principal fonte de geração de energia elétrica no Brasil, sendo responsável, historicamente, por mais de 60% da matriz elétrica nacional. O predomínio dessa fonte renovável deve-se à abundância de recursos hídricos disponíveis no país e ao seu custo competitivo de produção quando comparado a outras fontes de energia [10].

A geração hidrelétrica depende diretamente do regime de chuvas, que alimenta os reservatórios das usinas. Como consequência, a produção de energia apresenta variações sazonais ao longo do ano, refletindo a dinâmica dos períodos de seca e de chuvas. Além da sazonalidade anual, ocorrem oscilações interanuais associadas a fenômenos climáticos como El Niño e La Niña, que podem alterar de forma significativa o volume de água disponível para a geração [28].

Essas variações naturais introduzem desafios importantes para o planejamento energético, uma vez que a previsibilidade da geração hidrelétrica não depende apenas de fatores técnicos, mas também de condições meteorológicas. A análise e a previsão da geração hidrelétrica exigem, portanto, métodos capazes de lidar com a variabilidade dos dados ao longo do tempo.

Nesse contexto, o comportamento da série histórica da geração de energia deve ser estudado com atenção para identificar tendências, padrões sazonais e eventuais rupturas causadas por eventos excepcionais, como secas prolongadas ou crises no setor energético.

2.1.1 EVENTOS HIDROLÓGICOS CRÍTICOS E IMPACTOS NO SETOR ELÉTRICO

A geração de energia hidrelétrica no Brasil depende fortemente de fatores naturais, especialmente do regime de chuvas. Essa dependência torna o sistema vulnerável a eventos extremos, como estiagens prolongadas, variações climáticas intensas e falhas operacionais, que podem comprometer significativamente a oferta de eletricidade. Ao longo das últimas décadas, o país

vivenciou diferentes episódios que evidenciam essa fragilidade, com destaque para os eventos de 2001–2002, 2009, 2018, 2020, 2021 e 2023.

A crise do apagão de 2001–2002 foi provocada por uma combinação de estiagens severas e limitações na expansão da infraestrutura elétrica. O cenário resultou em medidas de racionamento em nível nacional e na drástica redução da geração hidrelétrica, impactando significativamente as séries históricas do setor [29]. Esse episódio evidenciou a importância de mecanismos de planejamento e previsão no setor elétrico, como forma de antecipar riscos e garantir a segurança no fornecimento de energia.

Em 10 de novembro de 2009, uma falha em três linhas de transmissão provenientes da Usina Hidrelétrica de Itaipu ocasionou o desligamento automático de 20 turbinas. Esse evento deixou cerca de 90 milhões de pessoas sem energia no Brasil e no Paraguai. Quatro estados brasileiros foram integralmente afetados, e aproximadamente 90% do território paraguaio também ficou sem luz, revelando a interdependência e vulnerabilidade das redes interligadas entre os dois países [8].

Em 2018, o país enfrentou um apagão de grandes proporções que atingiu todas as regiões do Brasil, com maior intensidade no Norte e Nordeste. A causa principal foi uma falha em linhas de transmissão, revelando a vulnerabilidade da infraestrutura de transporte de energia [15]. Dois anos depois, em 2020, o estado do Amapá sofreu um apagão de longa duração provocado por um incêndio em uma subestação elétrica, o que deixou grande parte da população sem energia por vários dias e expôs a fragilidade das redes isoladas [13].

Em 2021, o setor enfrentou a pior crise hídrica em 91 anos, com volumes críticos de precipitação nas bacias do Sudeste e Centro-Oeste, regiões responsáveis por grande parte da capacidade de armazenamento de água no país. Apesar da gravidade, não houve racionamento oficial, o que foi atribuído à diversificação da matriz elétrica desde 2001. A participação da energia hidrelétrica, que antes representava 85% da matriz, caiu para 61%, abrindo espaço para fontes complementares como eólica, solar, termelétrica e nuclear [11, 23].

Mais recentemente, em agosto de 2023, o Brasil registrou um novo apagão de grandes proporções, que afetou diversas regiões do país, incluindo o Norte e o Nordeste. O evento foi atribuído a falhas em equipamentos de controle e à instabilidade no sistema elétrico nacional [8]. Esse episódio, embora de natureza distinta dos anteriores, reforça a necessidade de atenção contínua à confiabilidade da rede e à previsibilidade da geração de energia.

Entre os eventos analisados, destacam-se, no contexto deste estudo, aqueles de natureza hidrológica que afetaram diretamente a geração de energia hidrelétrica, em especial as crises de 2001–2002 e de 2021. Os demais episódios — ocorridos em 2009, 2018, 2020 e 2023 — estiveram associados a falhas operacionais ou estruturais no sistema elétrico, sem relação direta com a escassez hídrica, e, por isso, não constituem foco da presente análise.

Além desses casos emblemáticos, outros eventos pontuais, como secas regionais e sobrecargas no sistema, também contribuíram para oscilações na geração hidrelétrica ao longo do tempo. Esse histórico reforça a necessidade de incorporar variáveis hidrológicas e climáticas nas análises de geração de energia, bem como de adotar instrumentos de planejamento e previsão que

permitam antecipar padrões de produção, identificar riscos de escassez e otimizar a operação do sistema. Nesse cenário, métodos estatísticos de previsão, como os modelos ARIMA e SARIMA, assumem papel estratégico por sua capacidade de capturar tendências e sazonalidades em séries temporais. No presente estudo, optou-se pelo modelo SARIMA, dada sua adequação à estrutura sazonal mensal da geração de energia hidrelétrica no Brasil.

2.2 SÉRIES TEMPORAIS

As séries temporais representam conjuntos de observações de uma variável registradas ao longo do tempo, respeitando uma sequência cronológica. Na estatística, uma série temporal é definida como um conjunto de observações $Z(t)$ que são tomadas em instantes de tempo específicos [24]. Quando as observações são coletadas continuamente, a série é dita contínua, quando são obtidas em intervalos discretos e regulares Δt , a série é denominada discreta. Um aspecto que distingue a análise de séries temporais de outras técnicas estatísticas é a dependência inerente entre as observações, tornando essencial preservar a ordem cronológica dos dados para garantir a validade da análise [12].

2.3 PROCESSOS ESTOCÁSTICOS

No contexto da análise de séries temporais, é importante compreender que os dados observados ao longo do tempo podem ser interpretados como uma realização particular de um processo estocástico. Um processo estocástico é definido como uma coleção de variáveis aleatórias $Z(t)$ indexadas por um conjunto de tempos T , geralmente o conjunto dos números inteiros ou dos números reais [24].

Formalmente, um processo estocástico é representado por:

$$Z = \{Z(t), t \in T\} \quad (2.1)$$

em que, para cada instante de tempo t , $Z(t)$ é uma variável aleatória definida em um mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$.

Ao fixar uma trajetória $\omega \in \Omega$, obtém-se uma sequência temporal de realizações, caracterizando o que chamamos de série temporal. Dessa maneira, uma série temporal pode ser vista como uma amostra particular de um processo aleatório subjacente.

Para fins de modelagem, é comum restringir o estudo a determinadas classes de processos estocásticos, principalmente:

- **Processos estacionários:** cuja distribuição de probabilidade não varia ao longo do tempo, sendo essencial para a aplicação de muitos métodos de séries temporais;
- **Processos normais ou não-normais:** dependendo se as variáveis seguem ou não a distribuição normal;

- **Processos markovianos:** caracterizados por uma dependência que se limita apenas ao valor imediatamente anterior da série.

Neste trabalho, o foco será nos processos estacionários, dada sua relevância para o ajuste de modelos de previsão como os da classe ARIMA e SARIMA.

2.4 COMPONENTES DE UMA SÉRIE TEMPORAL

Uma série temporal pode ser decomposta em diferentes componentes que representam os padrões sistemáticos e aleatórios observados ao longo do tempo. De forma geral, a decomposição clássica considera quatro principais componentes: tendência, sazonalidade, ciclo e ruído aleatório [24].

- T_t : componente de tendência, que representa a direção geral da série ao longo do tempo;
- S_t : componente de sazonalidade, associada a variações regulares e previsíveis dentro de um período específico;
- C_t : componente cíclica, referente a flutuações de longo prazo que não seguem uma periodicidade fixa, muitas vezes influenciadas por fatores econômicos, climáticos ou estruturais;
- ϵ_t : termo aleatório (ou erro), que representa as variações imprevisíveis da série, assumindo média zero e variância constante.

Essa decomposição é fundamental para a análise e a modelagem de séries temporais, pois a identificação e o tratamento dessas componentes possibilitam compreender melhor o comportamento dos dados e aplicar técnicas adequadas de previsão.

Existem diferentes formas de decomposição, sendo as mais comuns a aditiva e a multiplicativa. A decomposição aditiva é expressa por:

$$Z_t = T_t + S_t + C_t + \epsilon_t \quad (2.2)$$

Enquanto a decomposição multiplicativa é representada por:

$$Z_t = T_t \times S_t \times C_t \times \epsilon_t \quad (2.3)$$

A escolha entre os modelos aditivo e multiplicativo depende da análise da variabilidade dos dados: quando a amplitude das flutuações aumenta proporcionalmente ao nível da série, o modelo multiplicativo é mais adequado; caso contrário, adota-se o modelo aditivo, no qual as componentes são independentes e somadas linearmente, mantendo a variância aproximadamente constante ao longo do tempo.

Quando se identifica uma estrutura multiplicativa, é comum aplicar transformações nos dados, como a função logarítmica ou a raiz quadrada, para estabilizar a variância e permitir o uso de modelos estatísticos que assumem aditividade, como os modelos ARIMA e SARIMA.

Cabe destacar que, embora a decomposição teórica reconheça a existência de um componente cíclico, modelos estatísticos como o SARIMA não o tratam de forma explícita. Nesse tipo de abordagem, variações de longo prazo que poderiam ser atribuídas ao ciclo são geralmente capturadas pela estrutura autoregressiva e pelos termos de diferenciação do modelo.

2.4.1 TENDÊNCIA

A tendência é o componente que reflete movimentos de longo prazo na série, como um crescimento ou decréscimo persistente ao longo do tempo [24]. Ela pode ser linear ou não linear e é, geralmente, associada a fatores estruturais, como mudanças econômicas, tecnológicas ou demográficas.

Para identificar a presença de tendência, uma análise gráfica inicial é recomendada, complementada por testes estatísticos formais [3]. Neste trabalho, a verificação da existência de tendência foi realizada por meio do teste de Cox-Stuart.

O teste de Cox-Stuart baseia-se na comparação entre pares de observações da série, buscando identificar a existência de uma tendência persistente ao longo do tempo [9]. Sejam $Z_1, Z_2, \dots, Z_n$ as observações da série temporal. As observações são divididas em dois grupos: o primeiro com os $n/2$ primeiros valores, e o segundo com os $n/2$ últimos valores. Para cada par (Z_i, Z_{i+c}) , onde c é o deslocamento entre os grupos, verifica-se o sinal da diferença:

$$d_i = Z_{i+c} - Z_i \quad (2.4)$$

Define-se então:

- Um sinal positivo (+) se $d_i > 0$,
- Um sinal negativo (−) se $d_i < 0$,
- Ignorando casos de empate ($d_i = 0$).

A estatística do teste é o número de sinais positivos (T), com as hipóteses formuladas mostradas a seguir:

H_0 : Não existe tendência na série temporal

H_1 : Existe tendência crescente ou decrescente na série temporal

Sob a hipótese nula (H_0), assume-se que a quantidade de sinais positivos e negativos segue uma distribuição binomial $B(n, 0,5)$, caso ($n \leq 20$). Se o número de sinais positivos (ou negativos) for muito alto ou muito baixo em relação ao esperado, rejeita-se (H_0), indicando a presença de tendência.

Para amostras grandes ($n > 20$), a distribuição binomial pode ser aproximada por uma distribuição normal com os mesmos parâmetros de média e variância. Nesse caso, sob (H_0), tem-se:

$$\mu = \frac{n}{2}, \quad \sigma^2 = \frac{n}{4}$$

e a estatística de teste padronizada é:

$$Z = \frac{T - \mu}{\sigma}$$

em que T representa o número de sinais positivos observados.

O procedimento consiste em:

- Agrupar os pares (Z_i, Z_{i+c}) ;
- Contabilizar o número de sinais positivos (T);
- Calcular a estatística de teste;
- Comparar com os valores críticos da distribuição binomial (ou da distribuição normal padrão, se n grande) para o nível de significância α estipulado.

Se a estatística de teste ultrapassar o valor crítico, rejeita-se a hipótese nula de ausência de tendência.

2.4.2 SAZONALIDADE

A sazonalidade refere-se a padrões sistemáticos que se repetem em intervalos regulares de tempo, como ciclos anuais, trimestrais ou mensais. Esse comportamento é muitas vezes observado em séries econômicas, ambientais e de consumo de energia, como no presente estudo.

A sazonalidade pode ser de natureza determinística, quando apresenta comportamento regular e previsível, ou estocástica, quando suas flutuações variam ao longo dos ciclos [24].

A identificação da sazonalidade é essencial para a escolha de modelos adequados. Embora modelos ARIMA possam capturar certos padrões de repetição de forma indireta, os modelos SARIMA são mais indicados quando há sazonalidade explícita, pois incorporam esses efeitos diretamente em sua estrutura.

A presença da sazonalidade em uma série temporal pode ser removida por meio de um procedimento denominado ajuste sazonal. Este procedimento visa eliminar o padrão periódico, permitindo analisar a série sem a influência das oscilações sazonais. Para isso, primeiramente estima-se a componente sazonal S_t , por meio de uma estimativa $\hat{S}_t$, e em seguida ajusta-se a série observada. Dependendo da natureza do modelo, o ajuste pode ser realizado das seguintes formas:

- Para modelos aditivos:

$$Z_t^{SA} = Z_t - \hat{S}_t \quad (2.5)$$

- Para modelos multiplicativos:

$$Z_t^{SA} = \frac{Z_t}{\hat{S}_t} \quad (2.6)$$

Segundo Morettin e Toloi (2006) [24], a sazonalidade determinística ocorre quando o padrão sazonal é regular e estável, podendo ser previsto a partir de dados anteriores. Para séries com esse tipo de sazonalidade, modelos de regressão são adequados para a estimativa da componente sazonal. Por outro lado, na sazonalidade estocástica, as variações do padrão sazonal ao longo do tempo tornam o comportamento menos previsível, sendo recomendada a utilização de métodos baseados em médias móveis para a obtenção de $\hat{S}_t$.

Para testar a existência de uma estrutura sazonal determinística, pode-se utilizar o teste de Fisher baseado na análise espectral. Esse teste avalia a presença de uma frequência dominante na série temporal, o que indicaria a existência de um padrão periódico recorrente.

A estatística de teste é definida por:

$$g = \frac{\max I(\omega_j)}{\sum_{j=1}^{\lfloor n/2 \rfloor} I(\omega_j)} \quad (2.7)$$

em que $I(\omega_j)$ representa o valor do periodograma na frequência ω_j , e n é o número total de observações da série. O numerador corresponde ao maior pico do periodograma, enquanto o denominador representa a soma total da energia espectral da série. O teste verifica se a energia está significativamente concentrada em uma única frequência, sugerindo sazonalidade.

As hipóteses testadas são:

H_0 : a série é ruído branco, ou seja, não há componente sazonal dominante

H_1 : existe uma frequência dominante, indicando a presença de sazonalidade

A distribuição da estatística g sob a hipótese nula é derivada de resultados assintóticos e pode ser usada para obtenção do valor crítico do teste. O procedimento não exige suposições de normalidade ou homocedasticidade, mas pressupõe que a série seja estacionária sob (H_0).

2.5 ESTACIONARIEDADE

A estacionariedade é uma propriedade fundamental na análise de séries temporais, sendo condição necessária para a aplicação de diversos métodos de modelagem e previsão. Uma série temporal é dita estacionária quando suas características estatísticas, como média, variância e autocovariância, permanecem constantes ao longo do tempo [24].

Formalmente, uma série Z_t é considerada estacionária em sentido amplo se satisfizer as seguintes condições:

- A média $E(Z_t)$ é constante para todo t ;
- A variância $Var(Z_t)$ é finita e constante para todo t ;
- A autocovariância $Cov(Z_t, Z_{t+h})$ depende apenas do deslocamento h , e não do tempo t .

Além da estacionariedade em sentido amplo, existe também a definição de estacionariedade estrita, na qual todas as distribuições conjuntas de Z_t são invariantes em relação a mudanças

no tempo. No entanto, na prática, especialmente em aplicações envolvendo séries econômicas e financeiras, considera-se geralmente a estacionariedade em sentido amplo como suficiente para a modelagem.

Séries temporais reais, como aquelas relacionadas ao consumo de energia elétrica, frequentemente não são estacionárias em sua forma original, apresentando tendências, sazonalidades ou heterocedasticidade. A não estacionariedade pode comprometer a eficácia dos modelos de previsão, tornando necessária a aplicação de transformações que visem estabilizar a média e a variância da série.

Uma das técnicas mais comuns para tornar uma série estacionária é a diferenciação, que será abordada no próximo tópico.

2.6 DIFERENCIAÇÃO

A diferenciação é uma técnica utilizada para transformar séries temporais não estacionárias em séries estacionárias, requisito fundamental para a aplicação de modelos como ARIMA e SARIMA. O objetivo da diferenciação é eliminar tendências ou padrões sazonais que possam estar presentes nos dados, estabilizando a média da série ao longo do tempo [24].

A diferenciação simples de primeira ordem consiste em calcular as diferenças sucessivas entre valores consecutivos da série temporal, sendo definida como:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.8)$$

Em casos em que a primeira diferenciação não é suficiente para estabilizar a série, pode-se aplicar a diferenciação de segunda ordem:

$$\Delta^2 Z_t = \Delta[\Delta Z_t] = \Delta[Z_t - Z_{t-1}] \quad (2.9)$$

ou seja,

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad (2.10)$$

A diferenciação de ordem n é definida de maneira recursiva, aplicando a operação de diferenciação n vezes consecutivas sobre a série. Assim, a n -ésima diferença de uma série temporal Z_t pode ser expressa como:

$$\Delta^n Z_t = \Delta[\Delta^{n-1} Z_t] \quad (2.11)$$

em que $\Delta^{n-1} Z_t$ representa a diferença de ordem $(n - 1)$.

Essa generalização permite lidar com séries que apresentam estruturas mais complexas de não estacionariedade, exigindo diferenciações sucessivas para estabilizar suas propriedades estatísticas.

Em situações usuais, a aplicação de uma ou, no máximo, duas diferenças já é suficiente para tornar a série temporal estacionária, sendo recomendável evitar diferenciações excessivas que possam introduzir ruído ou instabilidade no modelo.

2.7 MODELOS ARIMA E SARIMA

Os modelos da classe ARIMA (AutoRegressive Integrated Moving Average) são amplamente utilizados na modelagem de séries temporais univariadas, especialmente quando o objetivo é capturar as dependências temporais e produzir previsões futuras. A metodologia ARIMA combina três componentes principais: autorregressivo (AR), de média móvel (MA) e de diferenciação (I) para estabilizar a série [24].

Um modelo $\text{ARIMA}(p, d, q)$ é caracterizado por três parâmetros:

- p : número de defasagens da parte autorregressiva (AR);
- d : número de diferenciações aplicadas para tornar a série estacionária (I);
- q : número de defasagens da parte de média móvel (MA).

Matematicamente, um modelo $\text{ARIMA}(p, d, q)$ pode ser representado por:

$$\phi(B)(1 - B)^d Z_t = \theta(B)\epsilon_t \quad (2.12)$$

em que:

- B : operador de defasagem, tal que $B^k Z_t = Z_{t-k}$;
- d : número de diferenciações aplicadas para tornar a série estacionária;
- $\phi(B)$: polinômio de ordem p associado à parte autorregressiva;
- $\theta(B)$: polinômio de ordem q associado à parte de média móvel;
- ϵ_t : erro aleatório (ruído branco) com média zero e variância constante.

Em muitas séries temporais, especialmente aquelas que exibem padrões sazonais, a simples modelagem ARIMA não é suficiente. Para esses casos, utiliza-se o modelo SARIMA (Seasonal ARIMA), que incorpora componentes sazonais explicitamente.

O modelo $\text{SARIMA}(p, d, q)(P, D, Q)_s$ adiciona três parâmetros sazonais:

- P : ordem da parte autorregressiva sazonal (SAR);
- D : número de diferenciações sazonais necessárias para tornar a série estacionária (SI);
- Q : ordem da parte de média móvel sazonal (SMA);
- s : período da sazonalidade (por exemplo, $s = 12$ para dados mensais).

A estrutura geral de um modelo $\text{SARIMA}(p, d, q)(P, D, Q)_s$ é expressa por:

$$\Phi(B^s)\phi(B)(1 - B)^d(1 - B^s)^D Z_t = \Theta(B^s)\theta(B)\epsilon_t \quad (2.13)$$

em que:

- $\Phi(B^s)$ e $\Theta(B^s)$: polinômios sazonais de ordens P e Q , respectivamente;
- $(1 - B)^d$: diferenciação regular;
- $(1 - B^s)^D$: diferenciação sazonal.

Dessa forma, o modelo SARIMA permite capturar tanto os padrões de curto prazo quanto as estruturas sazonais presentes na série temporal, sendo uma ferramenta robusta para previsão em diferentes contextos.

A construção de modelos ARIMA e SARIMA segue a Metodologia Box-Jenkins, a qual propõe um procedimento sistemático composto pelas etapas de identificação, estimação, verificação de adequação e previsão. Essa metodologia será abordada em seguida, detalhando as técnicas utilizadas para a seleção e validação dos modelos ajustados.

2.8 FUNÇÕES DE AUTOCORRELAÇÃO

A identificação de modelos ARIMA e SARIMA baseia-se, entre outros métodos, na análise das funções de autocorrelação e autocorrelação parcial, conhecidas respectivamente como ACF (Função Autocorrelação) e PACF (Função Autocorrelação Parcial).

A Função Autocorrelação (ACF) mede a correlação linear entre as observações de uma série temporal em diferentes defasagens (lags). A ACF de ordem h é definida como:

$$\rho(h) = \frac{\text{Cov}(Z_t, Z_{t-h})}{\text{Var}(Z_t)} \quad (2.14)$$

em que $\text{Cov}(Z_t, Z_{t-h})$ é a covariância entre Z_t e Z_{t-h} e $\text{Var}(Z_t)$ é a variância da série. A ACF é útil para identificar padrões de dependência ao longo do tempo, como a presença de tendências ou sazonalidades.

Já a Função Autocorrelação Parcial (PACF) mede a correlação entre Z_t e Z_{t-h} eliminando a influência das observações intermediárias. A PACF de ordem h representa a correlação entre Z_t e Z_{t-h} condicionada às observações $Z_{t-1}, Z_{t-2}, \dots, Z_{t-h+1}$.

Enquanto a ACF captura todas as correlações diretas e indiretas, a PACF isola apenas a correlação direta de uma defasagem específica, sendo especialmente útil para identificar o número de termos autorregressivos (p) em um modelo ARIMA.

A interpretação conjunta dos gráficos da ACF e da PACF permite orientar a escolha dos parâmetros p e q :

- A ACF é utilizada principalmente para sugerir o valor de q (ordem da parte de média móvel);
- A PACF é utilizada principalmente para sugerir o valor de p (ordem da parte autorregressiva).

Essa análise será utilizada no processo de identificação dos modelos, conforme descrito na metodologia de Box-Jenkins, e também na avaliação dos resíduos, contribuindo para verificar a adequação do modelo ajustado.

2.9 METODOLOGIA BOX-JENKINS

A Metodologia Box-Jenkins, proposta por Box e Jenkins (1970), é um procedimento sistemático para a construção de modelos de séries temporais, com foco em modelos da classe ARIMA e suas extensões sazonais (SARIMA). O processo é iterativo e composto por quatro etapas principais: identificação, estimação, diagnóstico e previsão [4].

2.9.1 IDENTIFICAÇÃO

Nesta etapa, busca-se determinar a estrutura adequada do modelo, identificando os valores iniciais para os parâmetros p , d , q , P , D e Q . A análise gráfica da série temporal, combinada com o estudo das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), auxilia na escolha dos possíveis modelos candidatos. Se a série apresentar tendência ou sazonalidade, pode ser necessário aplicar diferenciações regular e/ou sazonal para estabilizar a média.

2.9.2 ESTIMAÇÃO

Após a escolha preliminar da estrutura do modelo, procede-se à estimação dos parâmetros autorregressivos e de média móvel. A estimação é realizada, em geral, pelo método da máxima verossimilhança, que consiste em encontrar os valores dos parâmetros que maximizam a probabilidade de observação da série temporal sob o modelo ajustado. Esse método garante que os parâmetros encontrados sejam aqueles que melhor explicam os dados observados.

2.9.3 DIAGNÓSTICO

Nesta fase, verifica-se se o modelo ajustado é adequado. Para isso, analisam-se os resíduos do modelo, que devem se comportar como ruído branco, isto é, não apresentar autocorrelação significativa e possuir variância constante. Testes estatísticos, como o teste de Box-Pierce, são aplicados para avaliar a independência dos resíduos. Caso o modelo não seja satisfatório, o processo retorna à etapa de identificação para ajustes.

2.9.4 PREVISÃO

Com um modelo validado, realiza-se a geração de previsões futuras da série. As previsões são acompanhadas de intervalos de confiança, refletindo a incerteza associada às estimativas.

A Metodologia Box-Jenkins é iterativa por natureza, podendo exigir a repetição das etapas até que um modelo final adequado seja obtido.

2.10 CRITÉRIOS DE SELEÇÃO DE MODELOS

A escolha do modelo mais adequado para representar uma série temporal envolve o balançamento entre a qualidade do ajuste e a complexidade do modelo. Critérios de informação,

como o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC), são amplamente utilizados nesse processo.

O *Akaike Information Criterion* (AIC), proposto por Akaike [2], mede a qualidade de um modelo ajustado em termos de sua verossimilhança penalizada pelo número de parâmetros. O AIC é definido como:

$$AIC = -2 \ln(\hat{L}) + 2k \quad (2.15)$$

em que:

- $\hat{L}$: valor máximo da função de verossimilhança do modelo;
- k : número de parâmetros estimados, correspondendo à soma dos termos autorregressivos (p), de média móvel (q), sazonais (P e Q) e do termo da variância dos resíduos.

Modelos com menores valores de AIC são preferidos, pois indicam melhor equilíbrio entre ajuste e parcimônia.

O *Bayesian Information Criterion* (BIC), proposto por Schwarz [27], propõe uma penalização mais severa para a complexidade do modelo. O BIC é definido como:

$$BIC = -2 \ln(\hat{L}) + k \ln(n) \quad (2.16)$$

em que:

- n : número de observações da série temporal;
- $\hat{L}$ e k : definidos como anteriormente, em (AIC).

Assim como no AIC, modelos com menores valores de BIC são preferíveis. No entanto, por penalizar mais fortemente modelos complexos, o BIC tende a selecionar modelos mais simples do que o AIC.

A utilização conjunta dos critérios AIC e BIC permite uma avaliação robusta na escolha do modelo que melhor descreve a série temporal, considerando tanto o ajuste quanto a complexidade.

2.11 MÉTODOS DE COMPARAÇÃO DE PREVISÃO

A avaliação do desempenho dos modelos ajustados é fundamental para verificar a qualidade das previsões geradas. Para isso, são utilizadas métricas que quantificam o erro entre os valores observados e previstos. Entre as principais métricas adotadas neste trabalho estão a Raiz do Erro Quadrático Médio (RMSE) e o Erro Percentual Absoluto Médio (MAPE).

A Raiz do Erro Quadrático Médio (RMSE), do inglês *Root Mean Squared Error*, mede a raiz quadrada da média dos quadrados dos erros [31]. Sua fórmula é dada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Z_t - \hat{Z}_t)^2} \quad (2.17)$$

em que:

- Z_t : valor observado no tempo t ;
- $\hat{Z}_t$: valor previsto no tempo t ;
- n : número de observações.

O RMSE expressa o erro na mesma unidade dos dados originais, sendo sensível a grandes desvios devido à elevação ao quadrado dos erros.

O Erro Percentual Absoluto Médio (MAPE — *Mean Absolute Percentage Error*) calcula o erro percentual absoluto médio entre os valores previstos e observados [19]. É definido como:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right| \quad (2.18)$$

em que as variáveis são as mesmas definidas anteriormente.

O MAPE fornece uma medida do erro em termos percentuais, facilitando a interpretação e a comparação entre séries de diferentes escalas. Valores mais baixos de RMSE e MAPE indicam melhor desempenho preditivo do modelo.

Não há consenso entre os estatísticos sobre qual método é preferível. Assim, se os erros elevados de previsão forem inaceitáveis, o uso do RMSE faz-se necessário. O MAPE, por sua vez, é utilizado para comparar a precisão de duas séries temporais diferentes. Dessa forma, tanto o RMSE quanto o MAPE serão utilizados como critério de acurácia para a comparação dos modelos de previsão que serão aplicados na metodologia da pesquisa [14].

Essas métricas serão utilizadas para comparar e selecionar o modelo que apresentar o melhor ajuste e a maior capacidade de previsão para a série analisada.

2.12 TESTE DE ADEQUAÇÃO DO MODELO

Após o ajuste de um modelo ARIMA ou SARIMA, é essencial verificar a adequação do modelo aos dados. Para isso, analisa-se o comportamento dos resíduos, que devem se assemelhar a um ruído branco, ou seja, serem não autocorrelacionados e apresentarem variância constante.

O Teste de Box-Pierce [5] é um dos procedimentos clássicos utilizados para essa verificação. Esse teste avalia a hipótese de que as autocorrelações dos resíduos, até uma determinada defasagem, são simultaneamente nulas.

As hipóteses do teste são formuladas da seguinte maneira:

H_0 : Os resíduos não apresentam autocorrelação significativa.

H_1 : Os resíduos apresentam autocorrelação significativa.

A estatística de teste é definida como:

$$Q = n \sum_{h=1}^m r_h^2 \quad (2.19)$$

em que:

- n : número de observações;
- r_h : valor da autocorrelação dos resíduos na defasagem h ;
- m : número de defasagens consideradas no teste.

Sob a hipótese nula (H_0), a estatística Q segue aproximadamente uma distribuição qui-quadrado (χ^2) com $(m - k)$ graus de liberdade, onde k é o número de parâmetros estimados no modelo.

Se o valor observado de Q for superior ao valor crítico da distribuição χ^2 para o nível de significância estabelecido, rejeita-se H_0 , indicando que os resíduos apresentam autocorrelação e que o modelo pode ser inadequado.

A utilização do Teste de Box-Pierce, juntamente com a análise gráfica dos resíduos, por meio das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos, contribui para a validação da qualidade do modelo ajustado.

3. METODOLOGIA

3.1 FONTE DOS DADOS

O presente estudo utilizou dados mensais da geração de energia hidrelétrica no Brasil, medidos em Gigawatts-hora (GWh), coletados por meio da plataforma Ipeadata, mantida pelo Instituto de Pesquisa Econômica Aplicada (IPEA) [20], com base em informações do Operador Nacional do Sistema Elétrico (ONS) [25].

A série temporal analisada compreende o período de janeiro de 1996 a dezembro de 2024, totalizando 348 meses, equivalentes a 29 anos completos. Esse intervalo foi escolhido por corresponder ao período integralmente disponível na plataforma Ipeadata, garantindo a continuidade e a consistência dos dados utilizados na modelagem.

Para fins de modelagem, foram inicialmente utilizadas 336 observações, correspondentes ao intervalo de janeiro de 1996 a dezembro de 2023. As 12 observações restantes, relativas ao ano de 2024, foram reservadas para a etapa de validação, com o objetivo de avaliar a acurácia das previsões geradas pelos modelos.

Ao final do processo, os valores previstos foram comparados com os valores efetivamente observados, permitindo aferir a adequação do modelo adotado.

3.2 FERRAMENTAS COMPUTACIONAIS

Todas as análises e modelagens foram realizadas no software estatístico R, versão 4.3.1 [26], com o apoio dos pacotes `ipeadatar` [16], `randtests` [6], `GeneCycle` [1], `forecast` [18] e `metrics` [17]. O R é um software estatístico de código aberto, amplamente utilizado em pesquisas científicas, que dispõe de recursos computacionais específicos para a análise de séries temporais.

3.3 PROCEDIMENTOS METODOLÓGICOS

O processo metodológico envolveu as seguintes etapas:

1. **Importação dos dados:** Os dados foram primeiramente importados para o software R, utilizando o pacote `ipeadatar` [16], que permite a extração direta de dados do portal Ipeadata dentro do ambiente R.

2. **Separação dos Dados:** A série temporal foi dividida em dois conjuntos. O primeiro, referente ao período de janeiro de 1996 a dezembro de 2023, foi utilizado para a análise exploratória e a modelagem. O segundo, correspondente ao intervalo de janeiro a dezembro de 2024, foi reservado para a etapa de validação, com o intuito de verificar a acurácia das previsões geradas pelos modelos ajustados.
3. **Análise Descritiva:** Com o intuito de avaliar o comportamento geral da série, foram calculadas medidas estatísticas de tendência central e dispersão, como a média e o desvio-padrão. Essa etapa permitiu investigar a necessidade de transformações nos dados, como aplicação de logaritmo ou raiz quadrada, visando estabilizar a variância e adequar a série às premissas dos modelos de previsão.
4. **Análise de Tendência e Sazonalidade:** Foram analisadas as características estruturais da série com o intuito de verificar a presença de tendência e sazonalidade. Para isso, aplicaram-se os testes estatísticos de Cox-Stuart e de Fisher, cujas implementações foram realizadas por meio dos pacotes `randtests` [6] e `GeneCycle` [1], respectivamente, no ambiente R. O teste de Cox-Stuart testa a hipótese nula de ausência de tendência. Caso rejeitada, indica a necessidade de diferenciação da série. Já o teste de Fisher avalia a hipótese nula de ausência de sazonalidade, e sua rejeição implica na adoção de um modelo que contemple estrutura sazonal, como os modelos da classe SARIMA, entre outras abordagens possíveis. Neste estudo, optou-se pelo uso do modelo SARIMA em razão de sua adequação à estrutura da série analisada.
5. **Diferenciação da Série:** Com base nos resultados do teste de Cox-Stuart, procedeu-se à diferenciação regular de primeira ordem, com o objetivo de eliminar a tendência da série e torná-la estacionária. A efetividade da transformação foi avaliada por meio da reaplicação do teste de Cox-Stuart sobre a série diferenciada, bem como da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF). A convergência das defasagens nessas funções, observada graficamente, corroborou a remoção da tendência.
6. **Identificação dos Modelos:** Com base nas funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série temporal diferenciada, foram definidos os valores iniciais das ordens dos componentes autorregressivo (AR), de média móvel (MA) e sazonais (SAR e SMA), culminando na proposição de modelos da classe SARIMA, adequados para capturar a estrutura regular e sazonal da série.
7. **Estimação dos Modelos:** A estimação dos parâmetros foi realizada por meio do método da máxima verossimilhança, utilizando as funções do pacote `forecast` [18] no ambiente R. Inicialmente, empregou-se o comando `auto.arima`, que sugere uma estrutura de modelo com base em critérios de informação e testes estatísticos automáticos. No entanto, a escolha final do modelo foi conduzida de forma manual e iterativa, por meio de sucessivos ajustes, considerando o comportamento das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos, bem como a significância estatística dos coeficientes

estimados e a adequação dos diagnósticos residuais.

8. **Seleção do Modelo:** A escolha do modelo final considerou a análise conjunta dos critérios de informação AIC (*Akaike Information Criterion*) e BIC (*Bayesian Information Criterion*), bem como das métricas de avaliação da qualidade preditiva RMSE (*Root Mean Squared Error*) e MAPE (*Mean Absolute Percentage Error*). Estas métricas foram calculadas a partir dos valores previstos e observados no conjunto de validação, utilizando funções do pacote `Metrics` [17].
9. **Validação do Modelo:** A qualidade do ajuste foi verificada por meio da análise dos resíduos do modelo estimado. Foram avaliadas a ausência de autocorrelação e a constância da variância, com apoio do teste de Box-Pierce, cuja hipótese nula assume que os resíduos se comportam como ruído branco.
10. **Geração de Previsões:** Após a validação do modelo, foram geradas previsões para os doze meses do ano de 2024, com base nos parâmetros ajustados. As previsões foram acompanhadas de intervalos de confiança, permitindo avaliar a incerteza associada às estimativas.

Ao final desse processo metodológico, obteve-se um modelo ajustado que considera tanto a estrutura de dependência temporal quanto os padrões sazonais da série. Os resultados obtidos com a aplicação desse modelo, bem como a avaliação de sua performance preditiva, são apresentados no capítulo seguinte.

4. RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados obtidos a partir da análise e da modelagem da série temporal da geração mensal de energia hidrelétrica no Brasil, abrangendo o período de janeiro de 1996 a dezembro de 2023. Conforme ilustrado na Figura 4.1, a análise visual da série ao longo desse intervalo revela indícios claros de tendência e sazonalidade, indicando que a série não é estacionária.

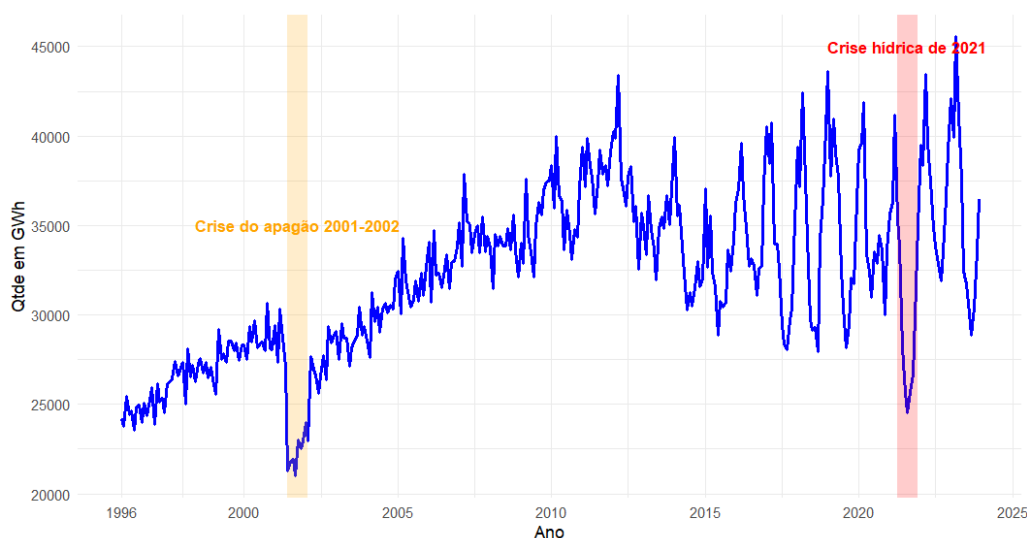


Figura 4.1: Série temporal da geração de energia hidrelétrica no Brasil de jan/1996 a dez/2023.

Fonte: Elaborado pelo autor

Um aspecto que merece destaque é a queda acentuada na geração de energia observada entre os anos de 2001 e 2002, e novamente em 2021. Ambos os períodos coincidem com crises hídricas amplamente documentadas, que impactaram significativamente o sistema elétrico nacional. No primeiro caso, medidas de racionamento foram impostas; no segundo, apesar da gravidade, a diversificação da matriz contribuiu para evitar cortes no fornecimento. Esses episódios, embora excepcionais, não alteraram de forma permanente a estrutura da série, que retoma seu padrão regular nos anos subsequentes [11, 29].

Com o objetivo de determinar a necessidade de aplicação de transformações estabilizadoras de variância, foi realizada uma análise da relação entre as médias e as amplitudes da série. Para isso, a série foi particionada em pares de observações, resultando em 168 médias e amplitudes correspondentes.

A inspeção visual da Figura 4.2 indica que a variância dos dados não depende da média, o que sugere um comportamento aditivo da série. Portanto, não foi necessária a aplicação

de transformações, como a transformação logarítmica ou a raiz quadrada, que são usualmente recomendadas para séries de natureza multiplicativa, em que a variabilidade cresce com o nível da série.

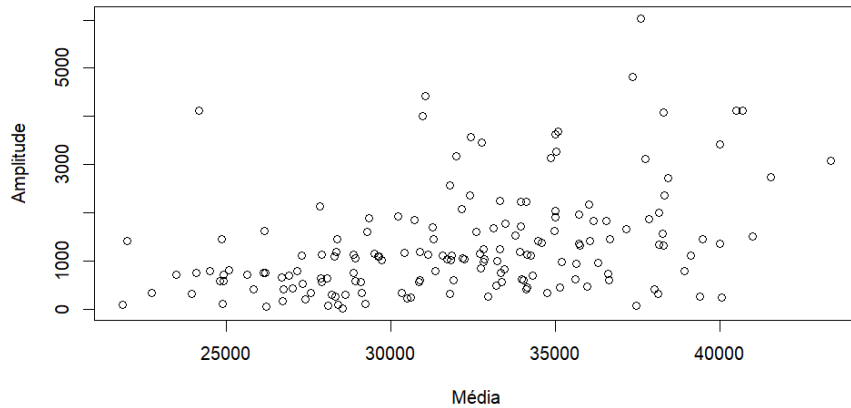


Figura 4.2: Relação das médias com as amplitudes de cada par de observações.

Fonte: Elaborado pela autora

Prosseguindo com as análises gráficas da série, a Função de Autocorrelação (ACF), ilustrada na Figura 4.3, auxilia na avaliação da estacionariedade. A presença de lags significativos e padrões repetitivos indica a existência de sazonalidade, enquanto o decaimento lento dos valores sugere a presença de tendência e, portanto, a não estacionariedade da série.

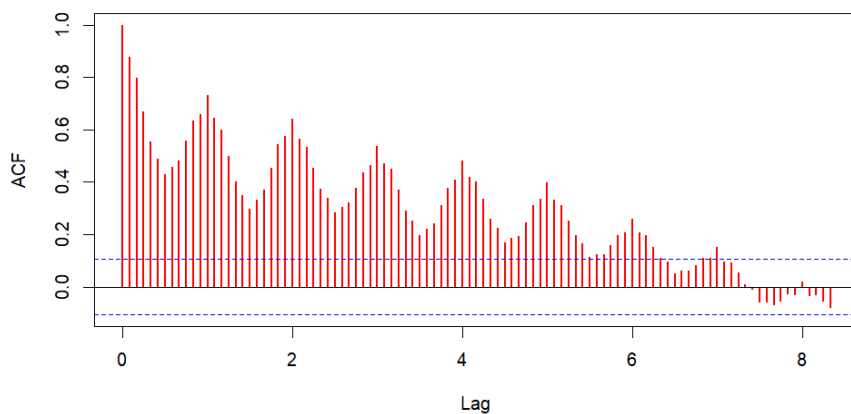


Figura 4.3: Função de Autocorrelação da série temporal (ACF).

Fonte: Elaborado pela autora

Para confirmar se realmente há a presença de tendência e sazonalidade na série temporal, foram aplicados o teste de Cox-Stuart e o teste de Fisher, respectivamente. O teste de Cox-Stuart, cuja hipótese nula assume a ausência de tendência, apresentou um valor-p ($2,2 \times 10^{-16}$) significativamente inferior ao nível nominal de 5%, permitindo a rejeição da hipótese nula, indicando então a existência de uma tendência significativa na série. Da mesma forma, o teste

de Fisher, cuja hipótese nula assume a ausência de sazonalidade, também apresentou um valor-p ($5,71 \times 10^{-39}$) significativamente inferior ao nível nominal de 5%, evidenciando a existência de sazonalidade na série.

Assim, confirmada a presença desses componentes, e considerando que a tendência compromete a estacionariedade da série, procedeu-se à diferenciação dos dados para estabilizar a média e eliminar essas componentes.

Foi aplicada na série uma diferenciação em grau 1, tendo suas observações substituídas pelas diferenças entre os dados originais adjacentes. O objetivo aqui é eliminar a componente de tendência, fazendo com que o processo se torne estacionário. A série resultante após a diferenciação é apresentada na Figura 4.4, e neste caso, fica nítida a mudança de comportamento, com a média constante e menor variação.

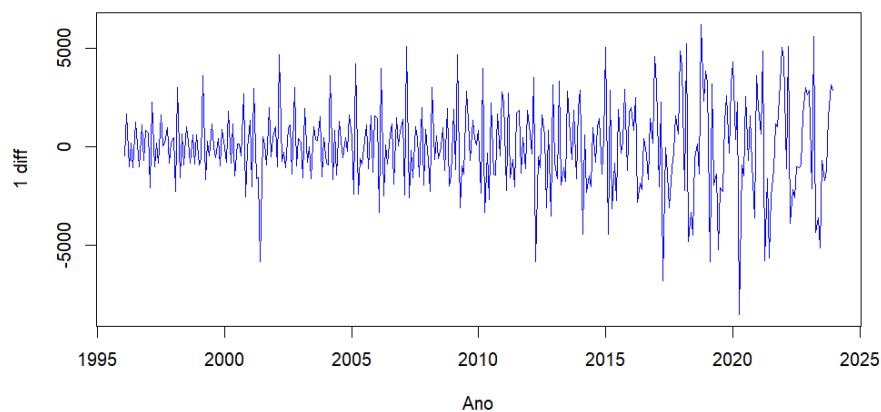


Figura 4.4: Série diferenciada em grau 1.

Fonte: Elaborado pela autora

Para verificar os efeitos da diferenciação, os testes estatísticos foram reaplicados. No caso da tendência, o teste de Cox-Stuart resultou em um valor-p = 0,2786, valor superior ao nível nominal de significância de 5%. Dessa forma, não se rejeita a hipótese nula de ausência de tendência, confirmando que a diferenciação de primeira ordem foi eficaz para eliminar esse componente da série temporal.

A eliminação da tendência é um passo fundamental para garantir a estacionariedade da série, permitindo o avanço para a etapa de identificação do modelo adequado de previsão. Logo, com a eliminação da componente tendência, conclui-se que a série passou a atender à condição de estacionariedade também.

Em relação à sazonalidade, o teste de Fisher aplicado à série diferenciada resultou em um valor-p de aproximadamente $1,60 \times 10^{-12}$, valor significativamente inferior ao nível nominal de 5%. Dessa forma, rejeita-se a hipótese nula de ausência de sazonalidade, indicando que, mesmo após a diferenciação regular, a série ainda preserva um padrão sazonal significativo.

A análise visual das Funções de Autocorrelação (ACF) e de Autocorrelação Parcial (PACF) da série diferenciada, apresentadas nas Figuras 4.5(a) e 4.5(b), reforça os resultados obtidos nos

testes estatísticos. Elas revelam que os lags deixaram de apresentar um padrão de decaimento lento, característico de uma série com tendência, confirmando que a diferenciação regular foi eficaz para torná-la estacionária. No entanto, observa-se a persistência de lags significativos em múltiplos de 12 na (ACF), o que evidencia a manutenção de uma estrutura sazonal. Esses padrões reforçam a necessidade de utilizar um modelo que incorpore componentes sazonais, como o modelo SARIMA.

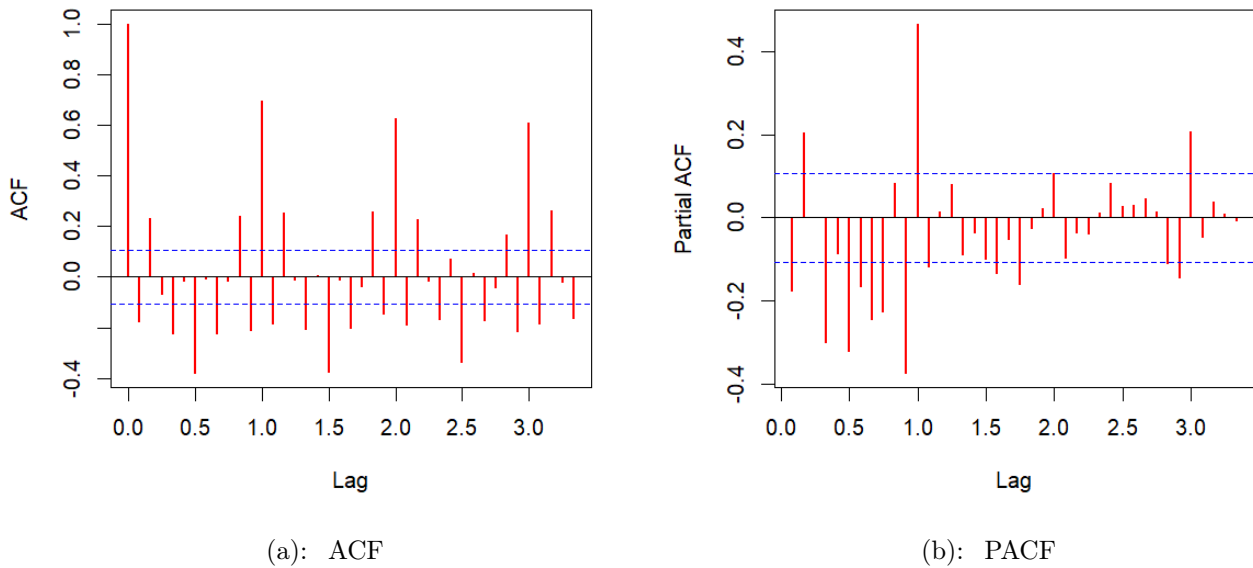


Figura 4.5: Autocorrelação (ACF) e autocorrelação parcial (PACF) da série diferenciada.

Fonte: Elaborado pelo autor

Com as componentes devidamente tratadas, prosseguiu-se para a etapa de modelagem da série, com o objetivo de identificar a estrutura mais adequada para representar seu comportamento. Inicialmente, foi utilizado o comando `auto.arima`, que seleciona automaticamente um modelo com base em critérios estatísticos e de informação. No entanto, para garantir maior rigor analítico, essa sugestão automática foi comparada com diversos outros modelos ajustados manualmente por meio da função `Arima`, testando diferentes combinações de parâmetros.

Apesar da sugestão automática fornecida pelo comando `auto.arima`, o modelo proposto não foi adotado neste estudo. A análise da significância estatística dos coeficientes indicou que nem todos eram significativos ao nível nominal de 5%, o que compromete sua validade. Diante disso, optou-se por descartar esse modelo.

Em seguida, foram testadas diversas combinações de parâmetros de forma manual, por meio da função `Arima`, aplicando um processo iterativo baseado em tentativa e erro. A seleção dos modelos foi guiada por três critérios principais: significância estatística dos coeficientes estimados, análise dos resíduos e verificação de que os resíduos apresentavam comportamento semelhante ao de ruído branco. Ao final desse processo, três modelos foram considerados adequados e seus resultados estão apresentados na Tabela 4.1.

Tabela 4.1: Modelos ajustados com parâmetros significativos e valores de critérios de seleção.

Modelo	AIC	BIC	RMSE	MAPE
SARIMA(2,1,1)(0,1,1) ₁₂	5576,49	5595,38	1287,92	2,84%
SARIMA(1,1,2)(0,1,1) ₁₂	5577,32	5596,21	1289,83	2,84%
SARIMA(0,1,0)(2,1,0) ₁₂	5620,16	5631,49	1399,00	3,05%

Fonte: Elaborado pela autora

Entre os modelos ajustados, a seleção do mais adequado baseou-se na comparação dos critérios de informação Akaike (AIC) e de Bayes (BIC), bem como das métricas de avaliação preditiva RMSE e MAPE. Conforme apresentado na Tabela 4.1, o modelo SARIMA(2,1,1)(0,1,1)₁₂ obteve os melhores desempenhos em todos os critérios considerados.

Dessa forma, esse modelo foi selecionado para representar a série temporal da geração de energia hidrelétrica no Brasil, por apresentar o melhor equilíbrio entre ajuste e capacidade preditiva.

A Tabela 4.2 apresenta os parâmetros estimados para o modelo SARIMA(2,1,1)(0,1,1)₁₂. Observa-se que todos os coeficientes possuem valores-p significativamente inferiores a 5%, indicando significância estatística para a estimação do modelo.

Tabela 4.2: Parâmetros estimados para o modelo SARIMA(2,1,1)(0,1,1)₁₂.

Modelo	Parâmetro	Estatística	Desvio Padrão	Valor-p
SARIMA(2,1,1)(0,1,1) ₁₂	ϕ_1	0,9737	0,0595	< 0,01
	ϕ_2	-0,1626	0,0587	< 0,01
	θ_1	-0,9630	0,0293	< 0,01
	Θ_1	-0,7333	0,0411	< 0,01

Fonte: Elaborado pela autora

Para validar a adequação final do modelo, foram analisadas as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos resíduos do modelo SARIMA(2,1,1)(0,1,1)₁₂. Os resultados indicaram, no máximo, dois lags estatisticamente significativos, o que, dentro de limites aceitáveis, sugere ausência de autocorrelações relevantes. Isso reforça que os resíduos são aproximadamente independentes entre si e que o modelo foi bem ajustado, sendo capaz de capturar a estrutura da série temporal de forma satisfatória, conforme ilustrado nas Figuras 4.6(a) e 4.6(b).

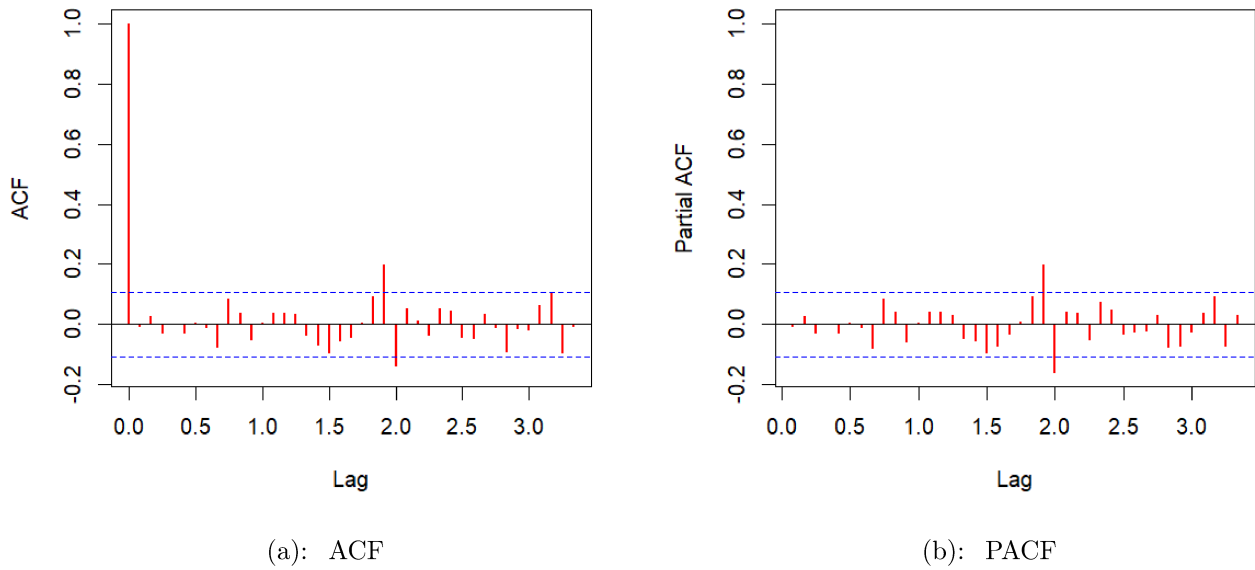


Figura 4.6: Autocorrelação (ACF) e autocorrelação parcial (PACF) residual.

Fonte: Elaborado pelo autor

O teste de Box-Pierce foi aplicado aos resíduos do modelo $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ com o objetivo de avaliar se estes se comportam como ruído branco, isto é, se não apresentam autocorrelação significativa. A hipótese nula do teste assume que não há autocorrelação nos resíduos. O resultado obtido apresentou um valor-p de 0,9205, valor significativamente superior ao nível de significância de 5%. Dessa forma, não se rejeita a hipótese nula, indicando que os resíduos podem ser considerados ruído branco.

Essa conclusão é corroborada pela análise gráfica do periodograma acumulado, representado na Figura 4.7, que reforça a ausência de padrões relevantes não modelados na série residual.

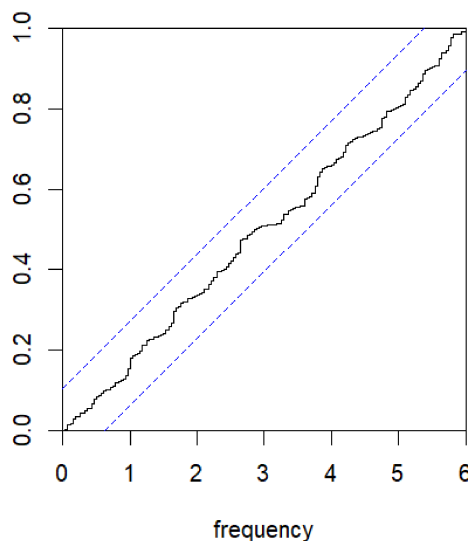


Figura 4.7: Periodograma acumulado residual.

Fonte: Elaborado pela autora

A Figura 4.8 apresenta a comparação entre os valores observados (realizados) e os valores previstos pelo modelo SARIMA(2,1,1)(0,1,1)₁₂ para o ano de 2024. De maneira geral, observa-se que o modelo conseguiu representar bem o comportamento sazonal da série, especialmente a tendência de queda da geração de energia nos meses de outono e inverno (abril a setembro), seguida por uma recuperação a partir de outubro.



Figura 4.8: Valores observados e previsto do modelo SARIMA(2,1,1)(0,1,1)₁₂.

Fonte: Elaborado pela autora

Nos primeiros cinco meses do ano, de janeiro a maio, observa-se que o modelo estimou valores ligeiramente inferiores aos efetivamente observados. Essa diferença pode estar associada a oscilações atípicas no regime de chuvas ou a variações inesperadas na demanda por energia. Esse período corresponde, em grande parte, à estação chuvosa em regiões com forte concentração de usinas hidrelétricas, como o Sudeste e o Centro-Oeste. Nesses meses, a geração de energia pode ser influenciada por fenômenos climáticos não capturados pelo modelo, como desvios nas afluições ou anomalias associadas a eventos extremos (ex.: El Niño ou La Niña). Além disso, ajustes operacionais pontuais no sistema interligado nacional, realizados com base em condições hidrológicas em tempo real, também podem contribuir para essa diferença. Tais fatores reforçam a complexidade de prever a geração hidrelétrica no curto prazo, mesmo com modelos estatísticos bem ajustados.

A partir de maio, os valores previstos passam a se aproximar mais dos valores reais, com destaque para os meses de junho a agosto, novembro e dezembro, em que as curvas praticamente coincidem.

Essa boa correspondência, observada principalmente entre junho e agosto, meses tipicamente secos no calendário hidrológico, reforça a capacidade do modelo em capturar os padrões regulares da série temporal. Nos últimos meses do ano, os valores previstos se mantiveram bastante próximos dos observados, apresentando pequenas variações positivas. Além disso, todas as previsões permaneceram dentro dos intervalos de confiança de 95%, evidenciando a adequa-

ção do modelo na representação da sazonalidade e na identificação de momentos característicos de queda e recuperação na geração de energia hidrelétrica.

Os valores previstos, os realizados e os respectivos erros absolutos percentuais para cada mês de 2024 estão detalhados na Tabela 4.3, permitindo uma análise quantitativa do desempenho do modelo.

Tabela 4.3: Comparação dos valores realizados e previstos em (GWh), com seus erros absolutos.

Mês/Ano	Realizado (GWh)	Previsto (GWh)	Erro Absoluto (%)
Jan/2024	42.069,18	39.707,96	5,61%
Fev/2024	41.370,46	38.285,06	7,46%
Mar/2024	45.551,63	42.766,17	6,11%
Abr/2024	41.183,61	38.175,73	7,30%
Mai/2024	37.558,10	36.110,39	3,85%
Jun/2024	32.440,63	32.376,90	0,20%
Jul/2024	31.743,43	31.780,34	0,12%
Ago/2024	30.052,58	30.703,39	2,16%
Set/2024	28.850,32	30.472,47	5,62%
Out/2024	30.397,41	32.097,58	5,59%
Nov/2024	33.548,82	33.635,16	0,26%
Dez/2024	36.437,94	36.969,64	1,46%

Fonte: Elaborado pela autora

Os resultados obtidos neste estudo, com o modelo SARIMA(2,1,1)(0,1,1)₁₂, apresentaram métricas preditivas satisfatórias, evidenciadas pelos baixos valores de RMSE (1.287,92) e MAPE (2,84%), conforme apresentado na Tabela 4.1, e se mostram baixos quando comparados à magnitude dos dados observados, que giram em torno de dezenas de milhares de GWh. Além disso, a ausência de autocorrelação nos resíduos indica um bom ajuste ao comportamento sazonal da série.

Para além dos bons resultados obtidos no presente estudo, é importante refletir sobre esses resultados e colocá-los em perspectiva com outras abordagens existentes na literatura científica. Dentre os estudos relevantes, destacam-se os trabalhos de Silva (2023) [28] e Freitas (2022) [14], ambos voltados à previsão no setor elétrico brasileiro e que também utilizaram o modelo SARIMA como uma das metodologias principais de análise.

No primeiro caso, o estudo de Silva (2023), intitulado “*Geração de energia hidrelétrica: uma análise comparativa de métodos de previsão estatísticos e de inteligência artificial aplicados a dados da região Sul do Brasil*”, teve como objetivo avaliar o desempenho de diferentes modelos preditivos na estimativa da geração de energia hidrelétrica nos estados do Paraná, Santa Catarina e Rio Grande do Sul, utilizando dados mensais de junho de 2013 a junho de 2023. Entre os modelos testados estavam o SARIMA, o SVR (Support Vector Regression) e combinações simples por média dos dois métodos.

O estudo aplicou os modelos individualmente em cada estado analisado, considerando diferentes horizontes de previsão (1, 3, 6 e 12 meses), o que permitiu comparar o desempenho conforme a complexidade e o tempo projetado. Embora o modelo SARIMA tenha apresentado resultados satisfatórios, sobretudo em séries com menor variabilidade, o modelo SVR apresentou o melhor desempenho preditivo para o horizonte de 12 meses em todos os estados analisados, superando o SARIMA nos critérios de erro utilizados. Essa superioridade do SVR foi atribuída à sua capacidade de capturar padrões não lineares e flutuações locais, aspectos menos explorados pelos modelos estatísticos tradicionais.

Apesar disso, o desempenho do modelo SARIMA ainda foi considerado adequado dentro de contextos onde a série temporal apresenta estrutura sazonal bem definida. Quando comparado ao presente estudo, observa-se que, mesmo com a maior diversidade regional envolvida na análise da geração hidrelétrica em todo o território nacional, os resultados obtidos foram igualmente expressivos.

Esses indicadores reforçam a robustez da modelagem estatística na presença de componentes sazonais marcantes, ainda que sem o uso de técnicas mais sofisticadas de aprendizado de máquina. Assim como observado no presente estudo, o modelo SARIMA foi considerado eficaz por Silva (2023) na modelagem de séries temporais com padrões sazonais bem definidos.

No entanto, é importante destacar que existem diferenças significativas entre os dois trabalhos, especialmente no que diz respeito ao contexto geográfico e à natureza dos dados analisados. Enquanto o estudo de Silva (2023) se concentrou em estados da região Sul do Brasil, que apresentavam características climáticas e hidrológicas relativamente homogêneas, este trabalho abrange todo o território nacional, marcado por uma diversidade de biomas, regimes pluviométricos e fontes de geração hidrelétrica.

Essa amplitude territorial impõe desafios adicionais ao processo de modelagem, dada a maior complexidade estrutural da série. Ainda assim, os resultados confirmam a capacidade do modelo SARIMA de capturar a dinâmica da geração de energia hidrelétrica mesmo em um contexto mais amplo, consolidando sua aplicabilidade em diferentes escalas de análise.

Já o trabalho de Freitas (2022), intitulado *“Modelagem do consumo de energia elétrica na região Sudeste com séries temporais e redes neurais artificiais”*, teve como foco a previsão do consumo mensal de energia elétrica nos estados da região Sudeste do Brasil, com base em dados coletados entre janeiro de 2002 e dezembro de 2021.

Diferentemente do estudo anterior, que tratou da geração hidrelétrica, este abordou o lado da demanda, o que confere à série características próprias, influenciadas por aspectos demográficos, econômicos e setoriais.

A metodologia adotada envolveu a comparação entre três abordagens: o modelo estatístico SARIMA, uma rede neural LSTM treinada de forma estática (sem reposição) e uma rede LSTM com reposição, em que os dados eram continuamente atualizados durante o processo de previsão. As previsões foram avaliadas a partir das métricas RMSE e MAPE, permitindo verificar o desempenho relativo de cada técnica.

De modo geral, os resultados indicaram que o modelo LSTM com reposição apresentou o

melhor desempenho preditivo na série de consumo da região Sudeste, superando tanto o modelo SARIMA quanto a LSTM sem reposição. O modelo SARIMA, por sua vez, ocupou uma posição intermediária, com desempenho superior ao da LSTM sem reposição, o que reforça sua robustez em contextos com estrutura temporal mais estável.

Um ponto de destaque no estudo foi a constatação de que modelos baseados em inteligência artificial, como as redes neurais, são mais eficazes na presença de padrões não lineares ou rupturas na série histórica, características menos compatíveis com os pressupostos dos modelos estatísticos tradicionais. Ainda assim, o autor reconhece a robustez e a interpretabilidade do SARIMA, sobretudo em contextos com estrutura temporal bem definida e baixa complexidade.

No presente trabalho, que modela a geração de energia hidrelétrica em todo o território nacional, observou-se uma configuração com forte sazonalidade e relativa estabilidade ao longo do tempo, características que favoreceram a adoção do modelo SARIMA.

Assim, embora os modelos de aprendizado de máquina possam superar os modelos clássicos em cenários com maior variabilidade ou comportamento não linear, os resultados obtidos aqui corroboram os achados de Freitas (2022) ao demonstrar que o SARIMA continua sendo uma alternativa eficaz e confiável quando aplicado a séries com padrão estruturado e sazonalidade bem estabelecida.

De maneira geral, os três estudos analisados — o presente trabalho, o de Silva (2023) e o de Freitas (2022) — convergem ao evidenciar a importância da modelagem estatística com o uso do SARIMA em séries temporais do setor elétrico brasileiro.

Cada um, no entanto, aborda realidades distintas: enquanto este estudo concentra-se na geração hidrelétrica em escala nacional, o de Silva (2023) trata da geração regional no Sul do país, e o de Freitas (2022) se dedica ao consumo de energia na região Sudeste.

Apesar dessas diferenças, todos apontam que o modelo SARIMA é eficaz na presença de sazonalidade e tendência bem definidas, sendo especialmente vantajoso pela sua simplicidade, interpretabilidade e bom desempenho preditivo em contextos regulares.

Por outro lado, os trabalhos de Silva (2023) e Freitas (2022) também demonstram que, em situações com maior variabilidade ou complexidade estrutural, modelos mais sofisticados, como os baseados em inteligência artificial, podem superar os métodos tradicionais em termos de acurácia.

Assim, a análise comparativa reforça que a escolha do modelo deve estar alinhada às características específicas da série e ao objetivo do estudo, e que tanto os métodos estatísticos clássicos quanto os algoritmos modernos têm papel relevante e complementar nas estratégias de previsão aplicadas ao setor elétrico.

5. CONCLUSÕES

Os resultados obtidos demonstraram que o modelo SARIMA(2,1,1)(0,1,1)₁₂ foi eficaz na representação da dinâmica da geração de energia hidrelétrica no Brasil, capturando adequadamente tanto a tendência quanto a sazonalidade da série. O modelo apresentou bom desempenho preditivo, com baixos valores de erro, e os diagnósticos residuais indicaram a adequação do ajuste, sem padrões não modelados e com resíduos compatíveis com ruído branco.

As previsões para o ano de 2024 mostraram forte alinhamento com os valores efetivamente observados, especialmente nos meses mais secos, o que reforça a capacidade do modelo em antecipar os padrões de variação da série. Esse desempenho evidencia o potencial da modelagem estatística como ferramenta de apoio à gestão e ao planejamento energético, especialmente em um setor sensível a oscilações climáticas e à necessidade de previsibilidade operacional.

Espera-se que este estudo contribua para a consolidação do uso de técnicas de séries temporais no setor elétrico e que incentive futuras investigações com a incorporação de variáveis climáticas ou abordagens híbridas, incluindo modelos de aprendizado de máquina (*machine learning*), capazes de ampliar a precisão das previsões e auxiliar na tomada de decisões estratégicas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Ahdesmaki, M., Fokianos, K. e Strimmer, K.: *GeneCycle: Identification of Periodically Expressed Genes*, 2025. <https://CRAN.R-project.org/package=GeneCycle>, R package version 1.1.6, Acesso em: abril 2025.
- [2] Akaike, H.: *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6):716–723, 1974.
- [3] Andrade, C. A.: *Análise de séries temporais aplicada ao programa de coleta seletiva no município de Uberlândia (MG)*, 2020. Monografia (Graduação em Estatística) – Universidade Federal de Uberlândia.
- [4] Box, G. E. P. e Jenkins, G. M.: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [5] Box, G. E. P. e Pierce, D. A.: *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*. Journal of the American Statistical Association, 65(332):1509–1526, 1970.
- [6] Caeiro, F. e Mateus, A.: *randtests: Testing Randomness in R*, 2024. <https://CRAN.R-project.org/package=randtests>, R package version 1.0.2, Acesso em: abril 2025.
- [7] Castro, L. N. e Ferrari, D. G.: *Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações*. Editora Saraiva, São Paulo, 2016.
- [8] CNN Brasil: *Entenda tudo o que sabemos sobre o apagão que afetou todas as regiões do Brasil*, 2023. <https://www.cnnbrasil.com.br/nacional/entenda-tudo-o-que-sabemos-sobre-o-apagao-que-afetou-todas-as-regioes-do-brasil/>, Acesso em: abril 2025.
- [9] Cox, D. R. e Stuart, A.: *Some quick tests for the equality of survival distributions*. Biometrika, 42(1/2):80–95, 1955.
- [10] Empresa de Pesquisa Energética (EPE): *Balanço Energético Nacional (BEN) 2023: ano base 2022*. <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-748/topico-687/BEN2023.pdf>, 2023. Acesso em: janeiro 2025.

- [11] ENGIE Brasil: *Tudo o que você precisa saber sobre a crise hídrica de 2021*. <https://www.alemdaenergia.engie.com.br/tudo-o-que-voce-precisa-saber-sobre-a-crise-hidrica/>, 2023. Acesso em: abril 2025.
- [12] Escola de Administração Fazendária (ESAF): *Modelo de previsão para arrecadação tributária*, 2001. https://premios.tesouro.gov.br/1996-2005/2001_T3_P1.pdf, Acesso em: janeiro 2025.
- [13] Folha de S.Paulo: *Amapá tem novo apagão duas semanas após a primeira ocorrência; causas são investigadas*, 2020. <https://www1.folha.uol.com.br/cotidiano/2020/11/amapa-tem-novo-apagao-duas-semanas-apos-a-primeira-ocorrencia-causas-sao-investigadas.shtml>, Acesso em: abril 2025.
- [14] Freitas Silva, M. de: *Modelagem do Consumo de Energia Elétrica na Região Sudeste com Séries Temporais e Redes Neurais Artificiais*. Tese de Mestrado, Uberlândia, 2022.
- [15] G1: *Apagão atinge todas as regiões do país; só 5 estados não foram afetados*, 2018. <https://glo.bo/2FY5rte>, Acesso em: abril 2025.
- [16] Gomes, L. E. S. e Goltara, J. A. P.: *ipeadatar: API Wrapper for 'Ipeadata'*, 2022. <https://CRAN.R-project.org/package=ipeadatar>, R package version 0.1.6, Acesso em: abril 2025.
- [17] Hamner, B., Frasco, M. *et al.*: *Metrics: Evaluation Metrics for Machine Learning*, 2018. <https://CRAN.R-project.org/package=Metrics>, R package version 0.1.4, Acesso em: abril 2025.
- [18] Hyndman, R. J., Athanasopoulos, G. *et al.*: *forecast: Forecasting Functions for Time Series and Linear Models*, 2025. <https://CRAN.R-project.org/package=forecast>, R package version 8.24.0, Acesso em: abril 2025.
- [19] Hyndman, R. J. e Koehler, A. B.: *Another look at measures of forecast accuracy*. International Journal of Forecasting, 22(4):679–688, 2006.
- [20] Instituto de Pesquisa Econômica Aplicada (IPEA): *Ipeadata*. <http://www.ipeadata.gov.br>, Acesso em: abril 2025.
- [21] International Energy Agency: *World Energy Outlook 2020*. IEA, Paris, 2020. <https://www.iea.org/reports/world-energy-outlook-2020>, acessado em abr. 2025.
- [22] Kaytez, F., Taplamacioglu, M. C., Cam, E. e Hardalac, F.: *Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines*. International Journal of Electrical Power & Energy Systems, 67:431–438, 2015.
- [23] Ministério de Minas e Energia: *Lições aprendidas no enfrentamento da situação de escassez hídrica 2020-2021*. <https://www.gov.br/mme/pt-br/assuntos/secretarias/>

- [secretaria-nacional-energia-eletrica/licoes-aprendidas-no-enfrentamento-da-situacao-de-escassez-hidrica-2020-2021](#). Acesso em: abril 2025.
- [24] Morettin, P. A. e Toloi, C. M. d. C.: *Análise de Séries Temporais*. Edgard Blucher, São Paulo, 2006.
- [25] Operador Nacional do Sistema Elétrico (ONS). <http://www.ons.org.br>. Acesso em: abril 2025.
- [26] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. <https://www.R-project.org/>.
- [27] Schwarz, G.: *Estimating the dimension of a model*. The Annals of Statistics, 6(2):461–464, 1978.
- [28] Silva, A. A. da: *Geração de energia hidrelétrica: uma análise comparativa de métodos de previsão estatísticos e de inteligência artificial aplicados a dados da região sul do Brasil*. Trabalho de Conclusão de Curso (Licenciatura em Matemática), Universidade Tecnológica Federal do Paraná, Pato Branco, Brasil, 2023. <http://repositorio.utfpr.edu.br/jspui/handle/1/33931>.
- [29] Silveira, C. A. C. e Guerra, H. N.: *A crise Energética e o monitoramento de reservatórios hidrelétricos*. Em *XIV Simpósio Brasileiro de Recursos Hídricos*, Aracaju, Brasil, 2001.
- [30] Tolmasquim, M. T.: *Novo modelo do setor elétrico brasileiro*. Synergia, 2^a ed., 2011.
- [31] Willmott, C. J.: *On the validation of models*. Physical Geography, 2(2):184–194, 1981.

A. APÊNDICE

Rotina implementada no R para o procedimento de Box-Jenkins, resultando em um modelo do tipo SARIMA:

```
# 1º - Importando a base de dados do Ipeadata
library(ipeadatar)
ipeadatar::search_series(
  terms = 'Energia elétrica - geração - hidráulica', fields = c('name')
)
metadata('ONS12_HIDR12')
dados <- ipeadata(code = "ONS12_HIDR12", language = "br")

# 2º - Separando os dados de análise (time_series)
analysis_data <- dados[dados$date <= as.Date("2023-12-01"),]
datas = analysis_data$date
values <- analysis_data$value

# 3º - Criando a serie temporal de análise
time_series <- ts(analysis_data$value, start = c(1996, 1), frequency = 12)
ts.plot(time_series, col="blue", ylab="Qtde em GWh", xlab="Ano",lwd=2)

# 4º - Verificando a necessidade de transformação
N1 <-length(time_series)
salto <-2
N <-as.integer(N1/salto)
# Calculando a média
media<-vector()
j<-1
vet<-time_series
for(i in 1:N){
  media[i]<-(vet[j]+vet[j+1])/2
  j<-j+2
}
```

```
# Calculando o desvio padrão
desvio <- vector()
z<-1
for(i in 1:N){
  desvio[i]<-sd(c(vet[z],vet[z+1]))
  z<-z+2
}
plot(media,desvio,ylab="Amplitude",xlab="Média")
# Conclusão = não há dependencia entre media e amplitude,
# Logo, é Modelo Aditivo.
# Portanto, não há necessidade de transformação log ou de raiz.

# 5º - FAC e FACP serie de análise (time_series)
acf(time_series, lag.max = 100, main =paste(), col="red", lwd = 2)
pacf(time_series, lag.max = 100)

# 6º - Testando os componentes da série (time_series)

# TENDÊNCIA (Teste de COX-STUART)
# H0: Não existe tendência
# H1: Existe tendência
require(randtests)
cox.stuart.test(time_series)
# Conclusão = Rejeita-se H0, ou seja, tem tendência

# SAZONALIDADE (Teste de FISHER)
# H0: Não existe sazonalidade
# H1: Existe sazonalidade
library(GeneCycle)
fisher.g.test(time_series)
# Conclusão = Rejeita-se H0, ou seja, tem sazonalidade

# 7º - Diferenciando a Série
diff <- diff(time_series)
ts.plot(diff,ylab="1 diff",xlab="Ano",lwd=1, col="blue")

# 8º - Testando os componentes da série diferenciada (diff)
# TENDÊNCIA (Teste de COX-STUART)
cox.stuart.test(diff)
# Conclusão = Não rejeita-se H0, ou seja, tendência removida
```

```
# SAZONALIDADE (Teste de FISHER)
fisher.g.test(diff)
# Conclusão = Rejeita-se H0, ou seja, sazonalidade mantida

# 9º - FAC e FACP serie de diferenciada (diff)
acf(diff, lag.max = 40, main =paste(), col="red", lwd = 2)
pacf(diff, lag.max = 40, main =paste(), col="red", lwd = 2)

# 10º - Ajuste dos modelos
library(forecast)
require(Metrics)
n = length(time_series);n

# Modelo auto.arima - modelagem automatica
mod0 <- auto.arima(time_series, seasonal = T, d = 1, trace = T)
summary(mod0)
t1 <- 0.9609/0.0638
t2 <- -0.1283/0.0774
t3 <- -0.0429/0.0636
t4 <- -0.9531/0.0367
t5 <- -0.7428/0.0430
dt(t1, n); dt(t2, n); dt(t3, n); dt(t4, n); dt(t5, n)
# valor-p (ar1; ar2; ar3; ma1; sma1) respectivamente
# t2 e t3 não significativos

# Modelo 1 - modelo escolhido
mod1 <- Arima(time_series, order=c(2,1,1), seasonal=list(order=c(0,1,1), period=12))
summary(mod1)
t1 <- 0.9737/0.0595
t2 <- -0.1626/0.0587
t3 <- -0.9630/0.0293
t4 <- -0.7333/0.0411
dt(t1, n); dt(t2, n); dt(t3, n); dt(t4, n)
# valor-p (ar1; ar2; ma1; sma1) respectivamente
# todos significativos - ok
acf(mod1$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
pacf(mod1$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
```

```
# + 2 lags sig já nao é ruído branco
# Teste de Box-Pierce para Ruído Branco
Box.test(mod1$residuals, type=c("Box-Pierce")) # Não rejeita H0
cpgram(mod1$residuals)

# Modelo 2
mod2 <- Arima(time_series, order=c(1,1,2), seasonal=list(order=c(0,1,1), period=12))
summary(mod2)
t1 <- 0.7997/0.0554
t2 <- -0.8040/0.0729
t3 <- -0.1616/0.0598
t4 <- -0.7252/0.0405
dt(t1, n); dt(t2, n); dt(t3, n); dt(t4, n)
# valor-p (ar1; ma1; ma2; sma1) respectivamente
# todos significativos - ok
acf(mod2$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
pacf(mod2$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
# + 2 lags sig já nao é ruído branco
# Teste de Box-Pierce para Ruído Branco
Box.test(mod2$residuals, type=c("Box-Pierce")) # Não rejeita H0
cpgram(mod2$residuals)

# Modelo 3
mod3 <- Arima(time_series, order=c(0,1,0), seasonal=list(order=c(2,1,0), period=12))
summary(mod3)
t1 <- -0.5650/0.0516
t2 <- -0.3519/0.0511
dt(t1, n); dt(t2, n)
# valor-p (sar1; sar2) respectivamente
# todos significativos - ok
acf(mod3$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
pacf(mod3$residuals, lag.max = 40, main =paste(), col="red", lwd = 2) # 2 lags
# + 2 lags sig já nao é ruído branco
# Teste de Box-Pierce para Ruído Branco
Box.test(mod3$residuals, type=c("Box-Pierce")) # Não rejeita H0
cpgram(mod3$residuals)

# 11º - Previsão
# Separando os dados para previsão
previsao <- dados[format(dados$date, "%Y") == "2024", ]
```

```
datas = previsao$date
values <- previsao$value

# Mod1 - modelo escolhido
previsao1 <- forecast(mod1, 12, level=c(95))
plot(previsao1, xlab="Ano", ylab="Qtde em GWh")
summary(previsao1)
```