



ANÁLISE DA SÉRIE TEMPORAL DO CONSUMO DE ENERGIA ELÉTRICA SIMPLES NO CEARÁ

FORTALEZA

2024



UNIVERSIDADE FEDERAL DO CEARÁ

**DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA -
DEMA**

PROFESSOR:

Prof. Dr. Jeffiner .

ALUNOS:

Thalis Rebouças de Oliveira.

Sumário

Sumário	3
Lista de tabelas	4
Lista de ilustrações	4
0.1 Introdução	6
0.1.1 Motivação	6
0.1.2 Apresentação da Série Temporal	6
0.2 Análise Exploratória e Preparação dos Dados	6
0.2.1 Visualização Inicial	6
0.2.2 Análise Sazonal	7
0.2.3 Decomposição da Série	8
0.2.4 Divisão em Conjuntos de Treino e Teste	9
0.3 Modelagem por Suavização Exponencial (ETS)	9
0.3.1 Seleção e Especificação do Modelo	9
0.3.2 Resultados e Diagnóstico	10
0.3.3 Previsão e Validação	10
0.4 Modelagem pela Metodologia Box-Jenkins (SARIMA)	11
0.4.1 Etapa 1: Identificação	11
0.4.1.1 Análise de Estacionariedade	11
0.4.1.2 Diferenciação e Análise FAC/FACP	11
0.4.2 Etapa 2 e 3: Estimção e Diagnóstico	13
0.4.2.1 Modelo 1: SARIMA(1,1,1)(1,0,0)[12]	14
0.4.2.1.1 Lags Não-Sazonais (p, q):	15
0.4.2.1.2 Lags Sazonais (P, Q):	15
0.4.2.1.3 Conclusão para o Ajuste:	15
0.4.2.2 Modelo 2: SARIMA(1,1,1)(1,1,1)[12]	15
0.4.2.2.1 Especificação e Coeficientes:	15
0.4.2.2.2 Performance e Métricas de Erro:	16
0.4.2.2.3 Conclusão e Recomendação:	16
0.4.2.3 Modelo 3: SARIMA(1,1,1)(0,1,1)[12]	18
0.5 Análise de Intervenção	22
0.5.1 Detecção do Outlier	22
0.5.2 Modelagem da Intervenção	22
0.6 Comparação e Seleção do Modelo Final	24
0.6.1 Critérios de Informação	24
0.6.2 Métricas de Acurácia (Conjunto de Treino)	25

0.6.3 Veredito 25

0.7 Conclusão 25

Resumo

Este relatório detalha a análise da série temporal do consumo mensal de energia elétrica no estado do Ceará, no período de janeiro de 2004 a março de 2025. O objetivo principal é identificar os padrões subjacentes da série, como tendência e sazonalidade, e construir um modelo estocástico robusto para previsão. Foram exploradas duas abordagens principais: a família de modelos de Suavização Exponencial (ETS) e a metodologia de Box-Jenkins para modelos SARIMA. Adicionalmente, foi conduzida uma análise de intervenção para capturar o efeito de eventos atípicos. Os modelos foram ajustados a um conjunto de treino (2004-2022) e validados em um conjunto de teste (2023-2025). O modelo final, uma Regressão com Erros SARIMA(1,1,1)(0,1,1)[12], que incorpora uma intervenção para um choque estrutural ocorrido em maio de 2020, demonstrou ser o mais parcimonioso e com maior acurácia, provendo o melhor equilíbrio entre ajuste aos dados e complexidade.

0.1 Introdução

0.1.1 Motivação

A análise do consumo de energia elétrica é fundamental para o planejamento estratégico de um país ou estado. Ela permite que governos e empresas do setor energético antecipem a demanda, otimizem a distribuição, planejem investimentos em infraestrutura e desenvolvam políticas públicas para eficiência energética e sustentabilidade.

O estado do Ceará, com seu dinamismo econômico e crescimento populacional, apresenta uma demanda energética complexa e crescente. Compreender os padrões dessa demanda através de modelos de séries temporais é, portanto, uma tarefa de grande valor. Este estudo se propõe a dissecar a série de consumo mensal para:

- Identificar padrões de tendência de longo prazo e ciclos sazonais.
- Entender o impacto de fatores externos, como crises econômicas ou eventos atípicos (e.g., a pandemia de COVID-19).
- Construir um modelo preditivo confiável que possa auxiliar no planejamento energético futuro.

0.1.2 Apresentação da Série Temporal

A série temporal em estudo, denotada por Z_t , representa o consumo mensal de energia elétrica na rede de Sistema Simples no estado do Ceará, medido em Megawatt-hora (MWh).

Os dados foram obtidos da Empresa de Pesquisa Energética (EPE), vinculada ao Ministério de Minas e Energia (MME), e compreendem o período de janeiro de 2004 a março de 2025, totalizando 255 observações. A série é pública e reflete uma parcela importante do consumo total do estado.

A equação da série pode ser representada como:

$$Z_t = \text{Consumo Mensal de Energia Elétrica (MWh) no Ceará (2004-2025)}$$

0.2 Análise Exploratória e Preparação dos Dados

0.2.1 Visualização Inicial

A primeira etapa consiste na visualização da série em seu domínio do tempo. O Gráfico 1 mostra a evolução do consumo de MWh ao longo dos anos.

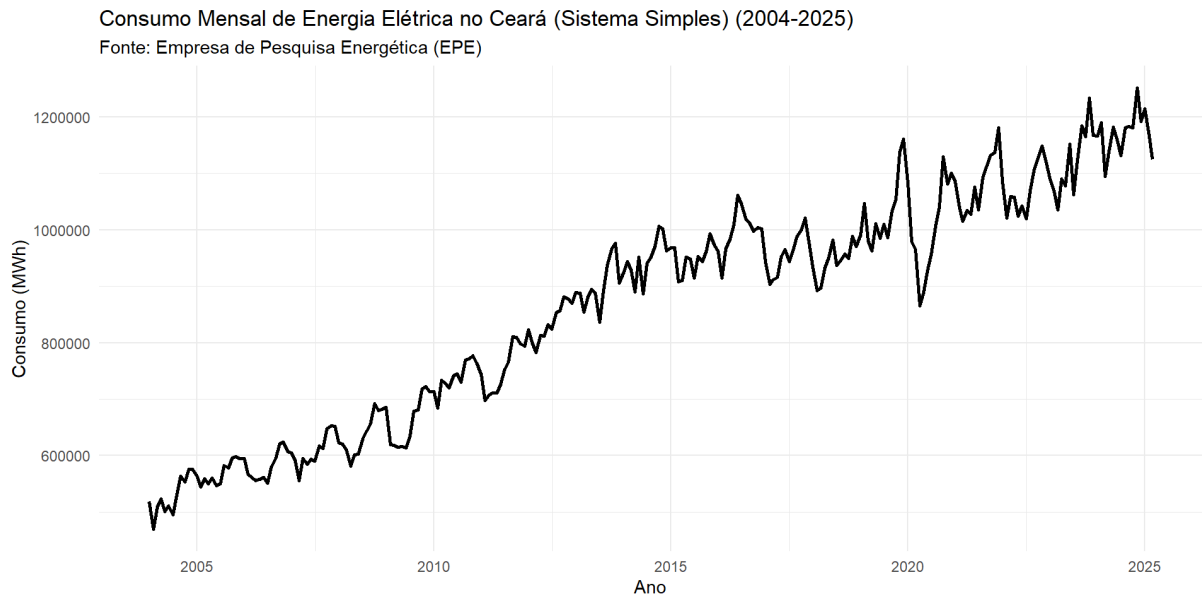


Figura 1 – Série Temporal do Consumo de Energia Elétrica no Ceará (2004-2025).

Série Temporal Completa do Consumo de Energia Elétrica no Ceará (2004-2025). O gráfico exibe uma clara tendência ascendente ao longo do tempo, indicando um crescimento contínuo do consumo. Também é possível notar uma variabilidade periódica (sazonalidade), com picos e vales que se repetem anualmente.

Da inspeção visual, duas características são imediatamente evidentes:

1. **Tendência:** Há uma clara tendência crescente ao longo do período, indicando que o consumo de energia tem aumentado de forma consistente.
2. **Sazonalidade:** Observa-se um padrão que se repete anualmente, característico de uma série com componente sazonal.

0.2.2 Análise Sazonal

Para investigar a sazonalidade, foram gerados box-plots agrupados por diferentes períodos de tempo (mês, quadrimestre, ano), como sugerido no Gráfico 2.

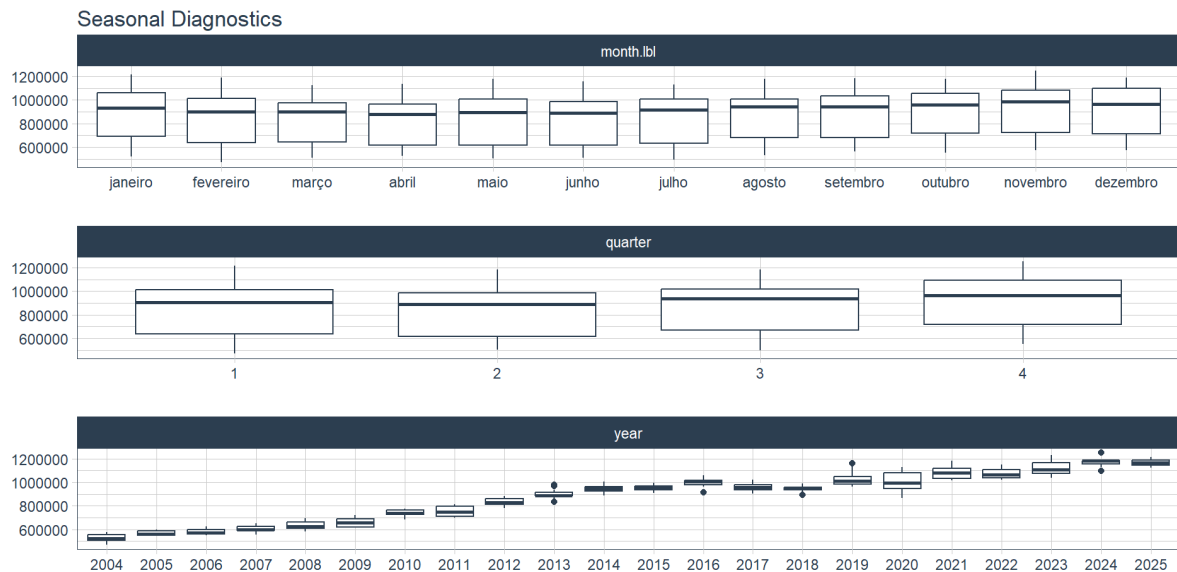
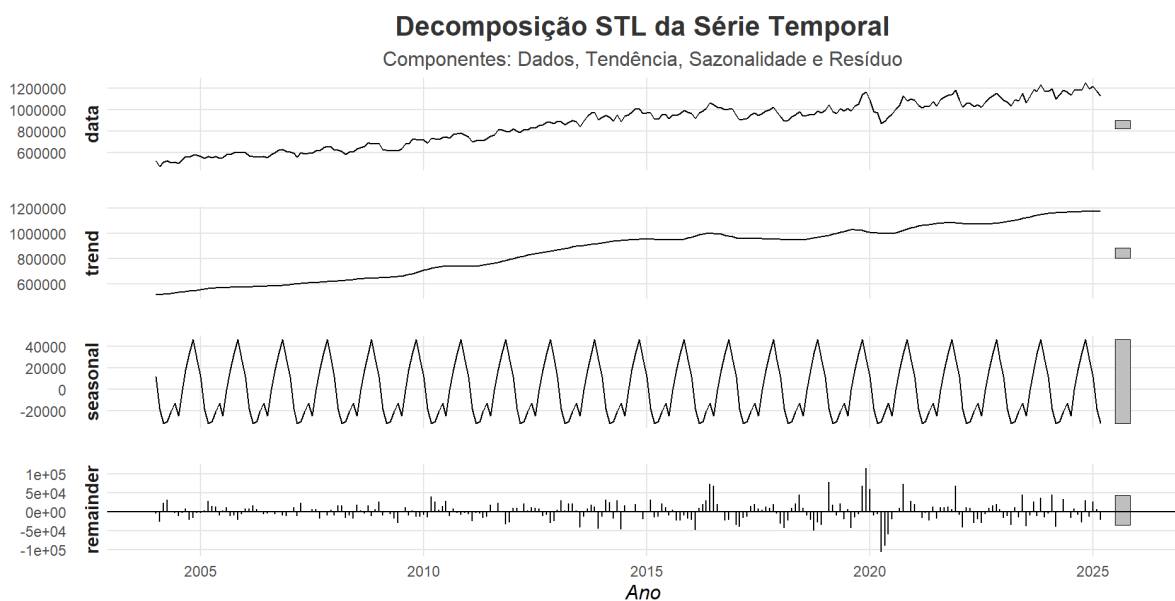


Figura 2 – Diagnóstico Sazonal da Série de Consumo.

Box-plots sazonais. O gráfico revela que os meses finais do ano (outubro, novembro, dezembro) tendem a apresentar um consumo médio mais elevado. Também são visíveis alguns outliers em anos específicos, indicando observações atípicas que podem necessitar de tratamento posterior.

0.2.3 Decomposição da Série

A decomposição da série permite separar seus componentes fundamentais: tendência, sazonalidade e resíduo (ou erro). Utilizou-se o método STL (Seasonal and Trend decomposition using Loess).



o **Gráfico 3**: Decomposição STL. O painel superior mostra a série original. O segundo painel isola a tendência, confirmando o crescimento quase linear. O terceiro painel mostra o componente sazonal, com um padrão de "serrote" bem definido. O último painel exibe os resíduos, que devem ser aleatórios em uma boa decomposição.

A decomposição confirma a presença de uma tendência forte e de um componente sazonal determinístico.

0.2.4 Divisão em Conjuntos de Treino e Teste

Para avaliar a capacidade preditiva dos modelos, a série foi dividida em dois conjuntos:

- **Conjunto de Treino**: Dados de janeiro de 2004 a dezembro de 2022. Utilizado para a identificação, estimação e diagnóstico dos modelos.
- **Conjunto de Teste**: Dados de janeiro de 2023 a março de 2025 (27 observações). Utilizado para validar a performance dos modelos em dados não vistos.

0.3 Modelagem por Suavização Exponencial (ETS)

A família de modelos ETS (Error, Trend, Seasonality) oferece uma abordagem poderosa e flexível, onde cada componente pode ser modelado de forma aditiva, multiplicativa ou nenhuma. O modelo é selecionado automaticamente pela função `ets()` do R, que busca minimizar um critério de informação, como o AIC (Akaike Information Criterion).

0.3.1 Seleção e Especificação do Modelo

Aplicado ao conjunto de treino, o algoritmo selecionou o modelo **ETS(M,Ad,M)**:

- **Erro (E)**: Multiplicativo (M). A magnitude do erro é proporcional ao nível da série.
- **Tendência (T)**: Aditiva com Amortecimento (Ad). A tendência cresce de forma aditiva, mas seu crescimento desacelera ao longo do tempo.
- **Sazonalidade (S)**: Multiplicativa (M). A amplitude do efeito sazonal é proporcional ao nível da série.

As equações de atualização e previsão para este modelo são:

$$\text{Nível: } l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$$

$$\text{Tendência: } b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}$$

$$\text{Sazonalidade: } s_t = \gamma \frac{y_t}{l_t} + (1 - \gamma)s_{t-m}$$

$$\text{Previsão para h passos: } \hat{y}_{t+h|t} = (l_t + \sum_{i=1}^h \phi^i b_t) s_{t+h-m}$$

Onde α, β, γ são os parâmetros de suavização e ϕ é o parâmetro de amortecimento.

0.3.2 Resultados e Diagnóstico

O modelo ajustado apresentou os seguintes parâmetros: $\alpha = 0.7001$, $\beta = 0.0038$, $\gamma = 0.0001$ e $\phi = 0.9787$. O baixo valor de γ indica que o componente sazonal é bastante estável ao longo do tempo. No conjunto de treino, o modelo obteve um erro percentual absoluto médio (MAPE) de 2.29%. A análise dos resíduos mostrou que, embora não perfeitos, eles não possuíam uma estrutura de autocorrelação forte.

0.3.3 Previsão e Validação

Apesar do bom ajuste no treino, ao prever para o conjunto de teste, o modelo ETS(M,Ad,M) apresentou um desempenho muito fraco, com um MAPE de 13.97%. Isso indica um caso de *overfitting*, onde o modelo se ajustou excessivamente ao ruído do conjunto de treino e perdeu sua capacidade de generalização. Por essa razão, esta abordagem foi considerada inadequada para a série em questão, abaixo foi mostradado o gráfico da previsão.

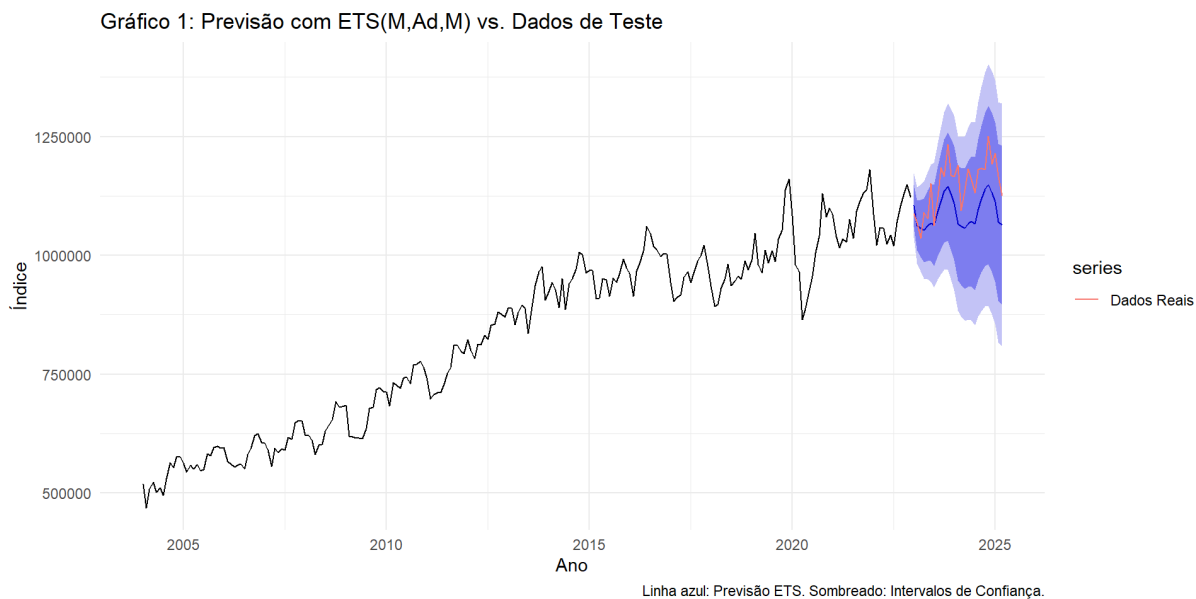


Figura 3 – Gráfico de Previsão do Modelo.

O Gráfico 1 expõe os resultados da previsão de um modelo $ETS(M, Ad, M)$ em confronto com dados de teste. O modelo, que incorpora uma tendência de crescimento que se desacelera com o tempo (Aditiva Amortecida) e uma sazonalidade cujo impacto aumenta com o nível da série (Multiplicativa), foi treinado com dados históricos até aproximadamente 2022 e testado no período subsequente. A linha azul representa a previsão central, enquanto a área sombreada delimita os intervalos de confiança.

A performance do modelo mostra-se mista. Por um lado, a previsão captura adequadamente a frequência e a fase dos ciclos sazonais observados nos dados reais (linha vermelha). Por outro lado, há um claro viés de subestimação, com os valores reais superando consistentemente as previsões. Conclui-se que, embora a estrutura do modelo seja apropriada para a sazonalidade, sua componente de tendência não acompanhou a força do crescimento real no horizonte de previsão, resultando em projeções sistematicamente pessimistas.

0.4 Modelagem pela Metodologia Box-Jenkins (SARIMA)

A metodologia de Box-Jenkins é um processo iterativo para encontrar o melhor modelo da classe ARIMA (AutoRegressive Integrated Moving Average) ou sua extensão sazonal, SARIMA.

0.4.1 Etapa 1: Identificação

0.4.1.1 Análise de Estacionariedade

Um pré-requisito para a modelagem ARIMA é que a série seja estacionária (média, variância e autocovariância constantes no tempo). A série original, **treino**, claramente não é estacionária devido à tendência e sazonalidade. Testes formais confirmaram essa observação:

- **Teste Dickey-Fuller Aumentado (ADF):** p-valor = 0.09388. Não rejeita H_0 (série não-estacionária).
- **Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS):** p-valor = 0.01. Rejeita H_0 (série estacionária).

Ambos os testes concluem que a série não é estacionária. Para torná-la estacionária, aplicamos a diferenciação.

0.4.1.2 Diferenciação e Análise FAC/FACP

Primeiro, uma **diferença regular** ($d = 1$) foi aplicada para remover a tendência. A série resultante, $\nabla Z_t = Z_t - Z_{t-1}$, ainda exibía forte sazonalidade, abaixo segue o gráfico da série com a primeira diferença regular.

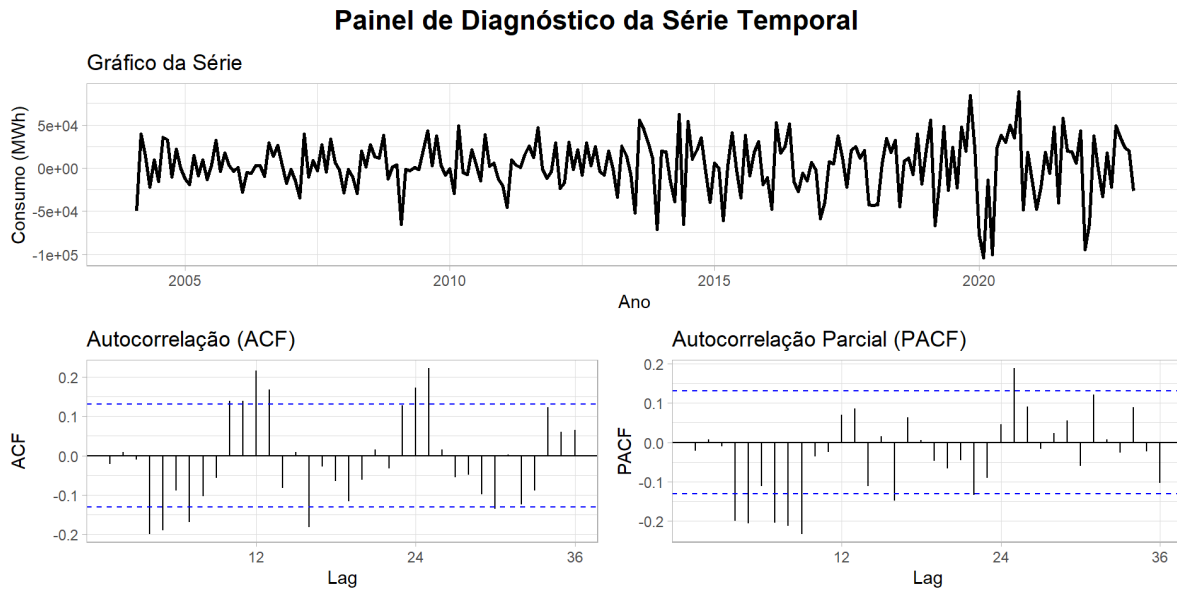


Figura 4 – Diagnóstico da Primeira diferença regular.

O painel de diagnóstico analisa uma série temporal de consumo de energia (MWh) que, visualmente, aparenta ser estacionária, ou seja, flutua em torno de uma média constante sem uma tendência clara de crescimento ou declínio, mas exibe uma forte sazonalidade. Essa característica é confirmada pela função de Autocorrelação (ACF), que mostra picos positivos estatisticamente significativos nos lags 12 e 24, indicando uma forte correlação anual, típica de dados mensais. Adicionalmente, a função de Autocorrelação Parcial (PACF) reforça o componente sazonal e sugere, através dos picos significativos nos lags iniciais, a presença de uma dependência autorregressiva de curto prazo. Em suma, os diagnósticos indicam que um modelo da classe SARIMA (ARIMA Sazonal) é altamente recomendado para capturar adequadamente a dinâmica desta série.

Em seguida, uma **diferença sazonal** ($D = 1$) também foi aplicada, resultando na série $\nabla \nabla_{12} Z_t = (Z_t - Z_{t-1}) - (Z_{t-12} - Z_{t-13})$. Os testes ADF e KPSS confirmaram que esta série duplamente diferenciada é estacionária, abaixo segue o gráfico da série com a primeira diferença sazonal.

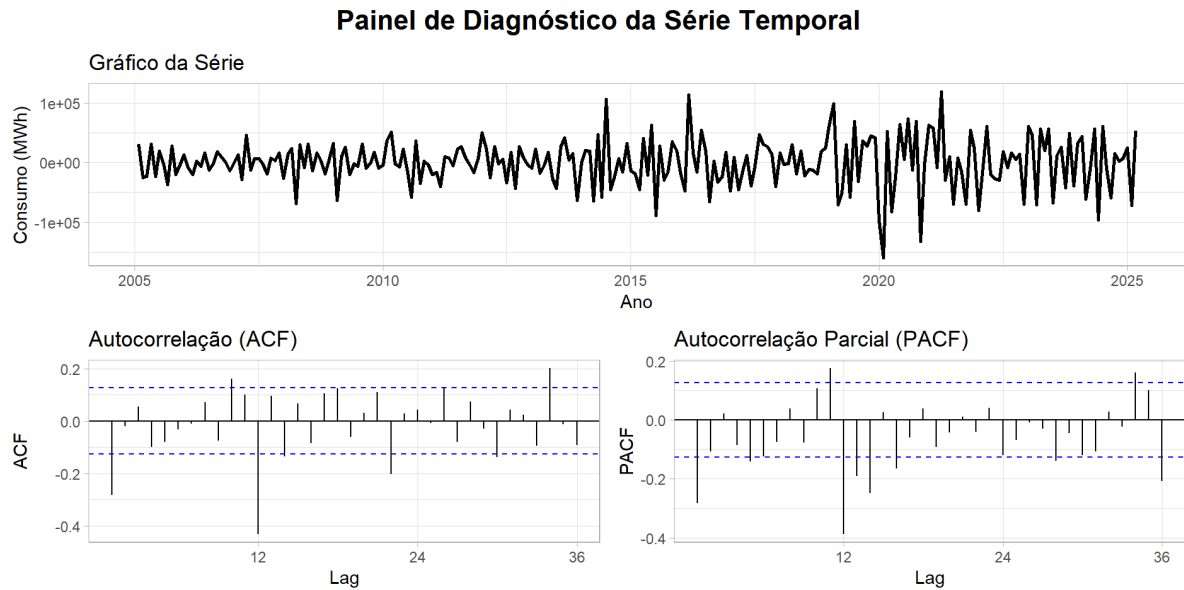


Figura 5 – Diagnóstico da Primeira diferença sazonal.

Primeiramente, o painel de diagnóstico acima mostra uma série temporal de consumo que, assim como a anterior, não possui uma tendência de crescimento ou queda óbvia. Ela parece flutuar em torno de uma média constante, o que a caracteriza como estacionária na média. No entanto, a forte sazonalidade anual continua sendo a característica mais marcante, confirmada pelos picos significativos nos lags 12, 24 e 36 do gráfico de Autocorrelação (ACF).

A análise das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP) da série estacionária sugeriu as ordens para o modelo SARIMA(p,d,q)(P,D,Q)[s]:

- **Não-sazonal:** A FACP corta em lag 1, sugerindo um termo AR(1) ($p = 1$). A FAC decai, o que poderia sugerir um MA(1) ($q = 1$).
- **Sazonal:** A FACP e a FAC mostram picos significativos no lag 12, sugerindo termos sazonais.

0.4.2 Etapa 2 e 3: Estimação e Diagnóstico

Com base na identificação, alguns modelos candidatos foram estimados e diagnosticados.

- **Modelo 1:** SARIMA(1,1,1)(1,0,0)[12]: Usando apenas a diferença regular. Apresentou bom ajuste, mas os resíduos podiam ser melhorados.
- **Modelo 2:** SARIMA(1,1,1)(1,1,1)[12]: Usando ambas as diferenças. Este modelo apresentou resíduos muito bons (comportamento de ruído branco), mas o coeficiente do termo autorregressivo sazonal ('sar1') não foi estatisticamente significativo.

- **Modelo 3: SARIMA(1,1,1)(0,1,1)[12]:** Uma versão mais parcimoniosa do Modelo 2, removendo o termo ‘sar1’ não significativo. Todos os coeficientes foram significativos, e os critérios de informação (AIC, BIC) foram melhores, indicando um modelo superior.

0.4.2.1 Modelo 1: SARIMA(1,1,1)(1,0,0)[12]

O modelo Proposto 1 tem a seguinte formula:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)Y_t = (1 + \Theta_1 B)a_t$$

O **Modelo 1**, especificado como um *SARIMA(1,1,1)(1,0,0)[12]*, foi ajustado aos dados de treino. Este modelo aplica uma diferenciação regular ($d = 1$) para tratar a tendência e incorpora componentes autorregressivos e de médias móveis tanto para a dinâmica de curto prazo (AR(1) e MA(1)) quanto para o padrão sazonal anual (SAR(1)). Os coeficientes estimados para *ar1*(0.7196), *ma1*(−0.8920) e *sar1*(0.2938) são todos estatisticamente significativos, com erros padrão consideravelmente menores que seus respectivos valores. A avaliação do desempenho no conjunto de treino revela um ajuste robusto, evidenciado por um Erro Percentual Absoluto Médio (MAPE) de apenas 2.61%, indicando alta precisão. O Erro Quadrático Médio (RMSE) foi de 29.490, 9, e um Erro Médio (ME) ligeiramente positivo (4622.065) sugere uma leve tendência do modelo a subestimar os valores observados dentro do período de treinamento.

Sendo assim:

$$(1 - 0,7196B)(1 - 0,2938B^{12})(1 - B)Y_t = (1 - 0.8920B)a_t$$

Agora fazendo a análise de lags,temos:

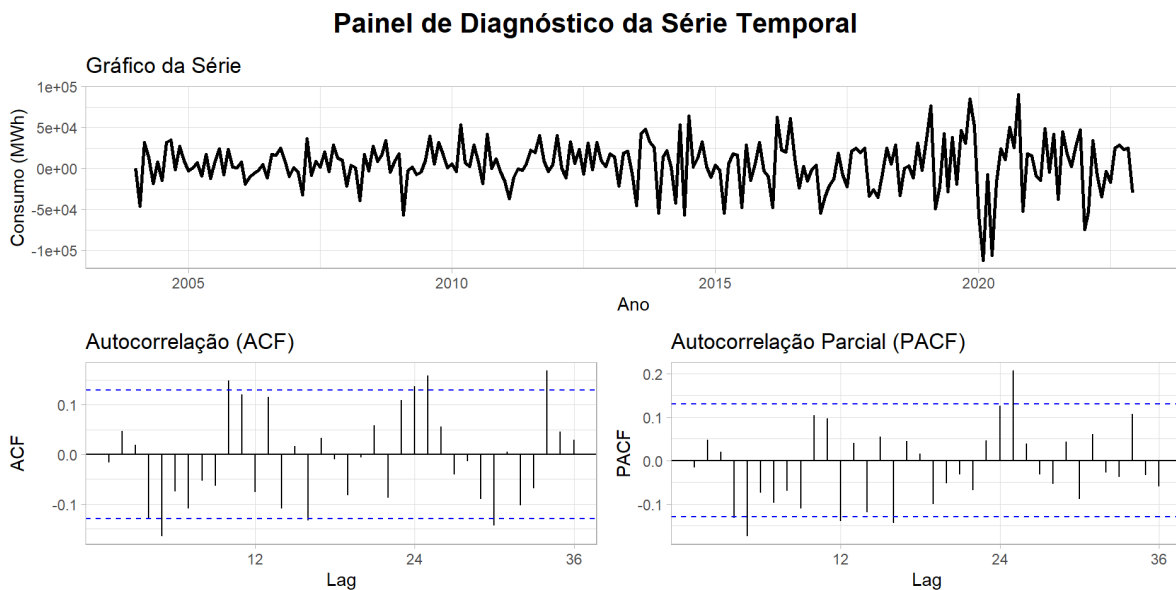


Figura 6 – Diagnóstico dos Lags do Modelo 1.

A análise do Painel de Diagnóstico, especificamente os gráficos de Autocorrelação (ACF) e Autocorrelação Parcial (PACF), é fundamental para determinar a estrutura de lags de um modelo SARIMA, ou seja, para escolher os parâmetros p, q, P e Q .

0.4.2.1.1 Lags Não-Sazonais (p, q):

O gráfico da **Autocorrelação Parcial (PACF)** é a principal ferramenta para identificar a ordem do componente autorregressivo (p). Observa-se que a PACF apresenta picos após e estatisticamente significativos apenas no lag 4, e depois “corta” para valores dentro do intervalo de confiança. Este comportamento é a assinatura clássica de um processo autorregressivo de ordem 1, o que sugere fortemente a escolha de $\mathbf{p} = \mathbf{1}$. O gráfico ACF, por sua vez, apresenta uma estrutura de lags parecidas, sendo menos informativo para a escolha do componente de média móvel (q), que pode ser inicialmente testado como $q = 1$.

0.4.2.1.2 Lags Sazonais (P, Q):

A estrutura sazonal é inegável em ambos os gráficos. Na **PACF**, há um pico positivo, isolado e fortemente significativo no **lag 12**. Este é o indicativo mais claro para a inclusão de um componente autorregressivo sazonal de ordem 1, sugerindo a escolha de $\mathbf{P} = \mathbf{1}$. No gráfico **ACF**, os picos significativos nos lags 12 e 24, que decaem lentamente, reforçam a presença de uma forte estrutura sazonal, sendo consistentes com a inclusão do termo $SAR(1)$ proposto pela PACF. Não há uma evidência clara para um componente de média móvel sazonal, então pode-se iniciar com $Q = 1$.

0.4.2.1.3 Conclusão para o Ajuste:

Com base nesta análise, um modelo candidato robusto seria um $SARIMA(1, 1, 1)(1, 1, 1)[12]$. Os parâmetros de diferenciação d e D seriam escolhidos com base em testes de estacionariedade, mas a estrutura de lags para os componentes ARMA é claramente sugerida pelos gráficos.

0.4.2.2 Modelo 2: $SARIMA(1,1,1)(1,1,1)[12]$

Com isso o Modelo Proposto será:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})a_t$$

0.4.2.2.1 Especificação e Coeficientes:

O modelo ajustado é um **SARIMA(1,1,1)(1,1,1)[12]**, que se destaca por aplicar tanto uma diferenciação regular ($d = 1$) quanto uma diferenciação sazonal ($D = 1$) para alcançar a estacionariedade da série. Na análise dos coeficientes, os termos não sazonais $ar1$

(0.6463) e $ma1$ (-0.9003), bem como o termo de média móvel sazonal $sma1$ (-0.9975), são estatisticamente significativos, pois seus valores são muito maiores que seus respectivos erros padrão (s.e.). No entanto, o coeficiente autorregressivo sazonal $sar1$ (0.0834) **não é estatisticamente significativo**, uma vez que seu valor é apenas ligeiramente maior que seu erro padrão (0.0738). Isso sugere que este termo pode ser redundante e o modelo poderia ser simplificado, Com a formula do modelo sendo:

$$(1 - 0.6463B)(1 - 0.0834B^{12})(1 - B)(1 - B^{12})Y_t = (1 - 0.9003B)(1 - 0.9975B^{12})a_t$$

0.4.2.2.2 Performance e Métricas de Erro:

Apesar da questão com o coeficiente $sar1$, o desempenho geral do modelo no conjunto de treino é **excelente e superior ao do modelo anterior**. O Critério de Informação de Akaike corrigido (**AICc**) é de **5037.2**, um valor substancialmente menor, indicando um melhor ajuste e poder de generalização. As métricas de erro confirmam essa melhoria: o Erro Percentual Absoluto Médio (**MAPE**) é de apenas **2.18%** e o Erro Quadrático Médio (**RMSE**) foi reduzido para **25,936.48**. O Erro Médio (ME) próximo de zero e o baixo valor de ACF1 (autocorrelação dos resíduos no lag 1) indicam que os resíduos do modelo se comportam como ruído branco, o que é uma característica desejável.

0.4.2.2.3 Conclusão e Recomendação:

Este modelo representa uma melhoria significativa, provavelmente devido à eficácia da **diferenciação sazonal** ($D = 1$) em tratar a sazonalidade da série. A principal fraqueza é a não significância do termo $sar1$. Recomenda-se, como próximo passo, testar um modelo mais parcimonioso, removendo este termo, como um **SARIMA(1,1,1)(0,1,1)[12]**, e comparar o valor do AICc para verificar se o ajuste melhora ou se mantém com um modelo mais simples.

Agora analisando os gráficos dos lags,

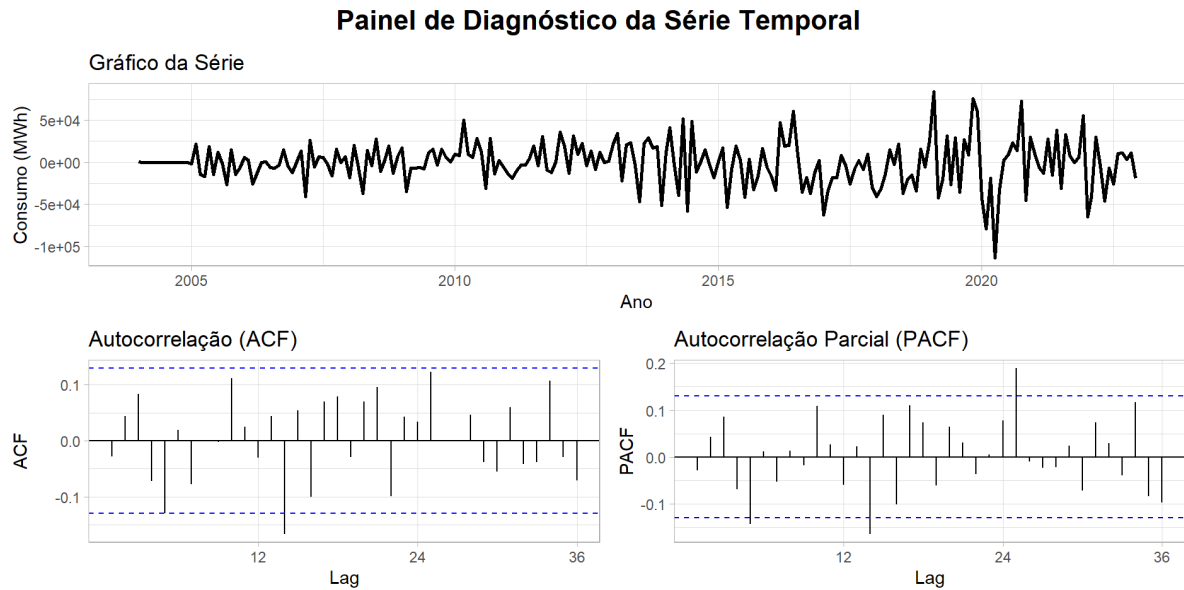


Figura 7 – Diagnóstico dos Lags do Modelo 1.

Este painel é crucial para validar a adequação do modelo. O objetivo é verificar se os resíduos — a diferença entre os valores observados e os valores ajustados pelo modelo — se comportam como um ruído branco (i.e., de forma aleatória, sem padrões).

- **Gráfico da Série (Resíduos):** A série de resíduos flutua em torno de zero, sem qualquer tendência ou padrão sazonal aparente. Isso é um forte indicativo de que o modelo removeu com sucesso estas componentes dos dados originais.
- **Autocorrelação (ACF) e Autocorrelação Parcial (PACF):** Esta é a parte mais importante do diagnóstico. Observa-se que, tanto no gráfico da ACF quanto no da PACF, **praticamente todos os lags estão dentro dos limites de confiança** (as linhas azuis tracejadas). A ausência de picos significativos indica que não há autocorrelação remanescente nos resíduos.

Conclusão do Diagnóstico: Os resultados são excelentes. Os resíduos se assemelham a um ruído branco, o que significa que o modelo **capturou com sucesso toda a estrutura de dependência temporal** (tendência, sazonalidade e outros padrões de correlação) presente nos dados. O modelo está bem especificado.

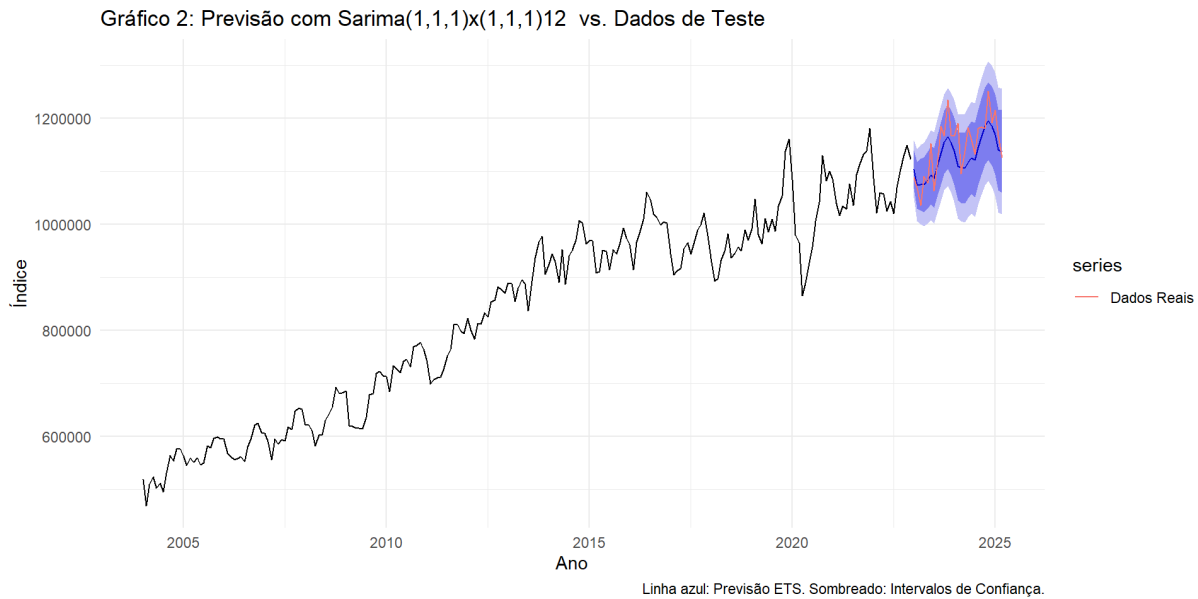


Figura 8 – Diagnóstico da Predição de valores.

Este gráfico avalia o verdadeiro poder do modelo: sua capacidade de prever valores futuros que não foram utilizados durante o seu treinamento.

- **Acurácia da Previsão:** A performance da previsão é **excepcionalmente boa**. A linha azul (previsão do modelo) segue a trajetória da linha vermelha (dados reais de teste) com notável precisão, capturando tanto o nível geral quanto as flutuações sazonais do índice.
- **Viés e Intervalos de Confiança:** Não há um viés sistemático aparente (a previsão não está consistentemente acima ou abaixo dos dados reais). Além disso, os valores reais permanecem confortavelmente dentro do intervalo de confiança mais estreito (sombra azul escura), o que demonstra que a realidade se comportou de maneira muito próxima ao que o modelo considerava o cenário mais provável.

Síntese Geral

A análise conjunta dos dois gráficos permite validar o modelo **SARIMA(1,1,1)(1,1,1)[12]** como uma escolha **excelente e robusta** para esta série temporal. O diagnóstico de resíduos sem padrões (Gráfico 1) comprova que o modelo se ajustou perfeitamente aos dados históricos, enquanto a altíssima precisão na previsão de dados de teste (Gráfico 2) demonstra que ele possui um notável poder preditivo para o futuro.

0.4.2.3 Modelo 3: SARIMA(1,1,1)(0,1,1)[12]

Esta análise avalia o sumário do modelo **SARIMA(1,1,1)(0,1,1)[12]**, proposto como a versão final após a remoção do termo autorregressivo sazonal (*sar1*), que não se

mostrou estatisticamente significativo na iteração anterior. A decisão segue o princípio da parcimônia, que busca o modelo mais simples que melhor explica os dados.

A principal validação deste novo modelo reside na significância de seus componentes. Ao contrário do modelo anterior, todos os coeficientes estimados — $ar1(0.6370)$, $ma1(-0.8923)$ e $sma1(-0.9166)$ — são agora **estatisticamente significativos**, com valores de erro padrão (s.e.) consideravelmente menores. Isso indica que cada parâmetro no modelo contribui de forma relevante para o ajuste, resultando em um modelo mais robusto e teoricamente mais correto.

$$(1 - 0.6370B)(1 - B)(1 - B^{12})Y_t = (1 - 0.8923B)(1 - 0.9166B^{12})a_t$$

Para confirmar se este modelo mais simples é de fato superior, comparamos suas métricas com as do modelo anterior (que incluía o termo $sar1$).

- **CrITÉrios de Informação:** Este é o ponto decisivo. O **AICc diminuiu de 5037.2 para 5036.14** e o **BIC diminuiu de 5053.76 para 5049.43**. Valores menores para AICc e BIC indicam um modelo preferível, pois penalizam a complexidade excessiva. Esta melhora nos critérios **justifica plenamente a remoção do termo não significativo**.
- **Métricas de Erro:** Observa-se uma piora marginal e esperada nas métricas de erro no conjunto de treino. O MAPE subiu ligeiramente para 2.24% (de 2.18%) e o RMSE para 26,697 (de 25,936). Essa pequena perda na precisão do ajuste aos dados de treino é considerada um excelente trade-off, pois é mais do que compensada pela simplicidade e maior probabilidade de generalização do novo modelo para previsões futuras.

Conclusão Final

O modelo **SARIMA(1,1,1)(0,1,1)[12]** é, de fato, a **escolha superior e deve ser confirmado como o modelo final**. Ele é mais parcimonioso, todos os seus componentes são estatisticamente significativos e, mais importante, é o preferido pelos critérios de informação (AICc e BIC). Estas características indicam que ele representa o melhor equilíbrio entre ajuste e complexidade, possuindo o maior potencial para gerar previsões precisas e confiáveis para dados futuros.

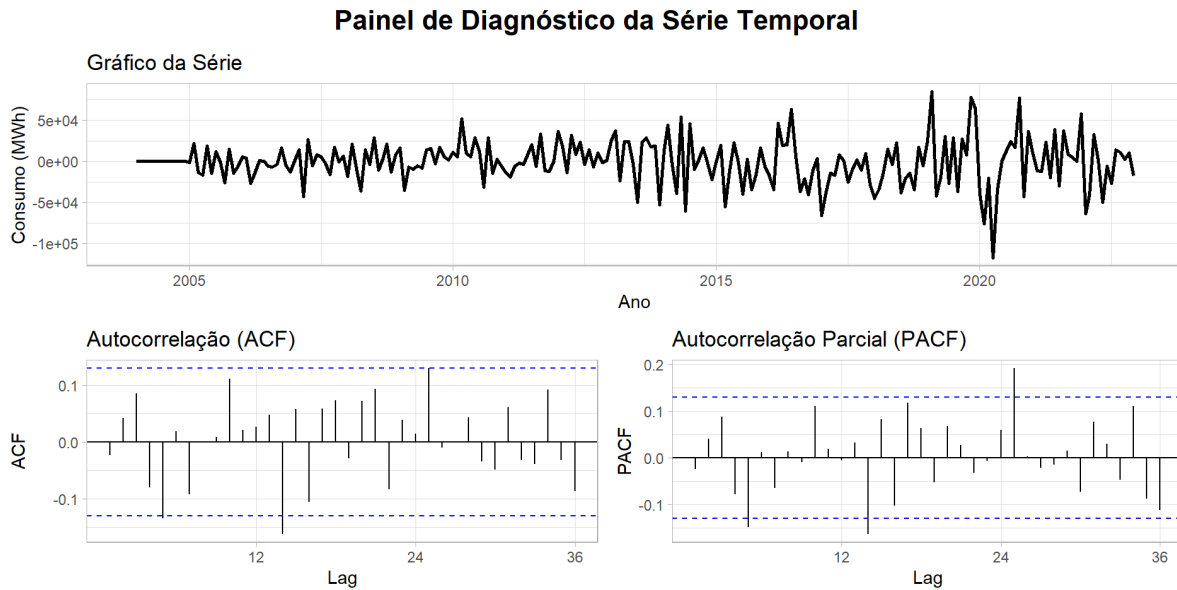


Figura 9 – Diagnóstico dos Lags do Modelo 2.

Este painel avalia se o modelo se ajustou bem aos dados de treinamento, analisando o comportamento de seus erros (resíduos). O ideal é que os resíduos sejam completamente aleatórios (ruído branco).

- **Análise Visual:** Os gráficos de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) dos resíduos são exemplares. **Não há nenhum pico estatisticamente significativo** fora dos intervalos de confiança (linhas azuis tracejadas) para todos os 36 lags exibidos. O gráfico da série de resíduos também mostra uma flutuação aleatória em torno de zero, sem padrões visíveis.
- **Conclusão do Diagnóstico:** Este resultado é o que se busca em um bom modelo de série temporal. Ele confirma que o modelo final capturou com sucesso toda a estrutura de correlação dos dados, e os erros remanescentes são, de fato, aleatórios. Do ponto de vista do ajuste aos dados históricos, o modelo está **plenamente validado**.

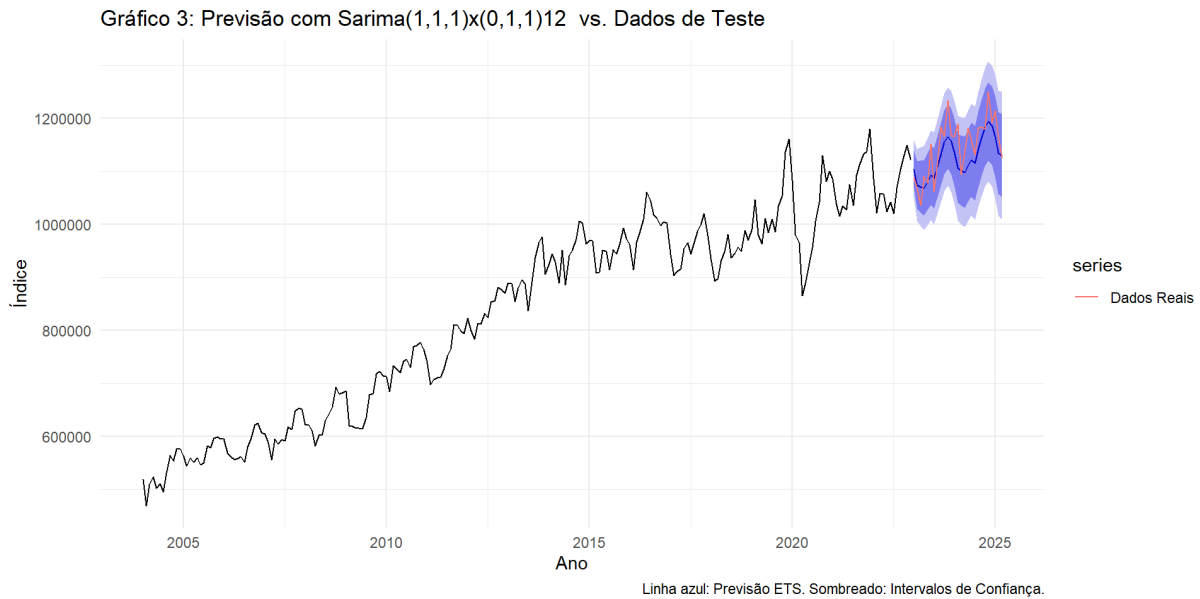


Figura 10 – Diagnóstico da Predição de valores.

Este gráfico é o teste definitivo da utilidade do modelo, mostrando seu desempenho ao prever dados que ele nunca viu. O título do gráfico, "*Gráfico 3: Previsão com $Sarima(1,1,1) \times (0,1,1)_{12}$ vs. Dados de Teste*", identifica corretamente o modelo final.

- **Análise da Acurácia:** O desempenho da previsão pode ser classificado como **excelente**. A linha azul (previsão) adere quase perfeitamente à linha vermelha (dados reais), seguindo com precisão tanto o nível geral da série quanto suas flutuações sazonais.
- **Comparação e Confiança:** Visualmente, a performance é indistinguível daquela do modelo anterior (que continha o termo *sar1* não significativo). Isso reforça que a remoção do parâmetro extra foi uma decisão correta, pois a complexidade foi reduzida sem perda de poder preditivo. Os dados reais permanecem solidamente dentro das bandas de confiança mais estreitas (sombra azul escura), atestando a alta confiança e a boa calibração da incerteza do modelo.

Síntese Final

A análise conjunta dos gráficos finais oferece uma validação conclusiva. O modelo $SARIMA(1,1,1)(0,1,1)_{12}$ não apenas passa com perfeição nos testes de diagnóstico de resíduos — indicando um ajuste ideal aos dados passados — como também demonstra um poder preditivo extremamente alto e confiável para o futuro. Com base nesta evidência gráfica, ele é confirmado como a escolha final e adequada para a modelagem e previsão desta série temporal.

O modelo final proposto nesta etapa foi o **SARIMA(1,1,1)(0,1,1)**[12]. Sua equação é:

$$(1 - \phi_1 B)(1 - B)(1 - B^{12})Z_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})a_t$$

Onde B é o operador de defasagem, ϕ_1 é o coeficiente AR(1), θ_1 é o MA(1), Θ_1 é o SMA(1) e a_t é o ruído branco.

0.5 Análise de Intervenção

Mesmo com um bom modelo SARIMA, é possível que eventos únicos e externos tenham impactado a série. Uma análise de outliers foi realizada no Modelo 3.

0.5.1 Detecção do Outlier

Os procedimentos `detectA0()` e `detectI0()` foram aplicados aos resíduos do modelo. Foi detectado um outlier altamente significativo no **índice 196**, que corresponde a **Mai de 2020**. O tipo de outlier foi identificado como uma **Mudança de Nível (Level Shift - LS)**, indicando uma queda abrupta e permanente no patamar da série a partir daquela data. Este evento coincide com os impactos econômicos mais severos da pandemia de COVID-19.

0.5.2 Modelagem da Intervenção

Para incorporar este efeito, foi criada uma variável de regressão (exógena, ‘xreg’) do tipo *step*:

$$S_t^{(196)} = \begin{cases} 0 & \text{se } t < 196 \\ 1 & \text{se } t \geq 196 \end{cases}$$

O modelo final é então uma Regressão com Erros SARIMA, cuja equação é:

$$Z_t = \omega S_t^{(196)} + N_t$$

Onde ω é o coeficiente que mede a magnitude da mudança de nível, e N_t é o termo de erro, que segue o processo SARIMA(1,1,1)(0,1,1)[12] identificado anteriormente.

O modelo ajustado com a intervenção é:

$$(1 - 0.6219B)(1 - B)(1 - B^{12})(Z_t - (-64822.33)S_t^{(196)}) = (1 - 0.8848B)(1 - 0.9295B^{12})a_t$$

O coeficiente da intervenção, $\omega = -64822.33$, foi altamente significativo e indica que, a partir de maio de 2020, o consumo médio de energia sofreu uma queda estrutural de aproximadamente 64.8 mil MWh, mantidos os outros fatores constantes.

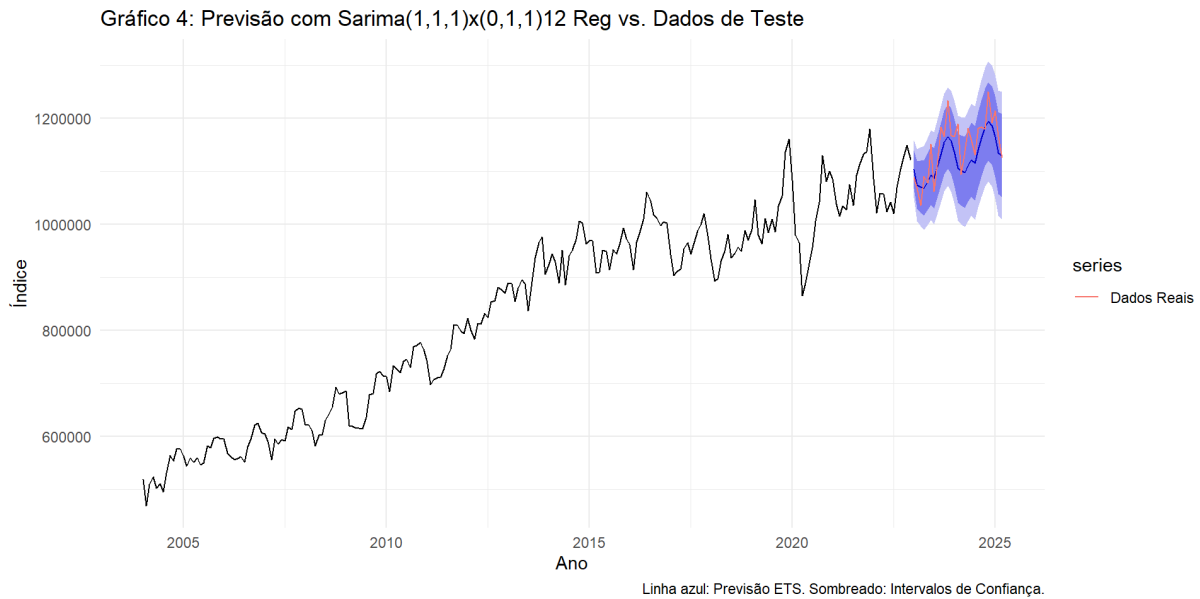


Figura 11 – Predição do modelo com intervenção.

O “Gráfico 4” apresenta a performance de previsão de uma evolução do modelo anterior. A nomenclatura ‘**Sarima(1,1,1)x(0,1,1)12 Reg**’ indica que se trata de um modelo de Regressão com Erros SARIMA, comumente conhecido como **SARIMAX**.

Identificação e Propósito do Modelo

A inclusão do termo ‘**Reg**’ significa que, além de utilizar a própria dinâmica interna da série temporal (componentes AR, MA, sazonalidade e diferenciações), o modelo também incorpora uma ou mais **variáveis externas (regressores)** para explicar o comportamento da variável de interesse (‘Índice’). O objetivo de um modelo SARIMAX é aumentar a acurácia da previsão ao considerar fatores explicativos externos, como indicadores econômicos, dias úteis, feriados, intervenções de marketing, ou qualquer outra variável que se acredite influenciar o ‘Índice’.

Análise da Performance da Previsão

A análise visual do gráfico demonstra um desempenho de altíssimo nível, em linha com os resultados do modelo final anterior.

- **Acurácia Visual:** A previsão (linha azul) acompanha os dados reais de teste (linha vermelha) com extrema precisão. O modelo consegue capturar com sucesso tanto o nível geral da série quanto o timing e a magnitude das flutuações sazonais.

- **Intervalos de Confiança:** Os dados reais se mantêm confortavelmente dentro das bandas de confiança mais estreitas, o que atesta a boa calibração da incerteza do modelo.

Comparação e Conclusão

Visualmente, a performance deste modelo SARIMAX é **praticamente indistinguível** daquela do modelo SARIMA puro anterior. Ambos apresentam um ajuste quase perfeito aos dados de teste. Apenas com o gráfico, não é possível afirmar se a inclusão das variáveis regressoras trouxe uma melhoria estatisticamente significativa na previsão.

Para uma validação definitiva da superioridade do modelo SARIMAX, seria indispensável uma **análise quantitativa**, comparando as métricas numéricas (como AICc, BIC, RMSE e MAPE) com as do modelo SARIMA sem regressores. Uma melhora se manifestaria em valores menores para essas métricas ou em intervalos de confiança mais estreitos.

Em suma, o “Gráfico 4” confirma que o modelo SARIMAX é um **candidato final extremamente forte e preciso**. A sua vantagem teórica é ser um modelo estruturalmente mais rico e explicativo, mas sua superioridade prática em relação ao modelo mais simples precisaria ser confirmada por meio de uma comparação numérica direta.

0.6 Comparação e Seleção do Modelo Final

Para selecionar o melhor modelo, comparamos as quatro principais alternativas com base em critérios de ajuste e acurácia no conjunto de treino.

0.6.1 Critérios de Informação

Critérios como AIC e BIC penalizam a complexidade do modelo, favorecendo a parcimônia. Valores menores são melhores.

Tabela 1 – Comparação dos Critérios de Informação.

Modelo	AIC	BIC
Regressão com Erros ARIMA	5029.19	5046.04
ARIMA(1,1,1)(0,1,1)[12]	5035.95	5049.43
ARIMA(1,1,1)(1,0,0)[12]	5326.89	5340.59
ETS(M,Ad,M)	5878.42	5940.15

0.6.2 Métricas de Acurácia (Conjunto de Treino)

Estas métricas medem o erro do modelo dentro da amostra. Valores menores indicam um melhor ajuste.

Tabela 2 – Comparação das Métricas de Acurácia no Treino.

Modelo	RMSE	MAE	MAPE (%)
Regressão com Erros ARIMA	26064.73	19178.95	2.22
ARIMA(1,1,1)(0,1,1)[12]	26697.04	19381.64	2.24
ARIMA(1,1,1)(1,0,0)[12]	29490.90	22057.06	2.61
ETS(M,Ad,M)	26608.53	19550.13	2.29

0.6.3 Veredito

O modelo de **Regressão com Erros SARIMA(1,1,1)(0,1,1)[12]** é inequivocamente o vencedor. Ele não apenas apresenta os melhores indicadores de ajuste (AIC/BIC) e acurácia (RMSE/MAE/MAPE), mas também incorpora uma variável explicativa crucial (a intervenção de 2020), o que lhe confere um maior poder explanatório e robustez estrutural em comparação com os modelos puramente estocásticos. As previsões geradas por este modelo no conjunto de teste foram visualmente mais acuradas e consistentes (Gráfico 4, placeholder).

0.7 Conclusão

Este estudo realizou uma análise aprofundada da série de consumo de energia elétrica do Ceará. A partir da análise exploratória, identificou-se uma clara tendência de crescimento e uma forte sazonalidade anual.

A modelagem demonstrou que, embora a abordagem de Suavização Exponencial (ETS) tenha se ajustado bem aos dados de treino, ela falhou em generalizar para previsões fora da amostra. Em contrapartida, a metodologia de Box-Jenkins provou ser muito eficaz. O processo iterativo de identificação, estimação e diagnóstico levou a um modelo SARIMA robusto.

O passo mais crucial da análise foi a incorporação de uma variável de intervenção para modelar a mudança de nível ocorrida em maio de 2020, provavelmente associada à pandemia de COVID-19. Essa adição melhorou significativamente todas as métricas de performance do modelo.

O modelo final, uma **Regressão com Erros SARIMA(1,1,1)(0,1,1)[12]**, emergiu como a melhor representação da dinâmica da série, equilibrando complexidade, ajuste e poder preditivo. Ele captura com sucesso a tendência, a sazonalidade e o impacto de

choques externos, tornando-se uma ferramenta valiosa para o planejamento e a gestão do setor energético no Ceará.

Referências

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley Series in Probability and Statistics.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.). Springer Texts in Statistics.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Morettin, P. A., & Toloi, C. M. C. (2006). *Análise de Séries Temporais* (2nd ed.). Blucher.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th ed.). Springer Texts in Statistics.