

UNIVERSIDADE FEDERAL DO PARANÁ

WILLIAN FELIPE DREHER QUEIROZ

OTIMIZAÇÃO DA ALOCAÇÃO DE CRÉDITOS NA GERAÇÃO DISTRIBUÍDA
BASEADA EM PREVISÃO DE CONSUMO VIA MODELO SARIMA

CURITIBA

2025

WILLIAN FELIPE DREHER QUEIROZ

OTIMIZAÇÃO DA ALOCAÇÃO DE CRÉDITOS NA GERAÇÃO DISTRIBUÍDA
BASEADA EM PREVISÃO DE CONSUMO VIA MODELO SARIMA

Trabalho de Conclusão de Curso apresentado ao curso de Especialização em Eficiência Energética e Geração Distribuída, Setor de Engenharia Elétrica, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Especialista em Eficiência Energética e Geração Distribuída.

Orientador: Prof. Dr. João da Silva Dias

CURITIBA

2025

TERMO DE APROVAÇÃO

WILLIAN FELIPE DREHER QUEIROZ

OTIMIZAÇÃO DA ALOCAÇÃO DE CRÉDITOS NA GERAÇÃO DISTRIBUÍDA BASEADA EM PREVISÃO DE CONSUMO VIA MODELO SARIMA

Trabalho de Conclusão de Curso apresentado ao curso de Especialização em Eficiência Energética e Geração Distribuída, Setor de Engenharia Elétrica, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Especialista em Eficiência Energética e Geração Distribuída.

Prof. Dr. João da Silva Dias

Orientador(a) – Departamento Engenharia Elétrica, UFPR

Prof. Dr. Rogers Demonti

Departamento Engenharia Elétrica, UFPR

Curitiba, 05 de dezembro de 2025.

RESUMO

A expansão da Geração Distribuída no Brasil e a consolidação do Sistema de Compensação de Energia Elétrica (SCEE) transformaram o setor elétrico, permitindo a compensação de energia entre unidades consumidoras. Porém, esse cenário trouxe desafios na alocação dos créditos de energia, onde a ineficiência causa perdas operacionais e financeiras. O presente trabalho propõe otimizar a gestão desses créditos substituindo métodos empíricos por modelagem estatística. O estudo de caso analisou dados reais de consumo mensal de energia de uma unidade comercial entre 2021 e 2024, utilizando a linguagem R para ajustar o modelo SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), capaz de capturar a tendência e a sazonalidade dos dados. A validação do modelo foi feita por testes estatísticos e indicadores de desempenho, como o MAPE (Erro Percentual Absoluto Médio), que resultou em 8,29%. A fim de confirmar sua aplicação na gestão de créditos de energia, foi realizada uma simulação comparativa de balanço energético entre a metodologia proposta e a prática comum de alocação utilizada no mercado, que é baseada apenas na média dos últimos doze meses. Os resultados demonstraram a superioridade da abordagem estatística, que melhorou o retorno financeiro em 10,23% em relação à metodologia comum. Conclui-se que o uso de métodos estatísticos para previsões de séries temporais é útil para melhorar a eficiência da gestão de ativos da geração distribuída.

Palavras-chave: Geração distribuída; Sistema Compensação de Energia Elétrica (SCEE); modelo SARIMA; previsão de consumo.

ABSTRACT

The expansion of Distributed Generation in Brazil and the consolidation of the Electric Energy Compensation System (SCEE) transformed the electricity sector, allowing energy compensation between consumer units. However, this scenario brought challenges in the allocation of energy credits, where inefficiency causes operational and financial losses. This work proposes to optimize the management of these credits by replacing empirical methods with statistical modeling. The case study analyzed real monthly energy consumption data from a commercial unit between 2021 and 2024, using the R language to fit the SARIMA (Seasonal AutoRegressive Integrated Moving Average) model, capable of capturing the trend and seasonality of the data. The model validation was performed using statistical tests and performance indicators, such as the MAPE (Mean Absolute Percentage Error), which resulted in 8.29%. In order to confirm its application in energy credit management, a comparative simulation of the energy balance was carried out between the proposed methodology and the common allocation practice used in the market, based solely on the average of the last twelve months. The results demonstrated the superiority of the statistical approach, which improved the financial return by 10.23% compared to the common method. It is concluded that the use of statistical methods for time series forecasting is useful to improve the efficiency of distributed generation asset management.

Keywords: Distributed generation; Electric Energy Compensation System (SCEE); SARIMA model; consumption forecasting.

LISTA DE FIGURAS

Figura 1 - Decomposição de uma série temporal.....	19
Figura 2 - Série temporal de consumo de energia elétrica da unidade comercial objeto deste estudo.....	31
Figura 3 - Série temporal dividida entre dados de treino e dados de teste	32
Figura 4 - Decomposição da série temporal de treino.....	33
Figura 5 - Gráficos de autocorrelação dos dados de treino.....	34
Figura 6 - Componente aleatório da série de treino	35
Figura 7 - Ajuste do modelo aos dados de treino	38
Figura 8 - Diagnóstico dos resíduos do modelo	39
Figura 9 - Comparativo entre Valores Previstos e Observados para o Período de Teste	41
Figura 10 - Previsão no Contexto da Série Temporal Histórica.....	42

LISTA DE TABELAS

Tabela 1 – Transformações a serem aplicadas conforme valor de lambda de Box-Cox.....	25
Tabela 2 - Análise descritiva dos dados de treino	34
Tabela 3 - Parâmetros do modelo estimado	36
Tabela 4 - Detalhes estatísticos do modelo.....	37
Tabela 5 - Resultados dos testes de diagnóstico dos resíduos do modelo estimado	39
Tabela 6 - Resultado da previsão para o ano de 2024.....	40
Tabela 7 – Indicadores de precisão do modelo	43
Tabela 8 – Indicadores de precisão das previsões mensais	44
Tabela 9 - Simulação de Demonstrativo de Saldo baseado na previsão por modelo SARIMA	45
Tabela 10 - Simulação de Demonstrativo de Saldo baseado na média dos últimos doze meses.....	46
Tabela 11 - Análise comparativa de desempenho das estratégias de alocação	47

SUMÁRIO

1 INTRODUÇÃO	9
1.1 PROBLEMA	9
1.2 OBJETIVOS	10
1.2.1 Objetivo geral	10
1.2.2 Objetivos específicos.....	10
1.3 JUSTIFICATIVA	10
2 REVISÃO DE LITERATURA	12
2.1 GERAÇÃO DISTRIBUÍDA.....	14
2.1.1 Sistema de Compensação de Energia Elétrica (SCEE)	15
2.1.1.1 Regras de Faturamento	16
2.2 SÉRIES TEMPORAIS	17
2.2.1 Decomposição de Séries Temporais.....	18
2.2.2 Modelagem de séries temporais	19
2.2.2.1 Modelo Autorregressivo (AR)	20
2.2.2.2 Modelo de média móvel (MA).....	20
2.2.2.3 Modelos ARIMA e SARIMA	21
2.2.2.4 Critérios de Seleção de Modelos	22
2.2.3 Testes estatísticos.....	22
2.2.3.1 Ljung-Box	23
2.2.3.2 Jarque-Bera	23
2.2.3.3 ARCH-LM	24
2.2.3.4 Lambda de Box-Cox.....	25
2.2.4 Indicadores de precisão	25
3 MATERIAL E MÉTODOS	27
3.1 OBJETO DE ESTUDO	27
3.2 SOFTWARE UTILIZADO	27
3.3 PREPARAÇÃO E SEGMENTAÇÃO DOS DADOS	28
3.4 ANÁLISE EXPLORATÓRIA DO CONJUNTO DE TREINO	28
3.5 MODELO SARIMA	29
3.6 PREVISÃO DE DADOS FUTUROS	29
3.7 APLICAÇÃO NA COMPENSAÇÃO DE ENERGIA ELÉTRICA	30
4 APRESENTAÇÃO DOS RESULTADOS	31

5 CONSIDERAÇÕES FINAIS	48
5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS	49
REFERÊNCIAS	50

1 INTRODUÇÃO

Ao longo dos últimos anos, a geração distribuída vem se destacando no setor elétrico brasileiro, impulsionada tanto pela evolução tecnológica quanto pelas mudanças regulatórias. A geração distribuída se refere à produção descentralizada de energia elétrica, onde a energia é gerada no próprio local de consumo ou em um local próximo, principalmente a partir de fontes renováveis, como a solar, eólica, biomassa, entre outras.

O mercado de geração distribuída no Brasil vem crescendo desde a REN 482/2012 da ANEEL, que definiu regras para o Sistema de Compensação de Energia Elétrica (SCEE), permitindo a injeção do excedente de energia produzida nas unidades consumidoras na rede de distribuição. Assim, os créditos de energia recebidos podem ser utilizados para abater o valor da energia na fatura, trazendo uma maior economia.

Neste contexto, uma modalidade de compensação de energia elétrica que se tornou comum, é a geração compartilhada, onde a energia é gerada em usinas de investimento que injetam a energia na rede, gerando créditos que serão distribuídos e compensados na fatura de diferentes consumidores, localizadas em endereços distintos dentro da área de concessão da distribuidora.

1.1 PROBLEMA

A gestão de créditos de energia na geração distribuída enfrenta desafios relacionados à eficiência na sua distribuição, devido a fatores como limitações regulatórias, padrões de consumo que incluem tendência e sazonalidade, que podem comprometer o retorno financeiro do investimento. Para maximizar a economia, a quantidade destinada a cada consumidor deve ser a mais próxima possível do seu consumo. Assim, surge o problema de pesquisa: Como otimizar a distribuição de créditos de energia na geração distribuída, considerando padrões de consumo e variações sazonais, de forma a minimizar perdas e maximizar o retorno financeiro do investimento?

Uma possível forma de lidar com esse problema é a aplicação de um método estatístico preditivo baseado na modelagem SARIMA para prever o consumo e permitir maior previsibilidade na alocação de energia.

1.2 OBJETIVOS

Buscando solucionar o problema apresentado, este trabalho estabelece um objetivo geral e seus objetivos específicos que guiam a metodologia e a análise dos resultados.

1.2.1 Objetivo geral

Implementar e testar um modelo preditivo de séries temporais, utilizando a metodologia SARIMA, para a previsão do consumo de uma unidade consumidora, visando a otimização da alocação de créditos de energia no sistema de compensação de energia elétrica da geração distribuída.

1.2.2 Objetivos específicos

- a) Realizar a análise exploratória da série temporal de consumo de energia elétrica;
- b) Estimar o modelo SARIMA mais adequado;
- c) Avaliar a capacidade preditiva do modelo ajustado;
- d) Analisar o impacto das previsões em um cenário de alocação de energia.

1.3 JUSTIFICATIVA

A crescente expansão da geração distribuída no Brasil apresenta novos desafios na gestão de créditos de energia gerados no Sistema de Compensação de Energia Elétrica (SCEE). Devido aos longos prazos regulatórios para alterações cadastrais, há necessidade de prever o consumo a fim de melhorar a eficiência da gestão, tarefa dificultada por padrões de consumo que apresentam tendência e sazonalidade.

A presente pesquisa se justifica por sua relevância prática e econômica, ao desenvolver uma ferramenta analítica para atender a essa demanda. Ao aplicar um modelo preditivo, o trabalho oferece uma solução para a necessidade de previsibilidade do consumo, permitindo otimizar a distribuição de créditos de energia,

reduzindo riscos operacionais e maximizando o retorno financeiro do investimento. Além de beneficiar os consumidores, fortalece o modelo de negócio de empresas que fazem a gestão desses ativos.

No âmbito acadêmico, a justificativa reside na aplicação de uma técnica estatística consolidada, como o modelo SARIMA, para resolução de um problema atual e de alta relevância no setor de geração distribuída, unindo a parte teórica de análise de séries temporais à parte prática de gestão de créditos de energia.

2 REVISÃO DE LITERATURA

A literatura sobre previsão de consumo de energia elétrica apresenta uma evolução, partindo de modelagem estatística clássica e avançando para o uso de técnicas de *machine learning*.

O trabalho de Stenghele et. al (2020) aplica a metodologia clássica de Box-Jenkins. Seu objetivo foi modelar a série temporal do consumo de energia elétrica da região sudeste do Brasil, utilizando a técnica SARIMA. Os dados de estudo, obtidos da Empresa de Pesquisa Energética (EPE), abrangem o período entre 2004 e 2018, totalizando 180 observações mensais. A natureza não-estacionária da série, por apresentar raiz unitária, e sua sazonalidade, foram problemas enfrentados pelos autores, que foram resolvidos com a aplicação de diferenciações regulares e sazonais. Para a análise dos dados, utilizou-se o software GRET (Gnu Regression, Econometrics and Time-series Library), um pacote estatístico de código aberto (*open-source*) especializado em análise econométrica. O modelo SARIMA foi selecionado com base nos indicadores de erro, critério de Akaike (AIC) e no coeficiente U de Theil. A abordagem se mostrou eficaz para previsões de curto prazo, apresentando valores previstos muito próximos dos valores reais observados no período de teste. Com isso, os autores concluíram que a abordagem SARIMA é uma alternativa viável para séries que apresentam sazonalidade e tendência.

Dando continuidade a essa linha de pesquisa, o trabalho de Santos e Chaucoski (2020) teve o objetivo de comparar o modelo SARIMA com uma rede neural *Long Shot-Term Memory* (LSTM), gerando previsões e focando no mesmo conjunto de dados da região sudeste, obtidos da Empresa de Pesquisa Energética (EPE). O período foi expandido, abrangendo desde janeiro de 2004 a maio de 2020, totalizando 197 observações mensais. Os modelos foram implementados em Python, utilizando bibliotecas como Keras e TensorFlow. Os resultados mostraram que tanto o modelo SARIMA quanto a rede neural *Long Shot-Term Memory* (LSTM) forneceram previsões confiáveis e próximas dos valores reais. O estudo concluiu que ambas as abordagens, estatística e de *machine learning*, são viáveis para previsões de curto prazo.

Já o trabalho de Paula, Júnior e Miranda (2020) amplia a técnica estatística clássica, utilizando o modelo SARIMAX, adicionando variáveis externas ao modelo SARIMA. Seu objetivo foi prever as despesas de energia do Tribunal Regional

Eleitoral do Rio Grande do Norte (TRE-RN), para auxiliar na elaboração do orçamento. Os dados cobriam o período de janeiro de 2009 a junho de 2020 e foram obtidos do setor de engenharia da própria instituição de estudo. Como os dados estão registrados em moeda, foi necessária a correção monetária pelo IPCA, para comparar períodos distintos. Além disso, para diminuir a influência de valores atípicos, foi aplicada uma transformação logarítmica de base dez. A abordagem *Walk Forward Validation* foi utilizada, onde o modelo é atualizado quando novos dados da série temporal são gerados. A principal mudança em relação aos outros trabalhos, foi o uso de variáveis externas para modelar eventos atípicos, como o período de mudança de sede do tribunal e o período de pandemia. Foram feitas previsões de um, seis, oito e vinte períodos a frente. O resultado foi uma precisão superior nas previsões comparado ao método de estimativa do órgão, reduzindo o erro absoluto percentual médio (MAPE) em aproximadamente 35,5%. Essa melhoria foi notável principalmente no curto prazo e espera-se que as previsões sejam mais precisas conforme o histórico de dados se expande.

Utilizando outros tipos de algoritmos, o trabalho de Vasconcelos et al (2017) comparou técnicas de regressão *Support Vector Regression* (SVR), redes neurais *perceptron* multicamadas (MLP) e regressão linear, com o objetivo de prever o consumo residencial no Brasil. Os dados anuais, de consumo de energia, PIB e população, abrangem o período de 1980 a 2016 e foram obtidos de fontes públicas como IPEA, IBGE e Banco Central. Avaliou-se o impacto do PIB e da população sobre o consumo de energia. As projeções contemplaram o intervalo de 2014 a 2016. O resultado mostrou que o SVR obteve a melhor precisão, com um erro absoluto percentual médio (MAPE) de 0,0239% para a previsão baseada no histórico de consumo, superando as demais técnicas. Porém, todas as técnicas superaram a abordagem *fuzzy* desenvolvida em um estudo anterior. Os autores afirmaram que nem sempre a inclusão do PIB e população melhoraram a previsão.

O trabalho de Kalimoldayev et al. (2020) teve como objetivo revisar e comparar diversas técnicas preditivas, analisando desde modelos como SARIMA até redes neurais e *Gradient Boosting*, com foco na sua aplicabilidade nos sistemas elétricos da Ucrânia e Cazaquistão. Os dados utilizados são de consumo horário de energia elétrica, de 2012 a 2017, totalizando seis anos de dados horários, obtidos de uma organização de transmissão atacadista dos Estados Unidos. Observando-se a influência de fatores externos, como clima e calendário, no consumo de energia,

foram incorporadas variáveis exógenas para representar a temperatura, a sazonalidade horária e os feriados. Os resultados mostraram que a rede neural *perceptron* multicamadas recorrente (RMLP) teve o melhor resultado, apresentando um Erro Percentual Absoluto Médio (MAPE) de 3,5%, superando as demais abordagens. Os autores destacam a evolução dos métodos preditivos para o *machine learning*, capaz de extrair automaticamente as características dos dados, porém requer maior capacidade computacional.

2.1 GERAÇÃO DISTRIBUÍDA

Historicamente, o setor elétrico foi baseado em um modelo de geração centralizada, onde a energia é gerada por grandes usinas, como hidrelétricas e termoelétricas, localizadas em regiões distantes dos centros de consumo, assim necessitando de longas redes de transmissão e distribuição de energia. A Geração Distribuída (GD) trouxe uma alternativa a esse modelo tradicional de geração, com a geração de energia elétrica no próprio local de consumo ou nas suas proximidades (Moreira, 2019).

Este modelo de geração utiliza fontes renováveis, como solar fotovoltaica, biomassa, eólicas ou centrais hidrelétricas de pequeno porte. Diversos benefícios são obtidos ao se utilizar este modelo de geração, tais como a redução de perdas técnicas no sistema de transmissão e distribuição, a diminuição da necessidade de investimento em longas linhas de transmissão e a diversificação da matriz energética (Agência Nacional de Energia Elétrica, 2021; Moreira, 2019).

No Brasil, o incentivo à Geração Distribuída se iniciou com regulamentações da Agência Nacional de Energia Elétrica (ANEEL), como as Resoluções Normativas 482/2012 e 1000/2021, mas consolidou-se com a promulgação da Lei 14.300, de 6 de janeiro de 2022, que instituiu o Marco Legal para microgeração e minigeração distribuída no país. No ano de 2023, a Resolução Normativa 1.059 aprimorou as regras definidas nas resoluções normativas anteriores (Agência Nacional de Energia Elétrica, 2021; Agência Nacional de Energia Elétrica, 2023; Brasil, 2022).

A legislação define microgeração como uma central geradora com potência instalada de até 75 kW (quilowatts) e minigeração como central geradora com potência superior a 75 kW e inferior a 3 MW (megawatts), conectadas à rede de

distribuição, que utilizam de fontes renováveis para geração de energia elétrica (BRASIL, 2022).

2.1.1 Sistema de Compensação de Energia Elétrica (SCEE)

O Sistema de Compensação de Energia Elétrica (SCEE) foi estabelecido pelas Resoluções Normativas 482/2012 e 1000/2021 da ANEEL, e consolidado pela Lei 14.300 de 2022. Este sistema permite que a energia excedente gerada pela unidade consumidora seja injetada na rede da distribuidora local e convertida em créditos de energia. Estes créditos de energia podem ser utilizados para abater o consumo da própria unidade consumidora que gerou a energia ou de outras unidades consumidoras localizadas na mesma região de concessão da distribuidora (Brasil, 2022; Moreira, 2019).

A legislação estabelece modalidades que permitem a compensação de créditos em unidades consumidoras distintas da unidade geradora, desde que situadas na mesma área de concessão. Dentre as modalidades, destacam-se o autoconsumo remoto, o empreendimento com múltiplas unidades consumidoras (EMUC) e a geração compartilhada (Brasil, 2022).

A modalidade de autoconsumo remoto permite a compensação de créditos de energia em unidades consumidoras sob a mesma titularidade da unidade geradora. Já a modalidade de empreendimento com múltiplas unidades consumidoras (EMUC) permite a compensação de energia em um conjunto de unidades consumidoras localizadas em uma mesma propriedade ou propriedades próximas, como um prédio ou condomínio, em que a unidade consumidora que engloba as instalações das áreas de uso comum, se conecta à microgeração ou minigeração distribuída. A geração compartilhada é a modalidade de compensação que permite a distribuição de créditos de energia entre consumidores participantes de consórcio, cooperativa ou associações civis (Brasil, 2022).

Dentre essas modalidades, a geração compartilhada é a mais utilizada por empresas que fazem a gestão de usinas que atendem consumidores interessados no benefício de compensação de energia elétrica gerada a partir de fontes renováveis. Sua vantagem em relação às demais modalidades é a possibilidade de enviar excedente de energia para unidades consumidoras localizadas em propriedades distantes, com titular diferente do titular da unidade geradora, apenas

necessitando da inscrição do titular em um consórcio, cooperativa ou associação que seja titular da unidade geradora.

2.1.1.1 Regras de Faturamento

O faturamento de unidades consumidoras integrantes do Sistema de Compensação de Energia Elétrica (SCEE) possui regras específicas definidas na regulamentação de geração distribuída. A Lei 14.300, de 2022, estabelece que o faturamento da unidade consumidora deve considerar a energia elétrica ativa consumida, deduzindo a energia excedente alocada nessa unidade, bem como eventuais créditos de energia acumulados em ciclos de faturamentos anteriores (Brasil, 2022).

Porém, a Lei 14.300 também impõe um limite mínimo para este faturamento, pois determina que a compensação dos créditos de energia não pode incidir sobre o valor mínimo faturável, denominado como Custo de Disponibilidade (Brasil, 2022). Conforme a Resolução Normativa nº 1.000, de 2021, o custo de disponibilidade é o montante mínimo de energia em quilowatts-hora (kWh) que deve ser cobrado do consumidor para cobrir os custos de disponibilidade da rede de distribuição da distribuidora. Esse montante mínimo varia conforme o tipo de fornecimento, sendo 30 kWh para circuitos monofásico ou bifásico a dois condutores, 50 kWh se bifásico a três condutores e 100 kWh se trifásico (Agência Nacional de Energia Elétrica, 2021).

Dessa forma, a regulamentação define um limite para a compensação de energia elétrica. A quantidade de energia que pode ser efetivamente compensada em um ciclo de faturamento é limitada ao montante total da energia elétrica consumida, subtraído o montante do custo de disponibilidade.

O excedente de energia que não é compensado dentro do ciclo de faturamento não é perdido e pode ser utilizado nos próximos ciclos de faturamento. A legislação define que o crédito de energia elétrica é o montante de energia que não foi compensado no ciclo em que foi gerado e fica disponível para alocação em faturamentos subsequentes. Estes créditos possuem um prazo de validade de 60 (sessenta) meses após a data do faturamento em que forma gerados (Agência Nacional de Energia Elétrica, 2021; Brasil, 2022).

Com base nessas regulamentações, o faturamento segue uma lógica que define a energia efetivamente compensada baseada em um cálculo condicional. Antes de detalhar essa lógica, é necessário definir alguns termos. O total de créditos disponíveis no ciclo de faturamento é igual à energia injetada (ou recebida de outra unidade consumidora) somada ao saldo de créditos de energia acumulado em ciclos de faturamentos anteriores. Já o limite de compensação equivale à quantidade total de energia consumida menos o custo de disponibilidade daquela unidade consumidora. Dessa forma, a quantidade de energia a ser compensada será sempre o menor valor entre o total de créditos disponíveis e o limite de compensação daquele ciclo de faturamento. Todo crédito disponível que exceder o limite de compensação não é utilizado naquele ciclo de faturamento e é armazenado como saldo acumulado, que fica disponível para compensação em ciclos de faturamento subsequentes.

2.2 SÉRIES TEMPORAIS

Primeiramente, é necessário definir o que é uma série temporal. Segundo Morettin e Toloí (2018), uma série temporal pode ser definida como qualquer conjunto de dados ordenados ao longo do tempo, como valores diários de visitantes, valores mensais de vendas, valores anuais de lucro, registro de acidentes de trânsito, entre outros. As séries temporais podem ser discretas, quando as observações são feitas em instantes de tempo específicos e igualmente espaçados, como no caso do consumo de energia, que é registrado mensalmente. Mas também existem as séries temporais contínuas, nas quais as observações são feitas sem interrupções, como o registro da temperatura ao longo de um dia. Porém, a fim de permitir uma análise estatística, as séries contínuas são usualmente transformadas em valores discretos, como calcular a média da temperatura ao longo do dia ou a cada hora.

Conforme Morettin e Toloí (2018), os principais objetivos da análise de séries temporais são: investigar o fenômeno gerador, descrever as suas características, prever valores futuros e, em determinados contextos, utilizar os modelos estimados para controlar processos.

Um conceito importante na análise de séries temporais é a estacionariedade. Uma série é considerada estacionária quando sua média e variância se mantêm

constantes ao longo tempo, fazendo com que ela flutue em torno de um valor constante (Morettin; Tolo, 2018).

Segundo Gujarati e Porter (2011), o estudo do comportamento de uma série temporal não estacionária se torna restrito apenas ao período em que os dados foram coletados, sendo impossível generalizar o comportamento da série. Conseqüentemente, modelos baseados em séries não estacionárias são ineficientes para previsão, já que assumem erroneamente que as propriedades estatísticas do passado serão repetidas no futuro. Porém, a maior parte das séries temporais encontradas no mundo real apresentam algum tipo de não estacionariedade. Para contornar a não estacionariedade, são necessárias transformações nos dados para tornar a série estacionária, como diferenciação ou transformação logarítmica (Morettin; Tolo, 2018).

2.2.1 Decomposição de Séries Temporais

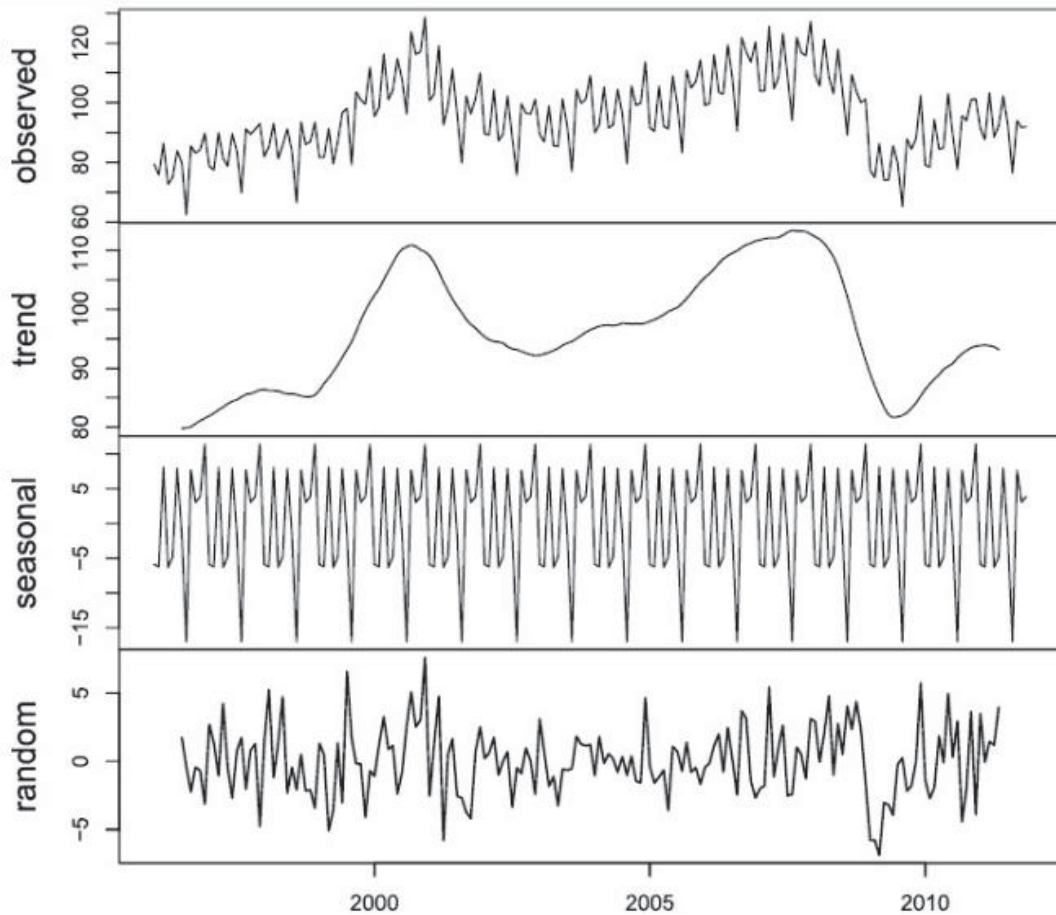
Segundo Ferreira (2017), uma série temporal é composta de diversos padrões que podem ser destacados ao se decompor a série temporal, a fim de isolar e analisar cada padrão de forma separada. A ideia é analisar cada componente separadamente para entender o comportamento da série. Os principais componentes a serem destacados são:

- a) Tendência: representa a direção que a série segue a longo prazo, podendo ser crescente ou decrescente, e não necessariamente linear;
- b) Sazonalidade: é o padrão que se repete em um período de tempo conhecido, como uma semana, um semestre ou um ano;
- c) Ciclo: semelhante à sazonalidade, são padrões que se repetem, mas sem período de tempo fixo, e geralmente são mais longos, durando mais de um ano, como fenômenos climáticos que duram anos;
- d) Componente aleatório (ou resíduo): são os movimentos da série que não podem ser explicados pelos outros componentes. É o resíduo obtido após a remoção dos demais componentes.

Pode-se observar na Figura 1, os diversos componentes presentes em uma série temporal. O gráfico no topo ilustra a série temporal completa, logo abaixo pode-se visualizar a sua tendência de crescimento e decrescimento. Já no terceiro

gráfico, notam-se os padrões sazonais existentes na série, e por último tem-se os resíduos que não podem ser descritos pelos demais componentes.

Figura 1 - Decomposição de uma série temporal



Fonte: Ferreira (2017).

2.2.2 Modelagem de séries temporais

Após identificar os padrões que compõem uma série temporal, através da análise exploratória, constrói-se o modelo estatístico que irá representar matematicamente o comportamento da série temporal. Para séries estacionárias, os modelos autorregressivos de média móvel (ARMA) são comumente utilizados. Já para séries não estacionárias, os modelos autorregressivos integrados de médias móveis (ARIMA) são mais recomendados, devido a integrarem a tendência e a sazonalidade ao modelo (Bueno, 2008; Ferreira, 2017; Morettin; Tolo, 2018). A seguir, serão detalhados os principais componentes que compõem esses modelos.

2.2.2.1 Modelo Autorregressivo (AR)

De acordo com Morettin e Tolo (2018), o modelo autorregressivo, denotado por $AR(p)$, é baseado no pressuposto de que o valor atual de uma série temporal pode ser explicado como uma função linear dos valores anteriores da série. Ou seja, a observação em um instante t é uma regressão das observações anteriores $t-1$, $t-2$, até $t-p$. A equação (1) descreve o processo autorregressivo.

$$Z_t = \theta_0 + \Phi_1 Z_{t-1} + \dots + \Phi_p Z_{t-p} + a_t \quad (1)$$

Onde, Z_t é o valor da série temporal no instante de tempo atual (t), θ_0 é um termo constante, relacionado à média da série, Φ_p são os coeficientes do modelo autorregressivo, Z_{t-p} são os valores anteriores da própria série temporal e a_t é o termo aleatório no instante t . A letra p representa ordem do modelo, indicando quantos valores anteriores serão utilizados para prever o valor atual.

2.2.2.2 Modelo de média móvel (MA)

Conforme descrito por Morettin e Tolo (2018), o modelo de médias móveis, denotado como $MA(q)$, diferente do modelo autorregressivo, não utiliza os valores anteriores da série temporal para explicar o valor atual, em vez disso, modela o valor atual como uma combinação linear dos erros de previsão de períodos anteriores. Este erro de previsão representa a parte do valor da série que não foi prevista pelo modelo nos valores anteriores. A equação (2) descreve o processo de médias móveis.

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2)$$

Onde, Z_t é o valor atual da série temporal, μ é a média constante do processo, a_t é o termo aleatório do período atual, a_{t-1} e a_{t-p} são os valores aleatórios de períodos anteriores, θ_1 e θ_p são os coeficientes do modelo de médias móveis. A letra q representa a ordem do modelo, indicando quantos erros de períodos passados serão incluídos no modelo.

2.2.2.3 Modelos ARIMA e SARIMA

Os modelos autorregressivo e de médias móveis são comumente utilizados para compor modelos mais complexos, capazes de lidar com diversas séries temporais encontradas na prática. Segundo Morettin e Toloi (2018), o Modelo Autorregressivo Integrado de Médias Móveis (ARIMA), denotado por $ARIMA(p, d, q)$, une os modelos autorregressivo e de médias móveis a um processo de diferenciação, permitindo a modelagem de séries não estacionárias. As letras p , d e q , presentes na sua notação, representam a ordem de cada um desses componentes que atuam juntos. A parte autorregressiva, $AR(p)$, se baseia na relação entre o valor atual e um número p de valores anteriores da série temporal, já a parte de médias móveis, $MA(q)$, representa a dependência entre o valor atual e um número q de erros anteriores de previsão. Completando o modelo, a parte de integração, $I(d)$, representa a quantidade d de diferenciações necessárias para tornar a série estacionária.

Dessa forma, a modelagem ARIMA permite a modelagem de séries temporais compostas por tendências, pois aplica a diferenciação (d) para remover a tendência, tornando a série estacionária e permitindo a aplicação da modelagem autorregressiva (AR) e de médias móveis (MA).

Nos casos em que a série é composta tanto por tendência quanto por sazonalidade, é necessário o uso de uma extensão do modelo ARIMA para realizar a correta modelagem da série. O Modelo Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA), denotado por $SARIMA(p, d, q)(P, D, Q)_s$ é capaz de modelar dependências não sazonais e sazonais simultaneamente. Os parâmetros P , D e Q incluídos neste modelo são os equivalentes sazonais dos parâmetros não sazonais, presentes no modelo ARIMA. Assim, P representa a ordem da parte autorregressiva sazonal, D representa a ordem das diferenciações sazonais e Q representa a parte de médias móveis sazonais. O termo s refere-se ao período da sazonalidade, por exemplo, para valores mensais, considerando o período anual, ele assume o valor doze. Esta classe de modelos, proposta originalmente por Box e Jenkins, se tornou uma das metodologias mais utilizadas para análise e previsão de séries temporais, sendo aplicadas em diversas áreas de estudo (Ferreira, 2017; Morettin; Toloi, 2018).

2.2.2.4 Critérios de Seleção de Modelos

A seleção de modelos é fundamental na análise de séries temporais, seu objetivo é identificar qual modelo apresenta o melhor equilíbrio entre o melhor ajuste aos dados e a complexidade do modelo. Com este fim, utilizam-se critérios de informação que penalizam a complexidade do modelo, como o Critério de Informação de Akaike Corrigido (AICc) e o Critério de Informação Bayesiano (BIC). Baseando a decisão nesses critérios, o melhor modelo será aquele em que o critério apresenta o menor valor numérico (Burnham; Anderson, 2002).

O Critério de Informação de Akaike (AIC) é um indicador padrão que estima a perda de informação devido ao uso de um modelo para representar a realidade. Porém, o AIC tradicional tende a selecionar modelos com excesso de parâmetros quando a amostra é pequena em relação ao número de parâmetros estimados. Para corrigir esse problema, utiliza-se o Critério de Informação de Akaike Corrigido (AICc) que incorpora um termo de penalidade adicional específico para casos de amostras finitas. A utilização do AICc é fortemente recomendada quando a razão entre a quantidade de observações e o número de parâmetros é baixa, garantindo uma seleção mais confiável do modelo (Burnham; Anderson, 2002).

O Critério de Informação Bayesiano (BIC) aborda a seleção de modelos sob uma outra perspectiva, priorizando a consistência estatística. Ao contrário do AICc, o BIC penaliza mais a complexidade conforme o tamanho da amostra cresce, favorecendo modelos com menos parâmetros. Embora evite o *overfitting*, pode levar à escolha de modelos que falham em capturar padrões relevantes dos dados, o que caracteriza um subajuste (Burnham; Anderson, 2002).

Portanto, a escolha adequada entre esses critérios é vital para atender ao princípio da parcimônia, buscando um modelo que seja simples o suficiente para ser interpretável, mas robusto o bastante para generalizar bem os dados futuros.

2.2.3 Testes estatísticos

Para complementar a análise visual da série temporal ou de seus resíduos, são aplicados testes estatísticos para validar as conclusões obtidas da análise visual. Alguns desses testes serão apresentados a seguir.

2.2.3.1 Ljung-Box

Segundo Bueno (2008), o teste de Ljung-Box busca avaliar a correlação entre os valores da série temporal. Em análise de séries temporais, é comumente aplicado aos resíduos de um modelo estimado para avaliar se eles se comportam como um ruído branco. Este teste avalia um conjunto de autocorrelações até uma determinada defasagem e é baseado nas seguintes hipóteses:

- a) Hipótese nula (H_0): As autocorrelações até a defasagem m são conjuntamente iguais a zero. Isso significa que não há autocorrelação nos resíduos da série e se comportam como um ruído branco;
- b) Hipótese alternativa (H_1): Pelo menos uma das autocorrelações até a defasagem m é diferente de zero, o que indica a presença de autocorrelação e que o modelo pode não ter capturado todo o comportamento da série.

A estatística de teste, conhecida como estatística Q , é calculada e comparada com uma distribuição qui-quadrado (χ^2). Para tomada de decisão, o valor de Q é convertido em uma probabilidade, conhecida como valor-p (ou *p-value*). O valor-p representa a probabilidade de se observar uma estatística Q mais extrema do que a calculada, assumindo que a hipótese nula é verdadeira (Bueno, 2008; Ferreira, 2017).

A decisão é baseada no comparativo do valor-p a um nível de significância (α) pré-estabelecido (normalmente definido como 0,05). Se o valor-p foi maior ou igual ao nível de significância, não há evidências para rejeitar a hipótese nula, e conclui-se que os resíduos se comportam como ruído branco. Porém, se o valor-p for menor que o nível de significância, rejeita-se a hipótese nula, indicando a presença de autocorrelação significativa nos resíduos da série temporal (Bueno, 2008; Ferreira, 2017).

2.2.3.2 Jarque-Bera

Segundo Bueno (2008), o teste de Jarque-Bera é um teste de aderência utilizado para verificar se os resíduos de um modelo seguem uma distribuição normal. A normalidade dos resíduos comprova a validade estatística de modelos,

pois garante que os testes de hipóteses sobre os coeficientes são confiáveis. Este teste verifica o quão distantes os valores de assimetria e curtose dos resíduos estão dos valores teóricos de normalidade. As seguintes hipóteses são definidas neste teste:

- a) Hipótese nula (H_0): os resíduos seguem uma distribuição normal. Para uma distribuição normal, espera-se que a assimetria seja igual a zero e a curtose igual a 3;
- b) Hipótese alternativa (H_1): os resíduos não seguem uma distribuição normal.

A estatística de teste, definida como JB , combina as medidas de assimetria e curtose em um único valor. Para tomada de decisão, esta estatística de teste é comparada aos valores de uma distribuição qui-quadrado (χ^2). Caso o valor-p obtido seja maior que o nível de significância, não há evidências para rejeitar a hipótese nula, indicando que os resíduos são normalmente distribuídos. No entanto, se o valor-p for menor que o nível de significância, representa que os resíduos não seguem uma distribuição normal. Nesse caso, o modelo pode apresentar problema de especificação (Bueno, 2008; Ferreira, 2017).

2.2.3.3 ARCH-LM

De acordo com Bueno (2008), o teste ARCH-LM (Multiplicador de Lagrange para Heterocedasticidade Condicional Autorregressiva) é utilizado para verificar se a variância dos resíduos não é constante ao longo do tempo. Se baseia na verificação da relação entre a variância dos resíduos de um período atual com os quadrados dos resíduos de períodos anteriores. As hipóteses definidas neste teste são as seguintes:

- a) Hipótese Nula (H_0): Não há efeitos ARCH, então os resíduos são homocedásticos (variância constante);
- b) Hipótese Alternativa (H_1): Existe pelo menos um efeito ARCH, que indica que a variância dos resíduos é heterocedástica e depende de resíduos passados.

A decisão é tomada a partir do valor-p associado à estatística de teste LM. Se o valor-p for maior que o nível de significância, não há evidências para rejeitar a

hipótese nula, indicando que os resíduos são homocedásticos, ou seja, tem variância constante. Mas se o valor-p for menor que o nível de significância, a hipótese nula é rejeitada, indicando que a variância nos resíduos não é constante.

2.2.3.4 Lambda de Box-Cox

Segundo Morettin e Tolo (2018), em alguns casos, quando a série temporal apresenta heterocedasticidade, ou seja, a variância não é constante, as previsões obtidas a partir de um modelo ARIMA podem ser prejudicadas. Para corrigir isso, há necessidade de aplicar uma transformação nos dados. A transformação de Box-Cox resolve esse problema, estimando um parâmetro chamado de lambda (λ) que indica qual a transformação de potência mais adequada para a série. Embora o valor de lambda possa ser usado para uma transformação direta, a interpretação dos resultados se torna mais intuitiva quando próxima a valores conhecidos. A Tabela 1 indica qual transformação deve ser aplicada conforme o valor de lambda.

Tabela 1 – Transformações a serem aplicadas conforme valor de lambda de Box-Cox

<i>Valor de lambda (λ)</i>	<i>Interpretação</i>	<i>Transformação a ser aplicada</i>
Próximo de 1	Variância da série já é considerada estável	Nenhuma transformação é necessária
Próximo de 0	A variância não é estável	Transformação logarítmica
Próximo de 0,5	A variância cresce junto com o nível médio dos dados	Transformação de Raiz Quadrada

Fonte: O autor (2025)

2.2.4 Indicadores de precisão

Para avaliar o desempenho do modelo e compará-lo a outros métodos, utilizam-se indicadores de precisão. O Erro Médio (ou *Mean Error* – ME) representa a média da diferença entre os valores reais e os valores previstos. Sua finalidade não é medir a magnitude do erro, mas sim a sua direção. Um resultado com valor positivo indica que o modelo subestimou os valores reais, enquanto um valor negativo aponta que o modelo superestimou a série (Silva, 2023).

O Erro Quadrático Médio (ou *Mean Squared Error* – MSE) é calculado como a média da diferença elevada ao quadrado entre o valor real e o valor previsto. Como os valores são elevados ao quadrado, são mais difíceis de interpretar. Essa métrica sempre terá um valor positivo e quanto mais próxima de zero, indica melhor precisão. Já a Raiz do Erro Quadrático Médio (ou *Root Mean Squared Error* - RMSE), é calculada aplicando a raiz quadrada no MSE, a fim de retornar a métrica para a mesma unidade dos dados originais, facilitando a compreensão da magnitude do erro (Spadini; Alencar, 2022).

O Erro Absoluto Médio (ou *Mean Absolute Error* – MAE) é calculado pela média da diferença absoluta entre o valor real e o valor previsto. O resultado é expresso na unidade original da série, sendo de fácil interpretação. O Erro Percentual Absoluto Médio (ou *Mean Absolute Percent Error* – MAPE) é obtido pelo cálculo da média dos erros absolutos divididos pelos valores reais da série. A principal vantagem dessa métrica é que, por ser uma medida percentual, ela permite comparar a precisão de previsões entre séries temporais de diferentes grandezas. De forma similar, o Erro Percentual Médio (ou *Mean Percentage Error* – MPE) calcula a média dos erros percentuais sem usar o valor absoluto. Enquanto o MAPE mede a magnitude do erro em porcentagem, o MPE indica se o modelo tende a superestimar ou subestimar os dados (Hyndman; Athanasopoulos, 2021; Spadini; Alencar, 2022).

O Erro Médio Absoluto Escalado (ou *Mean Absolute Scaled Error* – MASE) compara o erro absoluto médio do modelo com o erro de um método ingênuo na série de treino. Um valor menor do que um indica que o modelo proposto tem desempenho superior à previsão ingênua (Hyndman; Athanasopoulos, 2021).

O coeficiente U de Theil busca comparar o desempenho do modelo estimado com um método de previsão ingênuo, que utiliza o valor anterior observado para prever o período seguinte. Um resultado inferior a um indica que o modelo é superior ao método ingênuo, e um valor igual ou superior a um indica que a previsão ingênua tem desempenho igual ou melhor ao método proposto (Cook, 2019).

Já a Autocorrelação de Lag 1 (ACF1) mede a correlação linear entre os erros consecutivos de previsão. Em um modelo ideal, os resíduos não devem ter autocorrelação significativa. Um valor de ACF1 próximo de zero indica que o modelo capturou adequadamente as características da série temporal (Hyndman; Athanasopoulos, 2021).

3 MATERIAL E MÉTODOS

Esta pesquisa caracteriza-se como um estudo de caso quantitativo. A abordagem inicia-se com a apresentação dos dados utilizados no estudo e a forma como foram coletados. Em seguida, são detalhados os processos de análise e modelagem da série temporal, previsão de dados futuros, e da aplicação em uma situação típica de gestão de compensação de energia elétrica. Os indicadores de precisão utilizados na avaliação dos resultados também são especificados.

3.1 OBJETO DE ESTUDO

Os dados de estudo são registros do consumo mensal de energia elétrica, em quilowatt-hora (kWh), de um estabelecimento comercial do ramo alimentício, no período entre janeiro de 2021 e dezembro de 2024. As informações foram extraídas de um banco de dados de uma empresa especializada em gestão de energia na geração distribuída. Tais registros são provenientes dos históricos de consumo apresentados nas faturas de energia elétrica emitidas pela concessionária local, a Companhia Paranaense de Energia (Copel). Este caso específico foi escolhido para o estudo pois apresenta duas características bastante comuns em consumidores de baixa tensão, tendência e sazonalidade, em conjunto. Assim deve ser capaz de representar bem a capacidade de aplicação da modelagem nos demais consumidores.

3.2 SOFTWARE UTILIZADO

Toda a análise, modelagem e previsão dos dados de consumo foram implementadas utilizando a linguagem R, em sua versão 4.5.1. Como ambiente de desenvolvimento integrado (IDE), foi utilizado o Visual Studio Code (VSCode). A escolha por essa linguagem justifica-se por sua natureza de código aberto (*open source*), gratuidade e por ser uma ferramenta amplamente difundida no meio acadêmico.

Para a etapa de simulação da compensação de energia elétrica, os cálculos, montagem dos demonstrativos e tabelas comparativas foram realizados em uma planilha na plataforma Google Sheets.

3.3 PREPARAÇÃO E SEGMENTAÇÃO DOS DADOS

O primeiro passo da análise consistiu em salvar os dados em um arquivo de texto (.txt), contendo os dados de consumo em uma única coluna. Esses dados foram importados para o ambiente de programação e convertidos em uma série temporal de frequência mensal, com início em janeiro de 2021.

Com o objetivo de estimar um modelo preditivo e avaliar a sua precisão, a série temporal foi dividida em dois subconjuntos. O conjunto de treino abrangeu os dados entre janeiro de 2021 e dezembro de 2023, totalizando três anos de dados, que foram utilizados para a análise exploratória e ajuste do modelo SARIMA. O conjunto de teste, formado pelos dados referentes ao ano de 2024, serviu para validação do desempenho do modelo. A escolha de utilizar um ano completo como os dados de teste visou contemplar todo o período onde a sazonalidade atua, prevenindo interpretações incorretas dos resultados, que poderiam ocorrer caso a análise fosse restrita a um período menor. A base dos dados de treino deve representar a realidade e o modelo precisa ser capaz de capturar as suas características.

3.4 ANÁLISE EXPLORATÓRIA DO CONJUNTO DE TREINO

A análise exploratória dos dados de treino teve início na decomposição da série temporal em componentes de tendência, sazonalidade e aleatoriedade. Em seguida foi feita a análise descritiva dos dados, obtendo valores estatísticos como média, mediana, e os valores de máximo e mínimo. A fim de investigar a dependência temporal da série, foram gerados gráficos de autocorrelação e autocorrelação parcial para análise visual.

Com o intuito de determinar a necessidade de transformação dos dados para estabilizar a variância antes da modelagem, analisou-se a heterocedasticidade dos dados. Para isso, realizou-se a análise visual do componente aleatório da decomposição e, para confirmação estatística, a estimação do parâmetro da transformação de Box-Cox.

3.5 MODELO SARIMA

Os parâmetros do modelo SARIMA foram estimados automaticamente através da função “auto.arima” inclusa no pacote “forecast” da linguagem R. Esta função foi configurada para realizar uma busca completa do melhor modelo, sem atalhos, sem nenhuma transformação de Box-Cox e com a definição do melhor modelo baseado no menor valor de AICc (Critério de Informação de Akaike Corrigido).

Após o ajuste do modelo, realizou-se uma análise para verificar a sua aderência aos dados de treino. Esta análise foi aplicada aos resíduos do modelo, buscando verificar se eles se comportam como um ruído branco. O procedimento combinou a inspeção visual, por meio dos gráficos dos resíduos, autocorrelação e histograma de normalidade, com os testes estatísticos de Ljung-Box, Jarque-Bera e ARCH-LM. Para estes testes, a decisão foi baseada na comparação do valor-p com um nível de significância predefinido no valor de 0,05. Um valor-p superior ao nível de significância representa a não rejeição da hipótese nula, indicando que os resíduos atendem à premissa testada.

3.6 PREVISÃO DE DADOS FUTUROS

Após validar que os resíduos do modelo se comportam como ruído branco, realizou-se a previsão de consumo para os 12 meses à frente, correspondendo ao período de janeiro a dezembro de 2024. As projeções foram geradas utilizando-se da função “forecast”, pertencente à biblioteca “forecast” da linguagem R, aplicada ao modelo SARIMA estimado.

Com a previsão de consumo para 2024 finalizada, realizou-se a comparação entre os valores projetados e os dados reais do conjunto de teste. Neste contexto, o desempenho do modelo foi quantificado por meio do seguinte conjunto de indicadores de erro estatístico: Erro Médio (ME), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE), Erro Percentual Médio (MPE), Erro Percentual Absoluto Médio (MAPE), Erro Médio Absoluto Escalado (MASE), Autocorrelação dos erros no Lag1 (ACF1) e U de Theil. Além desses indicadores de análise geral de precisão, foram calculados o Erro Absoluto e o Erro Percentual para cada mês de consumo previsto.

3.7 APLICAÇÃO NA COMPENSAÇÃO DE ENERGIA ELÉTRICA

Todo esse processo de previsão de consumo de energia foi desenvolvido devido à necessidade de previsão de consumo no cadastro de distribuição de saldo excedente de usinas de micro ou minigeração, pertencentes à geração distribuída. Essa necessidade de previsão existe pela burocracia no cadastro da lista de distribuição da usina, onde a distribuidora tem até 30 dias para atender à solicitação de alteração do cadastro de lista de distribuição e deve aplicar até o próximo ciclo de faturamento, baseado na data de solicitação.

O método mais comum no mercado de gestão de saldo em geração distribuída é distribuir o saldo baseando-se na média de consumo dos últimos doze meses da unidade consumidora beneficiária. Porém, devido à sazonalidade e tendência, comuns nas séries temporais de consumo, esse método pode não ser muito eficiente, resultando na insuficiência de saldo para compensação total do consumo ou acúmulo de créditos que poderiam ser aproveitados em outros consumidores.

Dessa forma, simulou-se a alocação ideal do consumidor em uma usina, para os doze meses de 2024, distribuindo o saldo baseando-se no consumo previsto pela metodologia descrita acima. Além dessa simulação com o consumo previsto, também foi desenvolvida uma simulação baseada na média de consumo dos doze meses anteriores, ou seja, do ano de 2023, com o objetivo de comparar com a distribuição baseada em previsão com o método SARIMA.

As simulações foram conduzidas considerando que a energia injetada na unidade consumidora beneficiária fosse equivalente ao consumo previsto para aquele mês ou à média dos últimos doze meses. Ao confrontar a injeção simulada com o consumo real, obteve-se o valor de energia compensada em cada mês, considerando o limite do consumo mínimo cobrado pela concessionária. A partir desses valores, foi apurada a quantidade de saldo faltante ou excedente em cada mês. Estes cálculos foram baseados nas regras de faturamento definidas na regulamentação da geração distribuída.

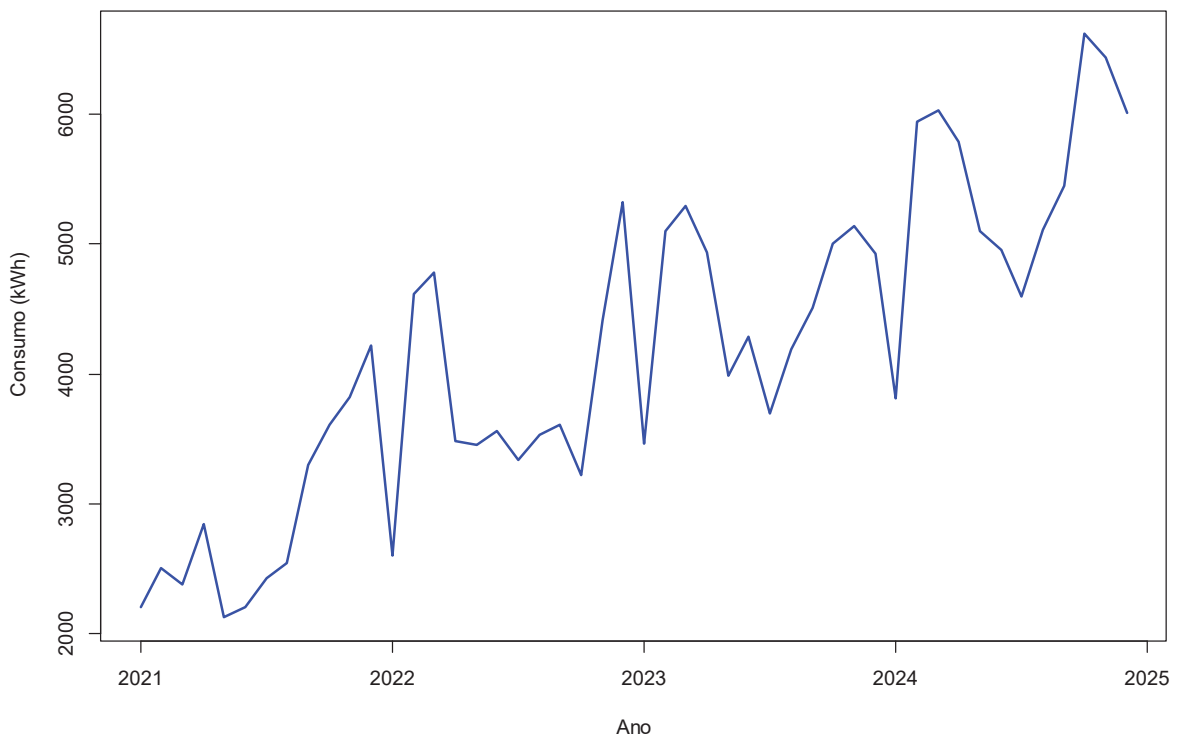
Concluídas as simulações, foram comparados os totais de energia injetada, energia compensada, saldo restante e faltante, bem como o valor total recebido. Para isso, calcularam-se a diferença absoluta e o percentual de variação entre os métodos.

4 APRESENTAÇÃO DOS RESULTADOS

Nesta seção, serão apresentados os resultados obtidos a partir dos procedimentos que foram executados e descritos na seção de metodologia, assim como a interpretação dos mesmos e a avaliação do desempenho do método preditivo e da distribuição de saldo excedente baseada nele.

A etapa de modelagem de uma série temporal se inicia pela análise visual do comportamento histórico dos dados. A Figura 2 apresenta a série temporal referente ao consumo mensal da unidade comercial objeto deste estudo, conforme descrito na metodologia, abrangendo o período entre janeiro de 2021 e dezembro de 2024.

Figura 2 - Série temporal de consumo de energia elétrica da unidade comercial objeto deste estudo

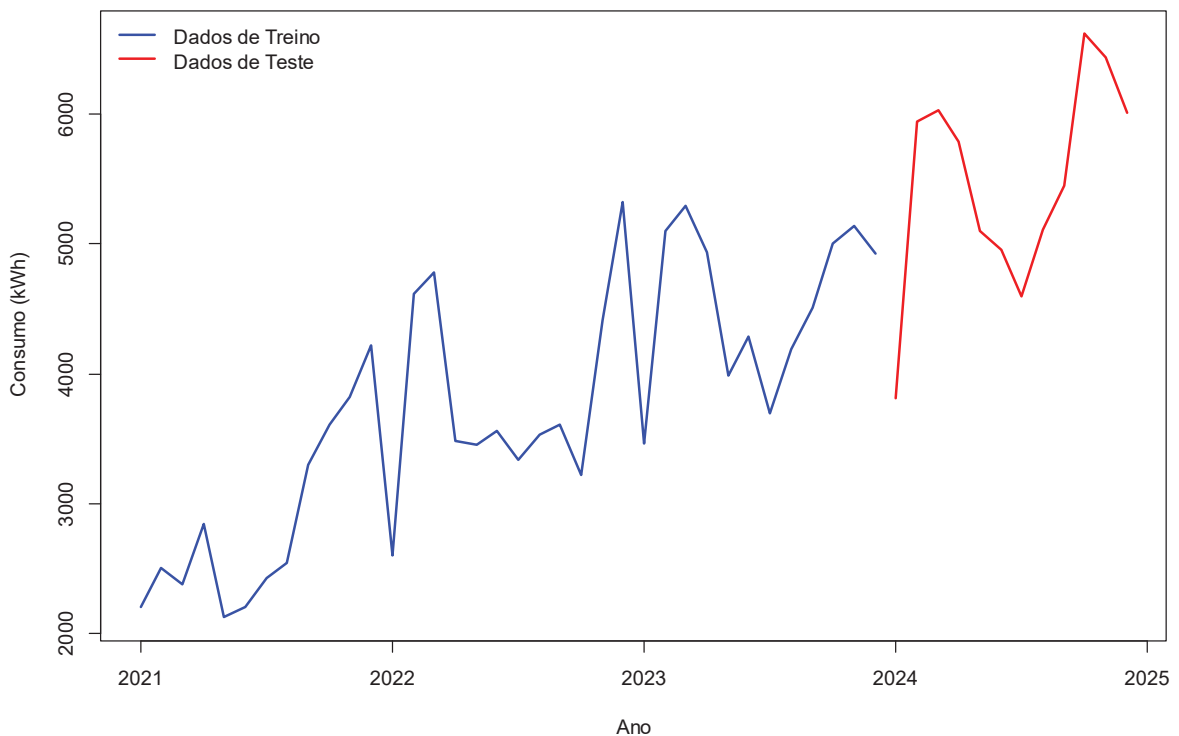


Fonte: O autor (2025).

Analisando a Figura 2, nota-se uma clara tendência de crescimento no consumo ao longo dos anos, assim como uma forte sazonalidade, com maior consumo nos períodos de final e início do ano, mas com uma queda relevante de consumo no primeiro mês do ano.

Após essa análise visual, os dados foram divididos em dados de treino e de teste, onde os dados de treino foram utilizados para estimar o modelo da série e os dados de teste serviram como um parâmetro para analisar a precisão da série. A Figura 3 mostra a série temporal dividida nestes dois subconjuntos, com o subconjunto de dados de treino.

Figura 3 - Série temporal dividida entre dados de treino e dados de teste



Fonte: O autor (2025).

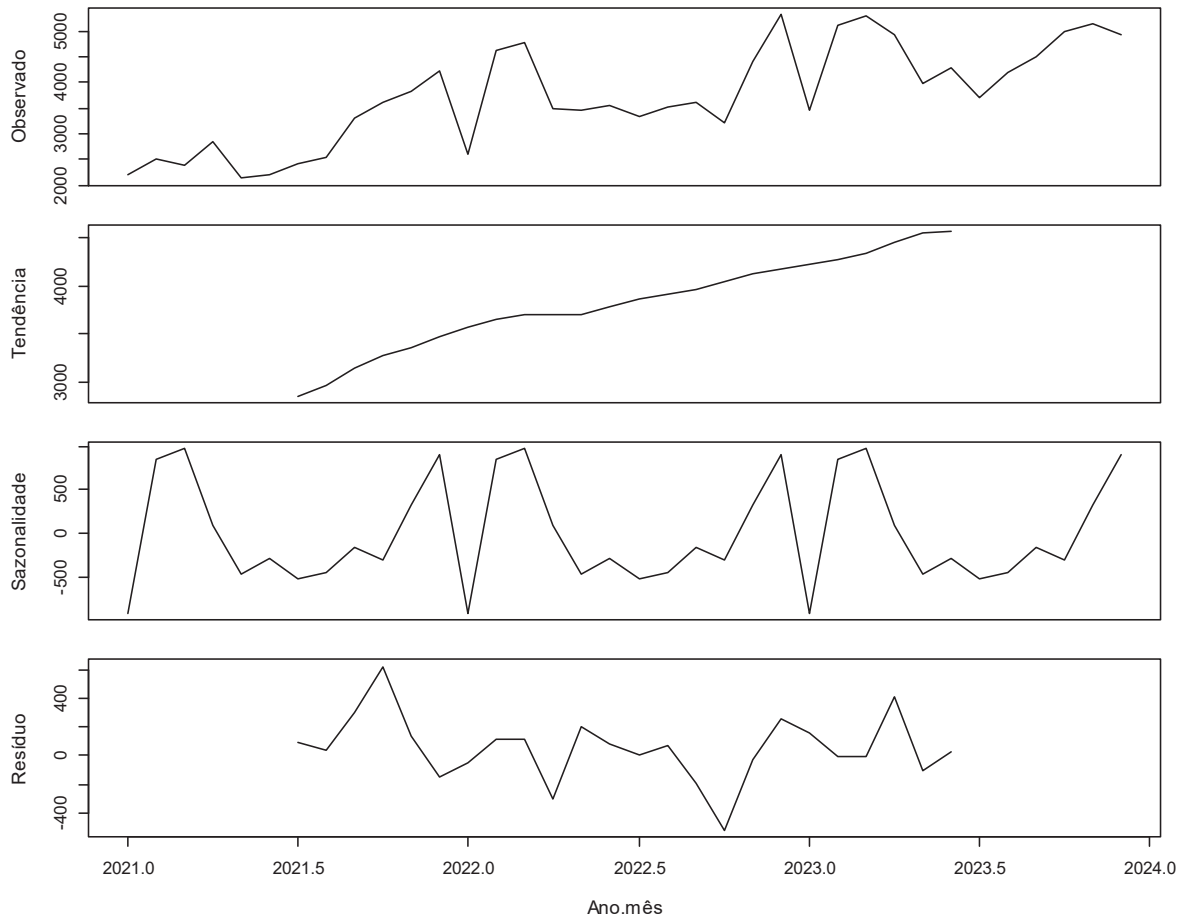
A Figura 3 mostra os dados de treino, representados em azul, que tem início em janeiro de 2021 e terminam em dezembro de 2023, junto dos dados de teste, representados em vermelho, no período entre janeiro e dezembro de 2024.

Após esta separação em dois subconjuntos, iniciou-se a análise do subconjunto de treino com o objetivo de estimar o modelo que melhor representa a série temporal, iniciando pela decomposição dos dados de treino, conforme a Figura 4.

Na Figura 4, pode-se visualizar o primeiro gráfico no topo, mostrando a série temporal dos dados de treino, ou seja, o conjunto de dados observado. No segundo gráfico, nota-se a componente de tendência extraída desse subconjunto, seguido

pela componente de sazonalidade. Por último, no painel inferior, tem-se os dados aleatórios que compõem a série observada.

Figura 4 - Decomposição da série temporal de treino



Fonte: O autor (2025).

Essas decomposições evidenciam como cada componente atua sobre a série de estudo, com a tendência de crescimento, a sazonalidade com maior consumo no verão (início e fim de ano), incluindo uma queda considerável de consumo que ocorre isoladamente no mês de janeiro de cada ano. Os dados aleatórios mostram a componente onde não se pode notar um padrão de sazonalidade ou tendência, o que indica que a decomposição foi eficaz em separar os componentes de tendência e sazonalidade da série.

A análise descritiva, apresentada na Tabela 2, apresenta as características da distribuição do consumo de energia no período de treino. Nota-se que o consumo oscila bastante entre o valor mínimo e máximo. A média, sendo superior à mediana,

por conta de influência da sazonalidade, mostra que não é uma boa referência para a previsão, justificando a necessidade de um método mais sofisticado de previsão.

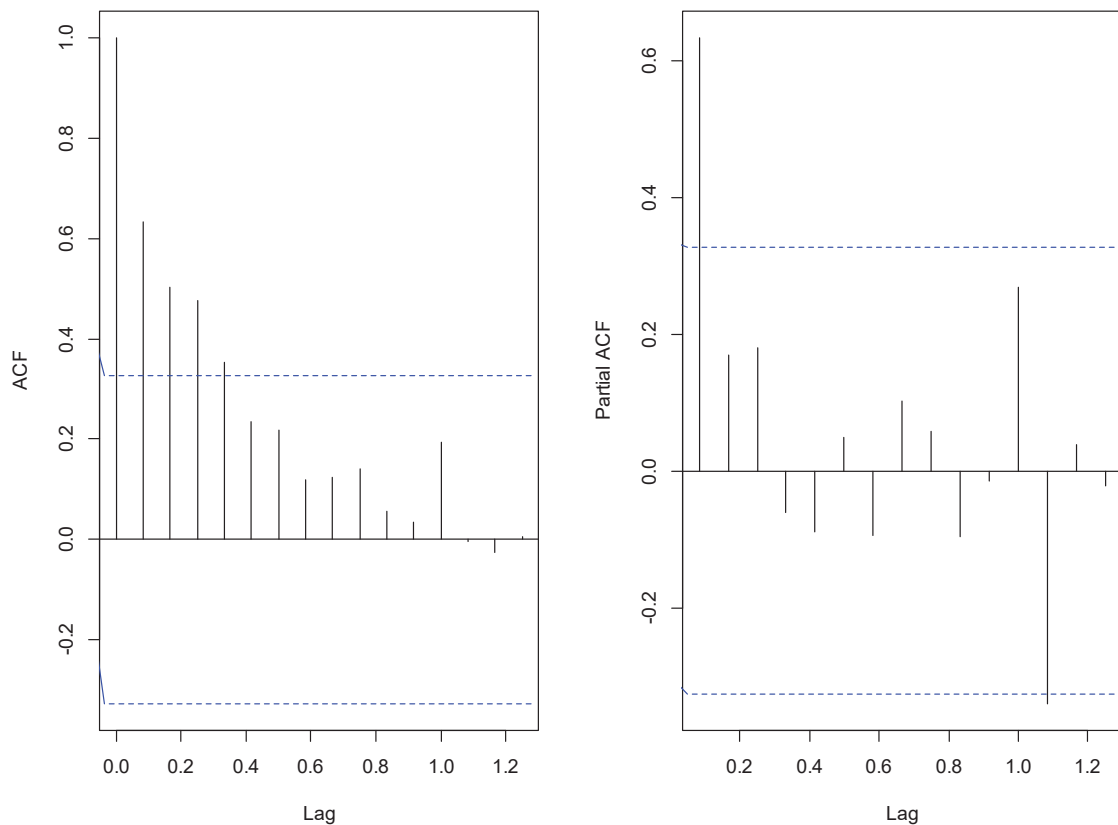
Tabela 2 - Análise descritiva dos dados de treino

<i>Medida Estatística</i>	<i>Valor (kWh)</i>
Mínimo	2127
1º Quartil (Q1)	3125
Mediana	3608
Média	3739
3º Quartil (Q3)	4530
Máximo	5325

Fonte: O autor (2025).

Em seguida, também foram gerados os gráficos de autocorrelação e autocorrelação parcial, conforme a Figura 5.

Figura 5 - Gráficos de autocorrelação dos dados de treino



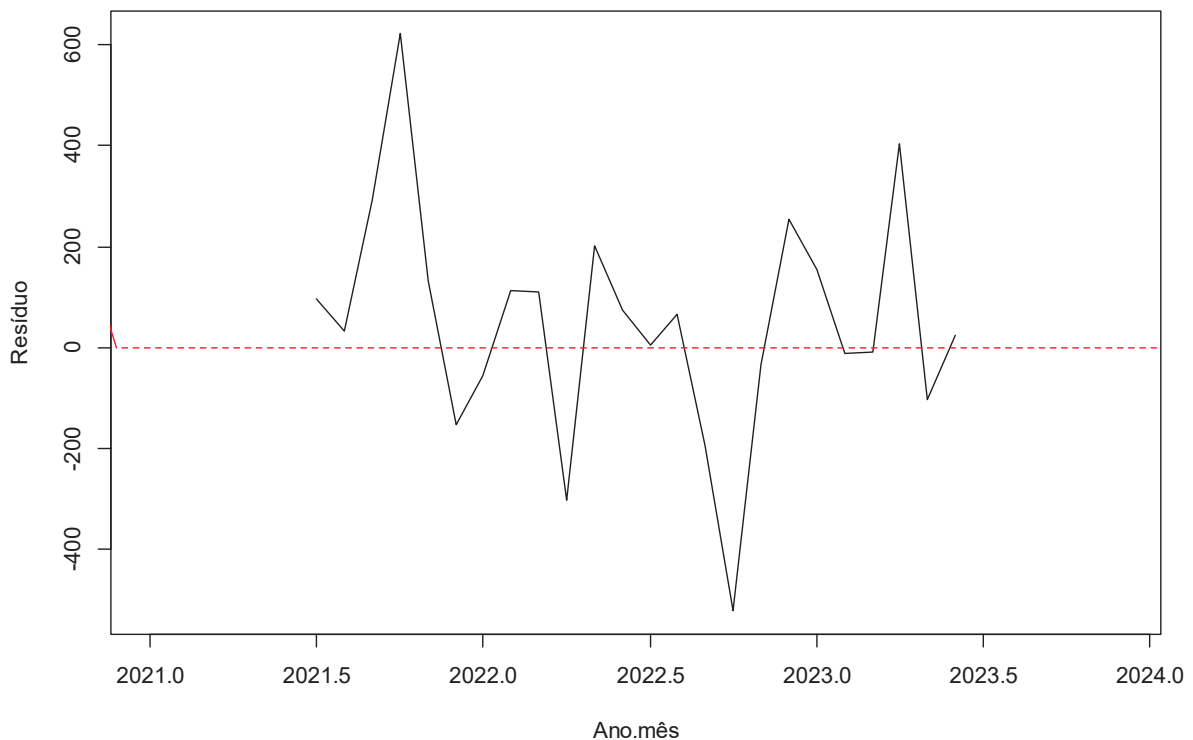
Fonte: O autor (2025).

Na Figura 5, o gráfico de autocorrelação é apresentado à esquerda e o gráfico de autocorrelação parcial à direita. O gráfico de autocorrelação mostra um decaimento lento, o que é um sinal de uma série não estacionária com forte tendência, confirmando a necessidade de diferenciação no modelo. Já o gráfico de autocorrelação parcial exibe um padrão ondulatório, que sugere a presença do componente sazonal e necessidade de diferenciação sazonal no modelo.

Essas análises indicam a necessidade de um modelo robusto, como o SARIMA, capaz de lidar tanto com a tendência quanto com a sazonalidade da série temporal de consumo.

Antes de partir pra modelagem, foi preciso avaliar a necessidade de transformação dos dados visando estabilizar a sua variância. O gráfico para análise dos resíduos da decomposição da série de treino é apresentado na Figura 6.

Figura 6 - Componente aleatório da série de treino



Fonte: O autor (2025).

Observando a Figura 6, não se nota uma clara diferença de magnitude ao longo tempo como um padrão de funil nos resíduos. Pela análise visual, não houve necessidade clara de estabilizar a variância, mas isto confirmou-se com um teste estatístico, estimando o parâmetro de Box-Cox, onde o valor do lambda estimado foi

de 1,225822. Como o valor de lambda foi bem próximo de um, isso indica que não houve necessidade de transformação dos dados para estabilizar a variância.

Após toda essa análise do conjunto de dados de treino, a função de estimativa automática dos parâmetros do modelo ARIMA foi aplicada aos dados de treino, resultando nos parâmetros do modelo exibidos na Tabela 3.

Tabela 3 - Parâmetros do modelo estimado

<i>Componente</i>	<i>Parâmetro</i>	<i>Ordem</i>
Não Sazonal (p,d,q)	--	--
Autoregressivo (AR)	p	0
Diferenciação (I)	d	1
Média Móvel (MA)	q	1
Sazonal (P,D,Q)[s]	--	--
Autoregressivo Sazonal (SAR)	P	1
Diferenciação Sazonal (SI)	D	1
Média Móvel Sazonal (SMA)	Q	0
Período Sazonal	s	12

Fonte: O autor (2025).

Os parâmetros estimados e apresentados na Tabela 3, podem ser interpretados da seguinte forma:

- a) **p = 0**: O modelo não usa valores passados não sazonais;
- b) **d = 1**: A série foi diferenciada uma vez para remover a tendência;
- c) **q = 1**: O modelo usa o erro da previsão do período anterior;
- d) **P = 1**: O modelo usa o valor do mesmo mês do ano anterior;
- e) **D = 1**: Foi aplicada diferenciação sazonal para remover a sazonalidade;
- f) **Q = 0**: O modelo não usa erros de previsões sazonais passadas;
- g) **s = 12**: O padrão sazonal se repete a cada doze meses.

O resultado dos parâmetros estimados comprova análises anteriores como a necessidade de diferenciação de *lag* igual a um e diferenciação sazonal, a fim de remover tendência e sazonalidade. Após estabilizar a série com estas diferenciações, a correlação remanescente foi modelada por dois componentes principais, a média móvel para capturar as dinâmicas de curto prazo e um termo

autorregressivo sazonal, que utiliza informações dos anos anteriores para melhorar a previsão. A ausência dos demais termos, como o autorregressivo e a média móvel sazonal, mostra que o modelo segue o princípio da simplicidade, utilizando somente o necessário para descrever o comportamento da série. Assim, evita o sobreajuste (ou *overfitting*), que ocorre quando um modelo se ajusta tanto aos dados de treino, que incorpora seus ruídos, tendo um ótimo desempenho para descrever o passado, mas gerando previsões imprecisas para o futuro.

Os detalhes estatísticos do modelo ajustado, como os valores dos coeficientes e as medidas de qualidade do ajuste, são apresentados na Tabela 4.

Tabela 4 - Detalhes estatísticos do modelo

Componente	Coefficiente	Valor Estimado	Erro Padrão (s.e.)
Média Móvel (MA)	Ma1	-0,6585	0,2125
Autorregressivo sazonal (SAR)	Sar1	-0,6402	0,1825
Medidas de Ajuste	Critério	Valor	--
Variância dos Resíduos	σ^2	315070	--
Critério de Akaike Corrigido	AICc	368,54	--
Critério Bayesiano	BIC	370,58	--

Fonte: O autor (2025).

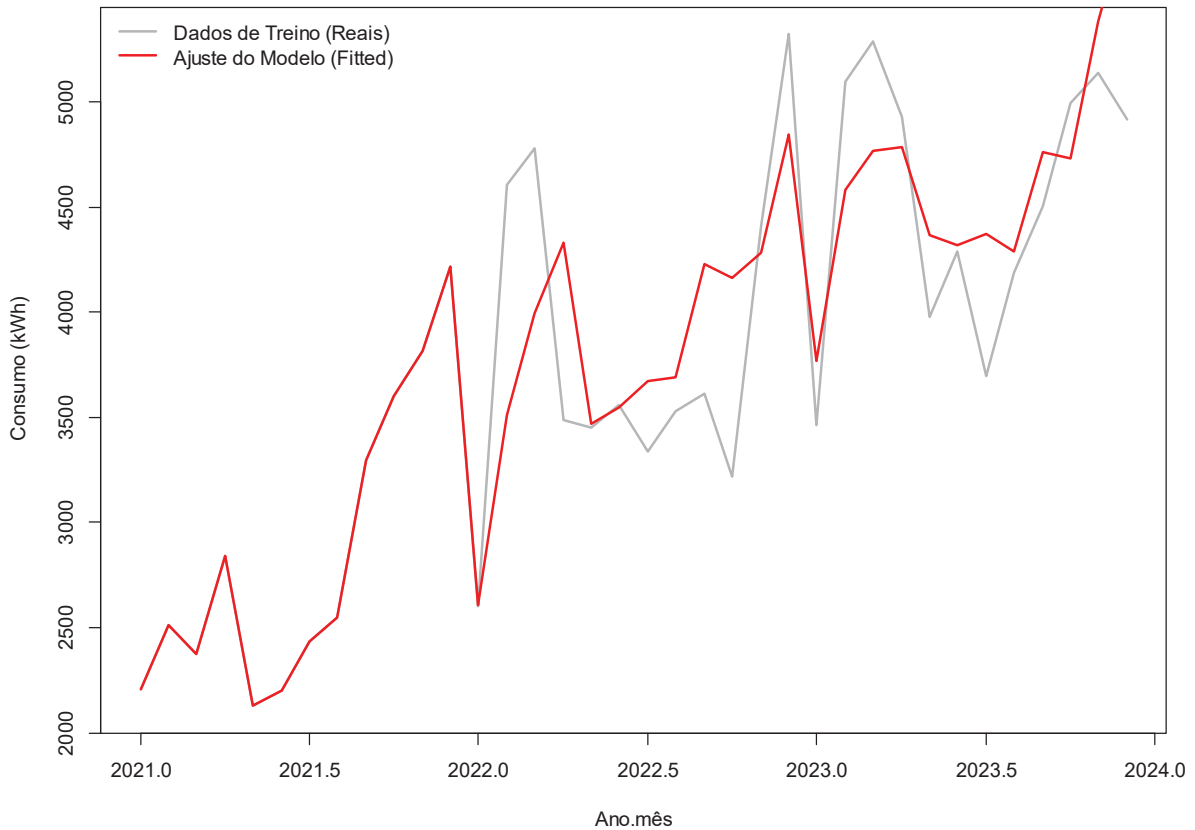
Na Tabela 4, os coeficientes de média móvel e autorregressivo sazonal representam os pesos que o modelo aplica ao erro da previsão anterior e ao valor do mesmo mês do ano anterior, respectivamente. O valor do Critério de Akaike Corrigido (AICc) foi o menor entre os modelos candidatos, validando a escolha do modelo.

A Figura 7 ilustra a aderência do modelo ajustado (linha vermelha) comparado aos dados reais de treino (linha cinza). A análise visual mostra que o ajuste não foi perfeito, mas o modelo foi capaz de capturar a dinâmica geral da série, seguindo a tendência de crescimento e os principais picos e vales sazonais. Acredita-se que ao utilizar uma base de dados maior como treino, o ajuste do modelo poderá ser mais próximo aos valores reais.

Para análise dos resíduos, foram gerados três gráficos, um com os resíduos do modelo, outro que exibe a autocorrelação entre os resíduos e por fim um que

mostra a normalidade dos resíduos. Todos estes gráficos estão ilustrados na Figura 8.

Figura 7 - Ajuste do modelo aos dados de treino

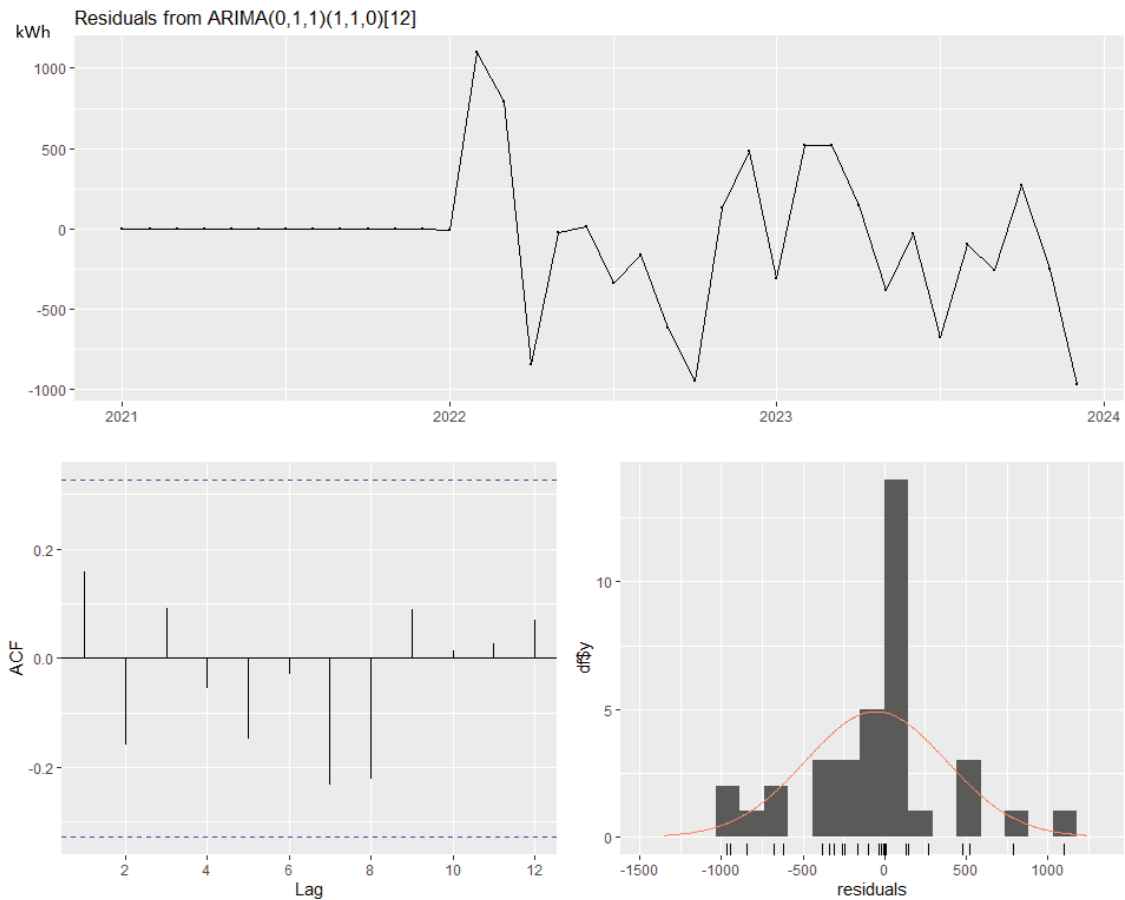


Fonte: O autor (2025).

Analisando a Figura 8, nota-se que o gráfico superior, que apresenta os resíduos do modelo, exibe os erros flutuando em torno de zero, sem apresentar tendências ou padrões sazonais. O gráfico de autocorrelação (ACF), mostra que nenhuma correlação dos resíduos é estatisticamente significativa, pois todas as barras permanecem dentro dos limites de confiança. O histograma, no painel inferior direito, mostra que a distribuição dos resíduos é aproximadamente normal e centrada em zero. Estas conclusões contribuem para confirmar a hipótese de que os resíduos se comportam como um ruído branco, o que indica que o modelo está bem ajustado, dando confiabilidade às previsões geradas a partir dele.

A fim de complementar a análise visual, foram aplicados testes estatísticos para confirmar se os resíduos se comportam como ruído branco. Os resultados desses testes estatísticos são apresentados na Tabela 5.

Figura 8 - Diagnóstico dos resíduos do modelo



Fonte: O autor (2025).

Tabela 5 - Resultados dos testes de diagnóstico dos resíduos do modelo estimado

Teste Estatístico	Hipótese Nula (H_0)	Estatística do Teste	Valor-p (p-value)	Conclusão ($\alpha = 0,05$)
Ljung-Box	Os resíduos são independentes (não há autocorrelação)	$Q^* = 6,0414$	0,3022	Não Rejeita H_0
Jarque-Bera	Os resíduos seguem uma distribuição Normal	$X^2 = 1,2092$	0,5463	Não Rejeita H_0
ARCH-LM	Os resíduos são homocedásticos (tem variância constante)	$X^2 = 7,5172$	0,8216	Não Rejeita H_0

Fonte: O autor (2025).

Conforme apresentado na Tabela 5, os valores-p para os testes de Ljung-Box, Jarque-Bera e o teste de Heterocedasticidade Condicional Autorregressiva (ARCH-LM) são todos significativamente superiores ao nível de significância definido ($\alpha = 0,05$). Estes resultados permitem a não rejeição das hipóteses nulas dos testes,

confirmando que os resíduos são independentes, seguem uma distribuição normal e apresentam variância constante. Desta forma, conclui-se que o modelo estimado é estatisticamente válido e adequado para gerar previsões.

Após a validação do modelo, foi realizada a previsão de consumo para os doze meses à frente. Os resultados dessa previsão são apresentados na Tabela 6.

Tabela 6 - Resultado da previsão para o ano de 2024

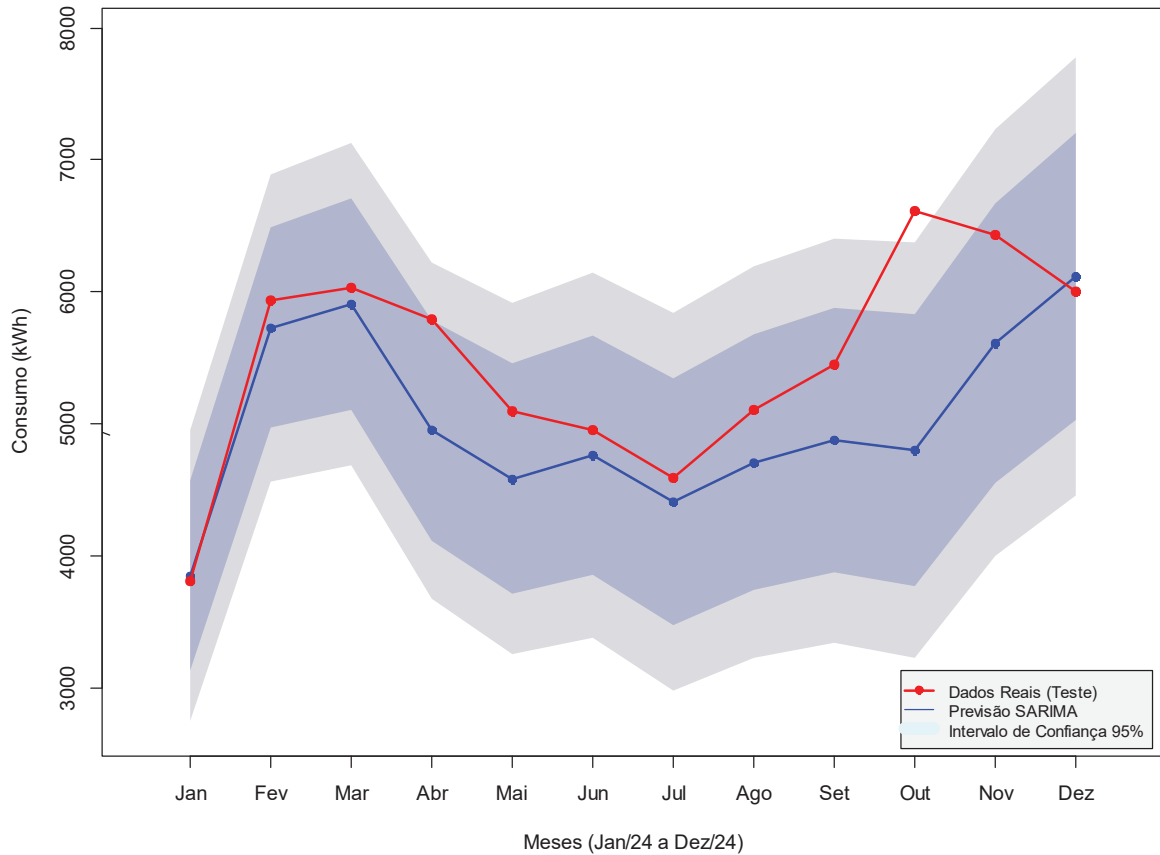
Mês	Valor Previsto	Limite Inferior (IC 80%)	Limite Superior (IC 80%)	Limite Inferior (IC 95%)	Limite Superior (IC 95%)
jan./24	3850	3130	4569	2750	4950
fev./24	5727	4967	6487	4564	6889
mar./24	5906	5107	6705	4684	7128
abr./24	4948	4112	5784	3670	6226
mai./24	4582	3711	5453	3250	5915
jun./24	4761	3856	5666	3377	6145
jul./24	4409	3471	5347	2975	5843
ago./24	4709	3740	5679	3226	6192
set./24	4874	3873	5874	3344	6403
out./24	4799	3770	5829	3224	6374
nov./24	5614	4555	6673	3995	7233
dez./24	6120	5033	7206	4457	7782

Fonte: O autor (2025).

A Tabela 6 apresenta os valores previstos para cada mês e os seus respectivos intervalos de confiança. Os intervalos de confiança de 80% e 95% quantificam a incerteza associada a cada previsão, fornecendo uma faixa de valores plausíveis para o consumo real. Pode-se observar que a amplitude desses intervalos aumenta para os meses mais distantes no futuro, revelando a maior incerteza das previsões de longo prazo.

Para uma melhor visualização dos resultados, foram gerados gráficos comparativos entre os dados reais e os dados previstos. A Figura 9 apresenta os dados previstos comparados aos dados do conjunto de teste, junto com os intervalos de confiança. Já a Figura 10, exibe uma visão geral da série temporal, acrescentando os dados reais de treino ao comparativo entre os dados previstos e os dados reais de teste.

Figura 9 - Comparativo entre Valores Previstos e Observados para o Período de Teste

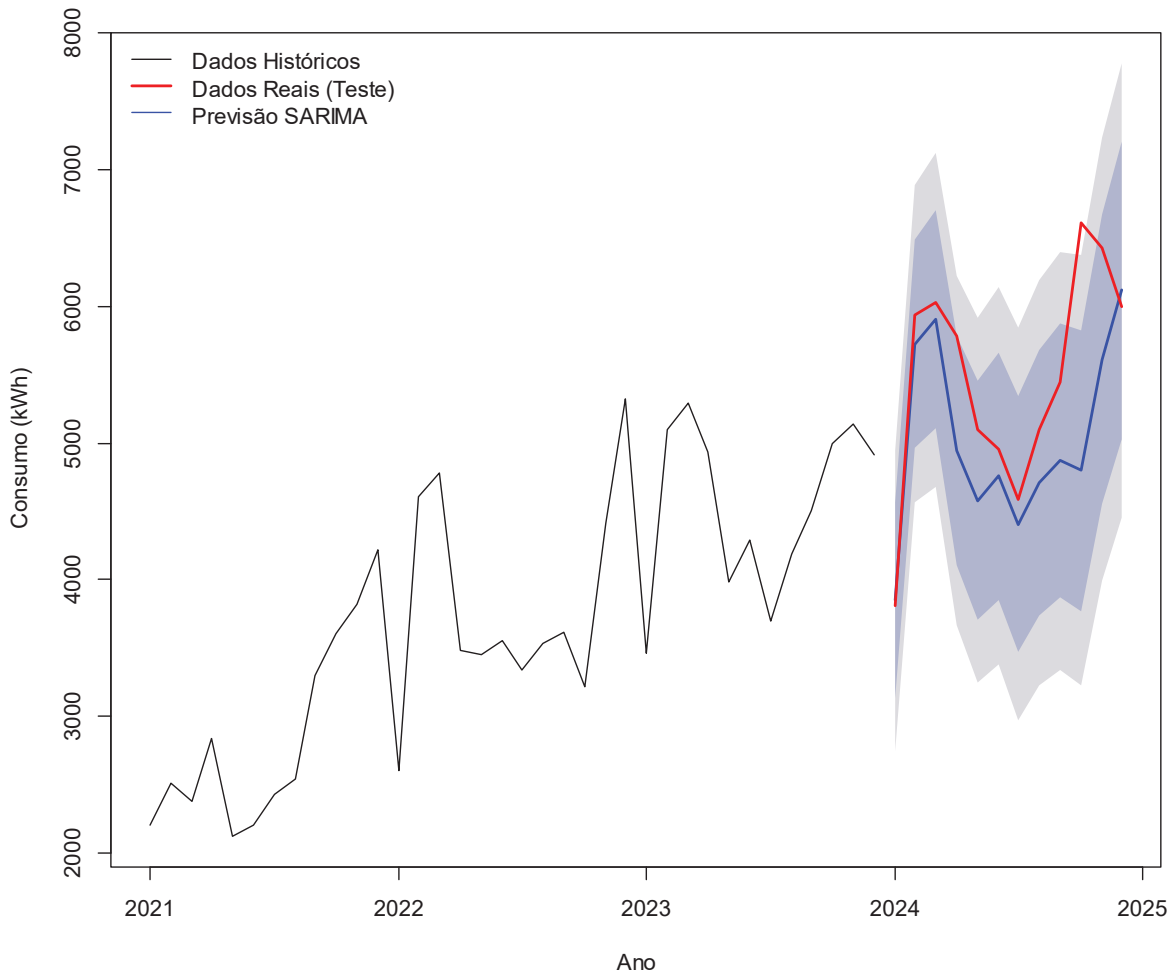


Fonte: O autor (2025).

A Figura 9 foca no período de teste, comparando a previsão (em azul) com os dados reais (em vermelho). Analisando visualmente, nota-se que o modelo foi eficiente em capturar a dinâmica da série temporal, prevendo o aumento de consumo no período de verão (início e fim de ano), incluindo o vale no mês de janeiro, e o menor consumo no período de inverno. Porém, a análise também revela duas limitações, uma tendência de subestimação, onde a previsão apresenta geralmente valores mais baixos do que os dados reais de consumo, e a incapacidade de prever o pico de consumo em outubro, onde apresentou o maior erro de previsão.

Enquanto isso, a Figura 10 apresenta uma visão geral da série temporal de consumo, demonstrando que a previsão para o ano de 2024 se comporta como uma continuação coerente, considerando tanto a tendência de crescimento quanto a sazonalidade observada nos dados do conjunto de treino.

Figura 10 - Previsão no Contexto da Série Temporal Histórica



Fonte: O autor (2025).

Apesar da sua tendência de subestimação, nota-se que a série do conjunto de teste se manteve quase integralmente dentro do intervalo de confiança de 95%, na área azul levemente mais escura. Apenas no mês de outubro, onde houve um pico inesperado de consumo, o valor não esteve dentro do intervalo de confiança. Isso indica que mesmo que o modelo não tenha sido perfeitamente preciso, conseguiu se aproximar de forma satisfatória dos valores reais, e foi eficiente em quantificar a sua incerteza, gerando a faixa de valores que abrangeram os dados reais de consumo na maioria dos meses.

A partir dessa previsão gerada, foram utilizados indicadores de precisão, com base nos dados reais do conjunto de teste, para avaliar a precisão dos dados previstos. A Tabela 7 apresenta os indicadores de precisão dos dados previstos como um todo.

Tabela 7 – Indicadores de precisão do modelo

Conjunto de dados	ME (kWh)	RMSE (kWh)	MAE (kWh)	MPE (%)	MAPE (%)	MASE	ACF1	U de Theil
Conjunto de treino	-54,06	428,71	274,45	-2,10	6,60	0,30	0,16	NA
Conjunto de teste	458,87	680,93	484,93	7,79	8,29	0,53	0,20	0,67

Fonte: O autor (2025).

Analisando a Tabela 7, obteve-se o desempenho do modelo ao gerar previsões. A métrica mais importante, o Erro Percentual Absoluto Médio (MAPE) para o conjunto de teste, foi de 8,29%, indicando que, em média, a previsão teve aproximadamente 8% de diferença para o valor real, o que é um resultado bom considerando a dinâmica dos dados e a aplicação a qual pretende-se utilizar esta previsão. A métrica de Erro Percentual Médio (MPE) de 7,79%, por ser um valor positivo, indica a tendência leve de subestimar o consumo, assim como notou-se na análise visual. Esse comportamento sugere que a intensidade do crescimento do consumo no período de teste superou a tendência histórica média capturada pelo modelo nos dados de treino. As métricas relativas, como MASE e U de Theil, são significativamente menores do que um valor unitário, indicando que o modelo SARIMA é superior à uma análise ingênua. Já a métrica de função de autocorrelação (ACF1), por ser um valor próximo de zero, indica uma correlação muito fraca entre os erros consecutivos de previsão, reforçando que o modelo foi bem ajustado.

Buscando uma análise mais detalhada, mostrando o desempenho mensal das previsões, foram calculados os erros para cada mês, os quais são apresentados na Tabela 8.

Observando a Tabela 8, nota-se que a previsão foi muito precisa entre janeiro e março, como também em dezembro. Já a previsão para o mês de outubro se destaca negativamente, por um erro percentual muito alto, de 27,44%. Este erro alto sugere a ocorrência de algum evento atípico não capturado no padrão histórico. Para os demais meses de previsão, o erro foi bem aceitável para a aplicação na compensação de energia elétrica.

Tabela 8 – Indicadores de precisão das previsões mensais

Mês	Consumo Real (kWh)	Previsão (kWh)	Erro Absoluto (kWh)	Erro Percentual (%)
Jan	3809	3849,80	40,80	1,07
Fev	5935	5726,94	208,06	3,51
Mar	6029	5906,06	122,94	2,04
Abr	5788	4947,94	840,06	14,51
Mai	5097	4582,34	514,66	10,10
Jun	4953	4760,94	192,06	3,88
Jul	4591	4408,89	182,11	3,97
Ago	5104	4709,12	394,88	7,74
Set	5447	4873,60	573,40	10,53
Out	6614	4799,40	1814,60	27,44
Nov	6434	5613,94	820,06	12,75
Dez	6004	6119,54	115,54	1,92

Fonte: O autor (2025).

Para contextualizar a relevância prática do método de previsão, foi realizado um estudo de caso simulando a alocação de energia no cenário da geração distribuída. Foram simulados dois demonstrativos de saldo para o ano de 2024, um com a energia injetada baseada na previsão de consumo gerada pelo modelo SARIMA e o outro com a energia injetada baseada na média de consumo de todo o ano de 2023. A Tabela 9 apresenta o demonstrativo simulado com a previsão baseada na previsão do modelo SARIMA. Já a Tabela 10 exibe o demonstrativo simulado baseado na média de consumo dos últimos doze meses.

Analisando a Tabela 9, nota-se pelos valores totais, que boa parte da energia injetada foi utilizada para compensação do consumo, não armazenando muito saldo ao longo dos meses. Porém, entre abril e novembro não houve saldo suficiente para compensar todo o consumo possível, totalizando um saldo faltante de 4526 kWh ao longo do ano todo. Este valor de saldo faltante não é considerado muito alto, sendo totalmente aceitável na prática, já que houve compensação de aproximadamente 93% de todo o consumo compensável.

Fica claro como a subestimação da previsão de consumo afetou a compensação, causando esta falta de saldo disponível para compensação total do consumo de energia. No entanto, este resultado não invalida o modelo, mas

demonstra a sensibilidade do balanço energético à precisão da previsão do consumo.

Tabela 9 - Simulação de Demonstrativo de Saldo baseado na previsão por modelo SARIMA

<i>Mês</i>	<i>Energia Injetada</i>	<i>Consumo Real</i>	<i>Consumo Mínimo</i>	<i>Energia Compensada</i>	<i>Saldo Mês Anterior</i>	<i>Saldo Mês</i>	<i>Saldo Final</i>	<i>Energia Faltante</i>
01/2024	3849	3809	100	3709	0	140	140	0
02/2024	5727	5935	100	5835	140	-108	32	0
03/2024	5905	6029	100	5929	32	-24	8	0
04/2024	4947	5788	100	4955	8	-8	0	733
05/2024	4582	5097	100	4582	0	0	0	415
06/2024	4761	4953	100	4761	0	0	0	92
07/2024	4409	4591	100	4409	0	0	0	82
08/2024	4709	5104	100	4709	0	0	0	295
09/2024	4874	5447	100	4874	0	0	0	473
10/2024	4799	6614	100	4799	0	0	0	1715
11/2024	5613	6434	100	5613	0	0	0	721
12/2024	6120	6004	100	5904	0	216	216	0
Total	60295	65805	1200	60079	--	--	216	4526

Fonte: O autor (2025)

A Tabela 10 detalha a simulação do balanço energético utilizando o método padrão de alocação, baseado na média de consumo dos últimos doze meses. Neste modelo a injeção de energia foi estática ao longo do ano, com o valor da média de consumo do ano de 2023, igual a 4.542 kWh. A análise do demonstrativo mostra a ineficácia desta abordagem enfrentando um consumo com tendência de crescimento e sazonalidade. A injeção de energia foi bem menor do que o consumo na maioria dos meses, mostrando a incapacidade de acumular saldo para eventuais picos de consumo inesperados e totalizando 10.101 kWh de saldo faltante para compensação de todo o consumo.

Tabela 10 - Simulação de Demonstrativo de Saldo baseado na média dos últimos doze meses

<i>Mês</i>	<i>Energia Injetada</i>	<i>Consumo Real</i>	<i>Consumo Mínimo</i>	<i>Energia Compensada</i>	<i>Saldo Mês Anterior</i>	<i>Saldo Mês</i>	<i>Saldo Final</i>	<i>Energia Faltante</i>
01/2024	4542	3809	100	3709	0	833	833	0
02/2024	4542	5935	100	5375	833	-833	0	460
03/2024	4542	6029	100	4542	0	0	0	1387
04/2024	4542	5788	100	4542	0	0	0	1146
05/2024	4542	5097	100	4542	0	0	0	455
06/2024	4542	4953	100	4542	0	0	0	311
07/2024	4542	4591	100	4491	0	51	51	0
08/2024	4542	5104	100	4593	51	-51	0	411
09/2024	4542	5447	100	4542	0	0	0	805
10/2024	4542	6614	100	4542	0	0	0	1972
11/2024	4542	6434	100	4542	0	0	0	1792
12/2024	4542	6004	100	4542	0	0	0	1362
Total	54504	65805	1200	54504	--	--	0	10101

Fonte: O autor (2025).

A Tabela 11 apresenta uma análise comparativa do desempenho dos métodos de alocação nas simulações. O indicador em que se nota maior diferença é a “Energia Faltante”, que representa a falta de saldo para compensação total do consumo compensável, e conseqüentemente a quantidade de energia que precisou ser adquirida da concessionária, apresentou uma redução drástica de 55,19% ao se utilizar o método de alocação por previsão baseada em SARIMA. Esta expressiva redução da energia faltante, reflete diretamente no indicador de “Eficiência de Compensação”, que representa o quanto de consumo compensável foi realmente compensado com a injeção de saldo, onde saltou de 84,36% para 92,99%. Já o impacto financeiro foi significativo, resultando uma economia adicional de R\$ 3.512,25, a qual representa um ganho de 10,23% em relação à abordagem pela média.

Tabela 11 - Análise comparativa de desempenho das estratégias de alocação

<i>Indicador</i>	<i>Método Previsão SARIMA</i>	<i>Método Média 12 meses</i>	<i>Diferença Absoluta</i>	<i>Diferença (%)</i>
Consumo Compensável	64.605 kWh	64.605 kWh	--	--
Energia Injetada	60.295 kWh	54.504 kWh	5.791 kWh	10,62%
Energia Compensada	60.079 kWh	54.504 kWh	5.575 kWh	10,23%
Energia Faltante	4.526 kWh	10.101 kWh	-5.575 kWh	-55,19%
Saldo Final	216 kWh	0 kWh	216 kWh	--
Taxa de Compensação	92,99%	84,36%	+8,63 p.p.	--
Valor Recebido (Economia)	R\$ 37.849,77	R\$ 34.337,52	R\$ 3.512,25	10,23%

Fonte: O autor (2025).

Estes resultados comprovam que a aplicação de um modelo de previsão baseado em SARIMA, que é capaz de capturar a sazonalidade e tendência da série histórica de consumo, gera uma compensação de energia mais eficiente e com isso também gera um ganho econômico significativo na gestão de créditos de energia.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo principal desenvolver um modelo preditivo de séries temporais, utilizando a metodologia SARIMA, para a previsão de consumo de uma unidade consumidora, visando otimização nas alocações de créditos de energia. O problema de pesquisa surgiu da necessidade de otimizar as alocações de energia, considerando que uma gestão baseada em métodos simples de previsão, como a média histórica, gera perdas financeiras significativas quando o padrão de consumo apresenta características como tendência e sazonalidade.

A fim de alcançar este objetivo geral, os objetivos específicos definidos foram sequencialmente atingidos. Primeiramente, a análise exploratória da série temporal de consumo permitiu a visualização e constatação de características importantes dos padrões dos dados. Esta etapa evidenciou a natureza não estacionária da série, revelando uma clara tendência de crescimento e uma forte sazonalidade anual, o que justifica a escolha da metodologia SARIMA.

Após esta análise exploratória, foi estimado o modelo SARIMA, cujos parâmetros foram capazes de modelar tanto a tendência quanto a sazonalidade da série. A análise dos resíduos deste modelo, utilizando testes estatísticos como Ljung-Box, Jarque-Bera e ARCH-LM, validou a robustez estatística do modelo.

A capacidade preditiva do modelo foi validada através dos indicadores de precisão calculados a partir da comparação dos valores previstos pelo modelo com os valores reais da série temporal. O Erro Percentual Absoluto Médio (MAPE), foi o principal indicador e apresentou um valor de 8,29%, considerado satisfatório para a aplicação prática na gestão de créditos de energia. O Coeficiente U de Theil comprovou que o modelo SARIMA é significativamente superior à previsão ingênua, validando sua capacidade preditiva, apesar da leve tendência de subestimação, constatada pelo Erro Médio (ME).

Ao analisar o impacto das previsões em um cenário de alocação de energia, a simulação demonstrou que a alocação de energia baseada na previsão de consumo baseada em SARIMA é significativamente superior ao método comum. O modelo proposto gerou uma economia adicional de R\$ 3.512,25, que representa um ganho de 10,23% ao longo de um ano.

Desta forma, o trabalho atende ao problema de pesquisa, comprovando que a aplicação de um modelo preditivo SARIMA é capaz de otimizar a distribuição de

créditos de energia na geração distribuída. Ao capturar a tendência e a sazonalidade, o modelo permite a alocação mais eficiente dos créditos de energia, minimizando perdas e maximizando o retorno financeiro do investimento.

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

O estudo realizado neste trabalho abre diversas possibilidades de expansão para estudos em trabalhos futuros. Uma expansão natural seria utilizar a metodologia SARIMAX, já explorada na literatura de referência. A inclusão de variáveis externas, como feriados e dados climáticos, permitiria avaliar se há melhoria na precisão das previsões e também se poderia melhorar a modelagem de picos atípicos de consumo.

Outra evolução para este trabalho, seria adotar técnicas de *machine learning* para modelagem da série temporal. Conforme explorado na revisão de literatura, seria interessante comparar a capacidade preditiva da modelagem SARIMA com modelos baseados em redes neurais. Estes modelos de *machine learning* tem a capacidade de capturar padrões não-lineares complexos, que o SARIMA pode não ter modelado totalmente, e assim poderia melhorar a precisão das previsões.

Além de comparar com outras técnicas de modelagem, trabalhos futuros podem se utilizar do método de modelagem deste estudo para implementar um fluxo automatizado, capaz de coletar novos dados de consumo para expandir a série de dados, retreinar o modelo e realizar previsões atualizadas periodicamente e apresentar estas previsões em um dashboard para o gestor de energia, transformando o estudo em uma ferramenta de auxílio à decisão.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Resolução Normativa nº 1.000, de 7 de dezembro de 2021**. Estabelece as Regras de Prestação do Serviço Público de Distribuição de Energia Elétrica e revoga as Resoluções Normativas nº 414, de 9 de setembro de 2010; nº 470, de 13 de dezembro de 2011; nº 901, de 8 de dezembro de 2020 e dá outras providências. Brasília: Diário Oficial da União, 2021. Disponível em: www2.aneel.gov.br/cedoc/ren20211000.pdf#page=2.67. Acesso em: 07 out. 2025.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Resolução Normativa nº 1.059, de 21 de março de 2023**. Estabelece os procedimentos e as condições para a outorga de autorização de que tratam o art. 8º da Lei nº 9.074, de 7 de julho de 1995, e o art. 1º da Lei nº 14.120, de 3 de março de 2021, e para a outorga de concessão de aproveitamentos hidrelétricos de que trata o art. 26 da Lei nº 9.074, de 1995. Brasília: Diário Oficial da União, 2021. Disponível em: <https://www2.aneel.gov.br/cedoc/ren20231059.pdf>. Acesso em: 07 out. 2025.

BAJAY, Sérgio et al. **Geração distribuída e eficiência energética**: reflexões para o setor elétrico de hoje e do futuro. Campinas: International Energy Initiative – IEI Brasil, 2018.

BRASIL. **Lei nº 14.300, de 6 de janeiro de 2022**. Institui o marco legal da microgeração e minigeração distribuída, o Sistema de Compensação de Energia Elétrica (SCEE) e o Programa de Energia Renovável Social (PERS); altera as Leis nºs 10.848, de 15 de março de 2004, e 9.427, de 26 de dezembro de 1996; e dá outras providências. Brasília: Presidência da República, 2022. Disponível em: www.gov.br/mme/pt-br/aceso-a-informacao/legislacao/leis/lei-n-14-300-2022.pdf. Acesso em: 07 out. 2025.

BUENO, Rodrigo de Losso da Silveira. **Econometria de séries temporais**. São Paulo: Cengage Learning, 2008.

BURNHAM, Kenneth P.; ANDERSON, David R. **Model selection and multimodel inference**: a practical information-theoretic approach. 2. ed. New York: Springer-Verlag, 2002.

COOK, S. **Forecast Evaluation using Theil's Inequality Coefficients**. Swansea: The Economics Network, 2019. Disponível em: https://economicsnetwork.ac.uk/showcase/cook_theil. Acesso em: 28 set. 2025.

FERREIRA, Pedro Costa. **Análise de séries temporais**: modelos dinâmicos e aplicações em R. Rio de Janeiro: Elsevier, 2017.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011. E-book. ISBN 9788580550511. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9788580550511/>. Acesso em: 29 nov. 2025.

- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 18 nov. 2025.
- KALIMOLDAYEV, Maksat et al. Analysis of modern approaches for the prediction of electric energy consumption. **Open Engineering**, [S. l.], v. 10, n. 1, p. 350-361, 2020. Disponível em: <https://doi.org/10.1515/eng-2020-0028>. Acesso em: 16 jul. 2025.
- MOREIRA, José Roberto Simões. **Energias renováveis, geração distribuída e eficiência energética**. Rio de Janeiro: LTC, 2019.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 1. ed. São Paulo: Blucher, 2018. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 27 set. 2025.
- PAULA, Diego Miranda de; JÚNIOR, João Carlos Xavier; MIRANDA, Kléber Formiga. APLICAÇÃO DE SÉRIES TEMPORAIS PARA PREVISÃO DE DESPESAS DE ENERGIA ELÉTRICA DO TRIBUNAL REGIONAL ELEITORAL DO RIO GRANDE DO NORTE. **Brazilian Journal of Development**, v. 6, n. 11, p. 87089–87112, 2020.
- SANTOS, José Airton Azevedo dos; CHAUCOSKI, Yasmin. Previsão do Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando SARIMA e LSTM. **Revista Cereus**, v. 12, n. 4, p. 93–104, 2020.
- SILVA, Fernando da. **MAE, RMSE, ACC, F1, ROC, R2? Avaliação de desempenho de modelos preditivos**. Análise Macro, 2023. Disponível em: <https://analisemacro.com.br/econometria-e-machine-learning/mae-rmse-acc-f1-roc-r2-avaliacao-de-desempenho-de-modelos-preditivos/>. Acesso em: 27 set. 2025.
- SPADINI, Allan; ALENCAR, Valquíria. **Séries temporais com Prophet: análise e previsão de dados**. São Paulo: Casa do Código, 2022. E-book.
- STENGHELE, Gabriel *et al.* Previsão de Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando Sarima. **Revista FSA**, v. 17, n. 1, p. 169–183, 1 fev. 2020.
- VASCONCELOS, Leonardo Pio *et al.* **UMA COMPARAÇÃO DE TÉCNICAS DE REGRESSÃO PARA A PREVISÃO DE CONSUMO DE ENERGIA RESIDENCIAL NO CENÁRIO NACIONAL**. Blumenau: [S.n.]. Disponível em: <https://btdtd.ucb.br:8443/jspui/handle/tede/2645>. Acesso em: 16 jul. 2025.