

Análise do Consumo Elétrico de Uma Residência

João Kramer ¹

Raul Matsushita ²

Departamento de Estatística, Universidade de Brasília, DF

Resumo. OBJETIVO: avaliar a capacidade preditiva de um modelo univariado SARIMA para representar e antecipar os padrões de consumo elétrico de uma residência, com base exclusivamente na estrutura temporal da série, sem considerar variáveis explicativas adicionais como número de moradores, potência instalada ou tamanho do imóvel.

MÉTODO: a série mensal do consumo médio diário, medida em kWh/dia, foi transformada por diferenciação regular e analisada com base nas funções de autocorrelação (FAC) e autocorrelação parcial (FACP). A seleção do modelo SARIMA foi conduzida por busca exaustiva em uma grade de ordens reduzidas, com base no critério bayesiano de informação (BIC). O modelo escolhido, SARIMA(0,1,1)(1,0,1)₁₂, foi ajustado por máxima verossimilhança, e sua adequação foi avaliada por meio da análise dos resíduos (FAC, FACP, teste de Ljung-Box e teste de Shapiro-Wilk). A capacidade preditiva foi mensurada com a raiz quadrada da média dos quadrados dos erros (RMSE) no conjunto de teste.

RESULTADOS: o modelo selecionado apresentou resíduos não autocorrelacionados e um valor de RMSE de aproximadamente 5,123 no horizonte de 12 meses à frente, com previsões que acompanharam adequadamente a variação sazonal e o nível da série real, mesmo na ausência de variáveis exógenas.

CONCLUSÃO: o modelo SARIMA univariado demonstrou ser uma ferramenta estatística eficaz para previsão de consumo elétrico residencial em curto prazo. Seus bons resultados, mesmo em um cenário simplificado, indicam potencial para aplicações práticas de monitoramento e planejamento, com possibilidades futuras de extensão para modelos multivariados mais explicativos.

Palavras-chave. Modelo SARIMA. Filtros Lineares. Consumo Elétrico Residencial. ACF/PACF, RMSE. Comparação BIC/AIC.

1 Introdução

O estudo do consumo elétrico residencial tem ganhado destaque em um contexto de crescente preocupação com eficiência energética, sustentabilidade e gestão inteligente de recursos. A análise de séries temporais que representam o uso de energia ao longo do tempo permite compreender o comportamento do consumidor, identificar padrões e sazonalidades, detectar anomalias e avaliar possíveis impactos de fatores externos. Tais informações são valiosas para o monitoramento contínuo do uso de energia, para o planejamento de políticas de consumo consciente, para a formulação de estratégias de tarifação mais justas e para o dimensionamento adequado de sistemas de geração e armazenamento, especialmente em contextos de energia distribuída, como a solar fotovoltaica.

O avanço das tecnologias de medição e armazenamento de dados — como os medidores inteligentes e sistemas baseados em Internet das Coisas (IoT) — possibilitou o acompanhamento em alta resolução do consumo elétrico residencial, fornecendo dados em tempo real e com alto grau

¹seu e-mail

²raulmta@unb.br

de detalhamento. Quando devidamente analisados, esses dados revelam padrões de uso típicos, ajudam a diagnosticar desperdícios e orientam medidas de eficiência energética personalizadas ao perfil de cada residência.

Embora se reconheça que o consumo de energia elétrica pode ser influenciado por diversas variáveis explicativas — como o tamanho do imóvel, a potência instalada dos equipamentos, o número de moradores e condições climáticas —, este estudo se propõe a realizar uma análise univariada, ou seja, baseada exclusivamente na estrutura estatística da série temporal de consumo, sem recorrer a informações externas. Essa abordagem, embora simplificada, ainda é capaz de oferecer importantes *insights* sobre a dinâmica do consumo ao longo do tempo.

Este relatório tem como objetivo principal realizar uma análise de séries temporais do consumo elétrico de uma residência, no período de junho de 1997 a maio de 2025. A análise visa compreender as características dinâmicas dessa série, com especial atenção para tendências, sazonalidades e flutuações aleatórias. Para tanto, será empregado um modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*), cuja estrutura é particularmente adequada para séries com componentes sazonais e dependência temporal (Morettin e Toloi, 2005; Cryer e Chan, 2008; Cowpertwait e Metcalfe, 2009; Shumway e Stoffer, 2017). Os resultados obtidos podem subsidiar ações de conscientização, planejamento e otimização do uso da energia elétrica em ambientes residenciais.

A Seção 2 apresenta brevemente a metodologia empregada neste trabalho, descrevendo os dados coletados (Seção 2.1) e o modelo de SARIMA (Seção 2.2). A análise é feita na Seção 3 e a Seção 4 conclui o estudo.

2 Metodologia

2.1 Dados

Os dados foram organizados com base nas informações disponíveis na fatura mensal de uma residência, registrando-se as datas da leitura atual e anterior, o número de dias (entre a leitura atual e a anterior), e o valor do consumo apurado no mês de referência, em kWh, que se encontram parcialmente ilustradas na Tabela 1. A série temporal abrange o período de junho de 1997 a maio de 2025, perfazendo 28 anos de observações.

Tabela 1: Exemplo de registro mensal de consumo elétrico e informações associadas.

Mês/Ano	Consumo (kWh)	Dias faturados	Data leitura atual	Data leitura anterior
jan/18	644	32	09/01/2018	08/12/2017
fev/18	519	29	07/02/2018	09/01/2018
mar/18	545	29	08/03/2018	07/02/2018
abr/18	571	32	09/04/2018	08/03/2018
mai/18	570	30	09/05/2018	09/04/2018
jun/18	564	30	08/06/2018	09/05/2018
jul/18	705	31	09/07/2018	08/06/2018
ago/18	562	30	08/08/2018	09/07/2018
set/18	591	33	10/09/2018	08/08/2018
out/18	542	29	09/10/2018	10/09/2018
nov/18	637	31	09/11/2018	09/10/2018
dez/18	689	31	10/12/2018	09/11/2018

Conceitualmente, o consumo elétrico C observado no mês t depende do número D de dias faturados, pois o valor registrado pelo medidor representa a integral do consumo instantâneo ao longo do intervalo de leitura. Seja $X(\tau)$ a taxa instantânea de consumo de energia elétrica em kWh/dia , registrada de forma cumulativa e contínua por um medidor. O consumo total no período entre duas leituras, ou seja, durante os D dias anteriores à data t , pode ser expresso como:

$$C_t(D) = \int_{t-D}^t X(\tau) d\tau. \quad (1)$$

Essa formulação evidencia que o consumo C_t não é apenas uma função do tempo, mas também do intervalo de faturamento D . Como o número de dias faturados varia, é preciso considerar tal variação na análise da série temporal do consumo. Assim, oscilações em C_t podem refletir variações tanto na taxa real de uso ($X(\tau)$), quanto na duração do intervalo de leitura. Observe que Eq. (1) pode ser expressa como uma soma na forma

$$C_t(D) = \int_{t-D}^t X_t(\tau) d\tau = \sum_{j=1}^D \int_{t-D+j-1}^{t-D+j} X_t(\tau) d\tau = \sum_{j=1}^D C_{t_j}(1), \quad (2)$$

na qual $C_{t_j}(1)$ expressa o consumo referente a um dia ($D = 1$) no instante de referência $t_j = t - D + j$. Logo, como critério de normalização dos dados, tem-se que

$$Y_t = C_t(1) = \frac{C_t(D)}{D} = \frac{1}{D} \sum_{j=1}^D C_{t_j}(1) \quad (3)$$

denota o consumo diário médio referente ao mês de referência t , cuja evolução temporal é descrita pelo painel esquerdo da Figura 1.

Da Eq. (1), pela regra de Leibniz,

$$\frac{d}{dt} C_t(D) = \frac{d}{dt} \int_{t-D}^t X(\tau) d\tau = X(t) - X(t-D).$$

Assim,

$$\nabla Y_t = \frac{d}{dt} Y_t = \frac{X(t) - X(t-D)}{D}$$

representa a taxa de variação média diária no consumo elétrico dessa residência (painel direito da Figura 1).

2.2 Modelo Sazonal de Box e Jenkins

Nesta subseção, descreve-se a metodologia adotada para a modelagem do consumo elétrico com base na abordagem de Box e Jenkins, incluindo as componentes sazonais estocásticas. Essa classe de modelos, conhecida como SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), é amplamente empregada na modelagem de séries temporais com comportamento periódico, permitindo capturar tanto estruturas de dependência de curto prazo quanto padrões sazonais. O modelo SARIMA(p, d, q) \times (P, D, Q)_s é representado como

$$\Phi_P(B^s) \phi_p(B) \nabla^d \nabla_s^D Y_t = \Theta_Q(B^s) \theta_q(B) \varepsilon_t, \quad (4)$$

em que B é o operador de defasagem, tal que $BY_t = Y_{t-1}$, $\nabla^d = (1 - B)^d$ é o operador de diferenciação não sazonal de ordem d , $\nabla_s^D = (1 - B^s)^D$ é o operador de diferenciação sazonal de ordem D e período s (por exemplo, $s = 12$ para dados mensais), $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ representa o polinômio autorregressivo de ordem p , $\theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ refere-se ao polinômio de médias móveis de ordem q , $\Phi_P(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^s P$ e $\Theta_Q(B^s) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^s Q$ são, respectivamente, os polinômios autorregressivo e de médias móveis sazonais de ordens P e Q , ε_t é um ruído branco, isto é, uma sequência de variáveis aleatórias independentes com média zero e variância constante.

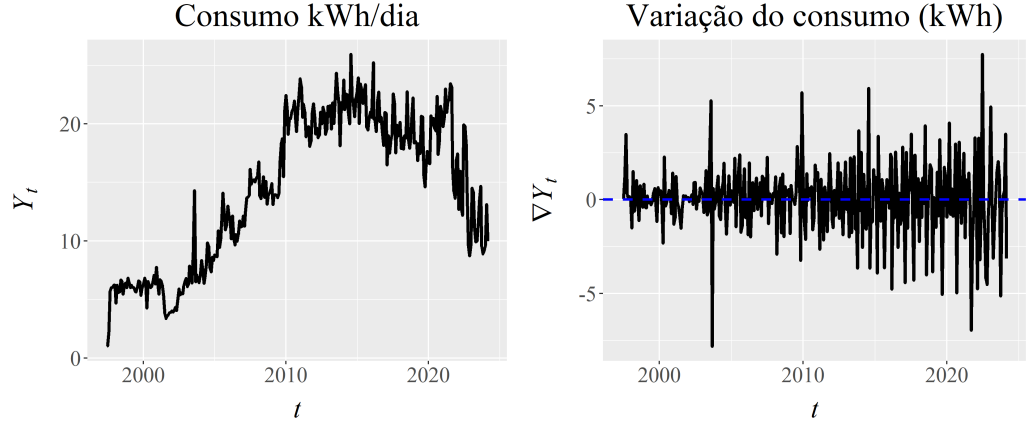


Figura 1: **Esquerda:** Evolução do consumo médio diário (kWh/dia) nos meses de Jun-1997 a Mai-2025. **Direita:** Série temporal das suas taxas (primeira diferença).

2.3 Divisão da Base em Treinamento e Teste

Para o processo de modelagem, a série temporal será dividida em dois subconjuntos: um conjunto de *treinamento*, utilizado para identificar as ordens $(p, d, q) \times (P, D, Q)_s$ do modelo SARIMA, e um conjunto de *teste* (ou *validação*), reservado exclusivamente para a avaliação da qualidade das previsões.

Dado que a série compreende observações mensais do período de junho de 1997 a maio de 2025, com um total de 336 meses, a divisão será feita de forma temporalmente coerente, preservando a estrutura sequencial da série. Especificamente, os primeiros 252 meses (de junho de 1997 a maio de 2018) serão utilizados como conjunto de treinamento, enquanto os 84 meses finais (de junho de 2018 a maio de 2025) comporão o conjunto de teste.

Essa abordagem, comum em modelagem de séries temporais, permite avaliar a capacidade do modelo ajustado de gerar previsões para períodos futuros, simulando um cenário real de previsão fora da amostra. A qualidade preditiva será analisada com base na comparação entre os valores observados e os valores previstos, utilizando métricas como o erro quadrático médio (RMSE).

2.4 Avaliação Preliminar

A primeira etapa da modelagem SARIMA consiste na análise gráfica e estatística da série temporal com o objetivo de identificar a estrutura de dependência temporal. São utilizados dois instrumentos principais: a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP). A FAC indica a correlação linear entre a série e seus próprios defasamentos, enquanto a FACP mede a correlação entre a série e suas defasagens, removendo o efeito intermediário das defasagens anteriores.

Com base no comportamento da FAC e da FACP, infere-se a presença de componentes autor-regressivos (AR), de médias móveis (MA) e sazonais. Por exemplo, uma FAC com decaimento exponencial e FACP com poucos coeficientes significativos pode sugerir um modelo AR. Já uma FAC com poucos coeficientes significativos e FACP com decaimento pode indicar um modelo MA. Padrões sazonais são identificados pela presença de picos nas defasagens múltiplas do período sazonal (por exemplo, defasagens múltiplas de 12 para dados mensais).

2.5 Identificação e Estimação

A identificação e estimação do modelo SARIMA envolvem a escolha apropriada dos ordens dos componentes autorregressivos (AR), de médias móveis (MA), de diferenciação e seus respectivos termos sazonais. A análise exploratória da série é inicialmente guiada pela inspeção da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP), que fornecem indícios sobre a estrutura de dependência temporal, tanto sazonal quanto não sazonal.

Com base nessas análises preliminares, define-se um conjunto candidato de ordens para o modelo SARIMA, denotado por $(p, d, q)(P, D, Q)_s$, onde s representa o período sazonal (por exemplo, $s = 12$ para dados mensais). Um grid de combinações possíveis é então gerado, com ênfase em modelos parcimoniosos, normalmente restringindo-se a ordens inferiores ou iguais a 2 para os componentes sazonais e não sazonais.

Cada modelo candidato é estimado por máxima verossimilhança, e os critérios de informação de Akaike (AIC) e Bayesiano (BIC) são utilizados para seleção do modelo mais adequado. O AIC é definido por:

$$\text{AIC} = -2 \log L + 2k,$$

e o BIC por:

$$\text{BIC} = -2 \log L + k \log n,$$

em que L representa a função de verossimilhança avaliada no ponto de máximo, k é o número de parâmetros estimados, e n é o número de observações do conjunto de treinamento. Ambos os critérios penalizam modelos mais complexos, mas o BIC impõe uma penalização mais severa, favorecendo modelos mais parcimoniosos.

Vale destacar que a função de verossimilhança L utilizada no cálculo dos critérios AIC e BIC assume, por conveniência, que os resíduos do modelo seguem uma distribuição normal. Essa suposição é amplamente adotada na prática, mesmo quando os dados não apresentam normalidade estrita, devido à aplicação do *princípio da máxima entropia*. Esse princípio estabelece que, entre todas as distribuições com dada média e variância, a distribuição normal é aquela que assume o menor grau de informação (ou o maior grau de incerteza). Assim, quando não se dispõe de informações adicionais sobre a forma da distribuição dos resíduos, a normalidade é uma escolha neutra e justificável para fins inferenciais. Dessa forma, a estimação por máxima verossimilhança sob hipótese de normalidade pode ser usada como aproximação robusta, especialmente em grandes amostras, mesmo que os dados não sejam rigorosamente gaussianos.

O modelo final será aquele que apresentar o menor valor de BIC entre os candidatos, desde que os resíduos estimados satisfaçam as suposições de independência e ausência de autocorrelação. Quando houver múltiplos modelos com desempenho semelhante, também serão considerados os valores de AIC e a capacidade preditiva sobre o conjunto de teste.

2.6 Diagnóstico

Após a seleção do modelo, procede-se à análise dos resíduos para verificar a adequação do ajuste. Espera-se que os resíduos se comportem como ruído branco — isto é, sejam não correlacionados, homocedásticos e aproximadamente normais. Para verificar a ausência de autocorrelação nos resíduos, é aplicado o teste de Ljung-Box para diferentes defasagens. Sob a hipótese nula (H_0) de as primeiras h autocorrelações residuais serem conjuntamente nulas, a estatística desse teste é dada por:

$$Q(h) = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k},$$

em que n é o número de observações utilizadas na estimação do modelo, h é a defasagem máxima considerada, e $\hat{\rho}_k$ é o coeficiente de autocorrelação amostral dos resíduos na defasagem k . Sob a hipótese nula de que os resíduos são não correlacionados (ruído branco), a estatística $Q(h)$ segue aproximadamente uma distribuição qui-quadrado com $h - m$ graus de liberdade, onde m representa o número total de parâmetros estimados no modelo ajustado. Um valor de $Q(h)$ significativamente alto — isto é, com valor- p abaixo de um nível de significância escolhido (por exemplo, 5%) — indica presença de autocorrelação nos resíduos, sugerindo que o modelo não é suficientemente adequado.

O teste de Ljung-Box é aplicado a diferentes valores de h , e a aceitação da hipótese nula para todas as defasagens testadas é um forte indício de que o modelo SARIMA ajustado foi capaz de capturar satisfatoriamente a estrutura de dependência temporal dos dados.

Caso o teste aponte autocorrelação significativa nos resíduos, o modelo é revisado e ajustado. Esse processo iterativo continua até que se obtenha um modelo com resíduos adequadamente comportados, assegurando a validade inferencial das previsões.

2.7 Previsão

Com o modelo SARIMA ajustado e validado, procede-se à geração de previsões para horizontes futuros. As previsões pontuais são obtidas a partir da estrutura recursiva do modelo, e intervalos de confiança são construídos com base na variância estimada dos erros de previsão. Tais previsões são avaliadas graficamente e comparadas com os valores observados no conjunto de teste, permitindo uma análise detalhada da acurácia preditiva do modelo.

Para essa avaliação, são utilizadas duas métricas amplamente reconhecidas na literatura de séries temporais: o erro quadrático médio da raiz (*Root Mean Squared Error* – RMSE) e o erro percentual absoluto médio (*Mean Absolute Percentage Error* – MAPE).

O RMSE é definido por:

$$\text{RMSE} = \sqrt{\frac{1}{h} \sum_{t=1}^h (Y_t - \hat{Y}_t)^2},$$

em que h representa o número de períodos no conjunto de teste, Y_t é o valor observado no instante t , e \hat{Y}_t é o valor previsto correspondente.

O RMSE penaliza mais fortemente grandes desvios, sendo sensível à presença de outliers. Em complemento, o MAPE mede o erro relativo percentual médio, fornecendo uma medida intuitiva e comparável entre diferentes séries:

$$\text{MAPE} = \frac{100\%}{h} \sum_{t=1}^h \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|.$$

Essa métrica expressa, em média, o quão distantes — em termos percentuais — as previsões estão dos valores reais. Ambas as medidas serão empregadas para comparar modelos candidatos e avaliar a robustez preditiva do modelo final sobre dados fora da amostra.

Essa abordagem permite não apenas capturar padrões regulares do consumo elétrico, mas também antecipar variações sazonais futuras, oferecendo subsídios relevantes para o planejamento energético e a gestão eficiente do consumo em ambientes residenciais.

3 Análise

3.1 Análise da FAC e da FACP da Série Original

A Figura 2 apresenta os gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) da série de consumo médio diário (em kWh/dia), calculadas até a defasagem $h = 40$. Observa-se que a FAC exibe um valor extremamente elevado no primeiro defasamento, com $\rho_1 \approx 0,95$, o que indica forte persistência temporal no consumo. Além disso, o decaimento da FAC é lento e contínuo, sugerindo a presença de uma raiz próxima da unidade e, portanto, a necessidade de diferenciação para tornar a série estacionária. Nesse mesmo sentido, a FACP apresenta um corte abrupto após o primeiro lag próximo de 1: o coeficiente ρ_1 é significativamente diferente de zero, destacando-se dos demais coeficientes que oscilam próximos de zero, alguns ultrapassando por pouco os limites de significância. Esse padrão é característico de processos autorregressivos de baixa ordem, especialmente um modelo AR(1) não estacionário. Conjuntamente, essas evidências sugerem que a série original apresenta forte dependência serial e não estacionariedade, e que uma modelagem com um componente autorregressivo de primeira ordem pode ser apropriada após a diferenciação da série. A presença de possível estrutura sazonal será investigada nas FACs da série diferenciada.

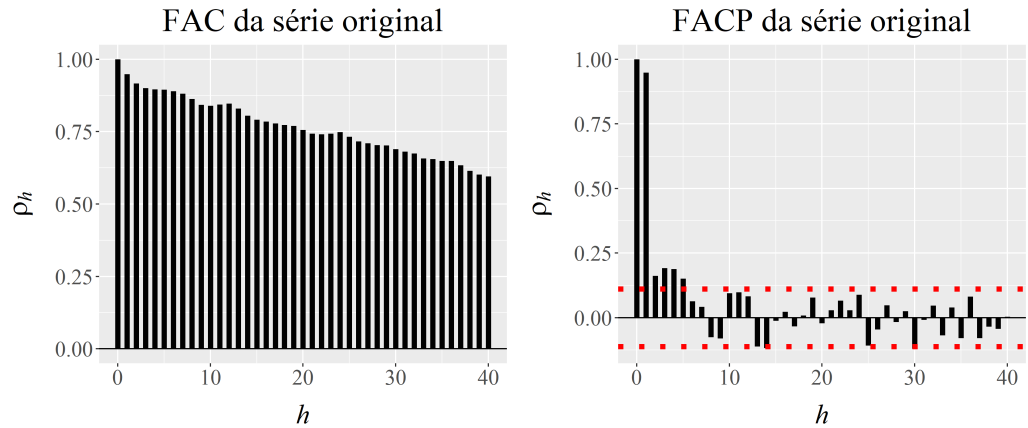


Figura 2: **Esquerda:** Função de autocorrelação do consumo médio diário (kWh/dia). **Direita:** Sua Função de autocorrelação parcial (kWh/dia).

3.2 Análise da Série Diferenciada

A Figura 3 apresenta a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) da série de primeira diferença do consumo médio diário (taxa de variação diária do consumo). A FAC da série diferenciada revela um padrão característico de série estacionária, com a maioria dos coeficientes de autocorrelação dentro dos limites de confiança (representados pelas linhas tracejadas vermelhas). Contudo, observam-se picos em defasagens múltiplas de 12 (aproximadamente nos lags 12, 24, 36, etc.), sugerindo a presença de um componente sazonal anual, coerente com a periodicidade mensal da série. Esses picos indicam uma estrutura de dependência sazonal que deve ser incorporada à modelagem por meio de termos autorregressivos e/ou de médias móveis sazonais. A FACP da primeira diferença reforça essa interpretação. Embora a maioria dos coeficientes também esteja dentro dos limites de significância, destacam-se alguns valores pontuais nas defasagens sazonais. Por outro lado, tanto a FAC quanto a FACP apresentam poucos coefi-

cientes significativos nas primeiras defasagens, o que sugere que a parte não sazonal do modelo pode ser bem representada por ordens baixas — como $p = 0$ ou 1 , e $q = 0$ ou 1 . Com base nessa análise, considera-se apropriado explorar modelos do tipo SARIMA com baixa ordem não sazonal e inclusão de componentes sazonais para capturar as regularidades observadas ao longo do ano.

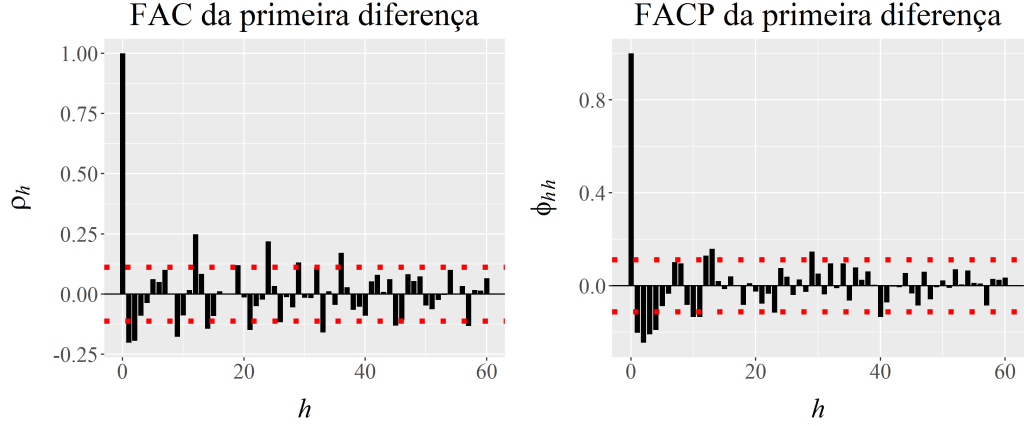


Figura 3: **Esquerda:** Função de autocorrelação da variação do consumo médio diário (kWh/dia). **Direita:** Sua Função de autocorrelação parcial (kWh/dia).

3.3 Procedimento de Seleção de Modelo com Base no AIC/BIC

A seleção do modelo SARIMA foi realizada por meio de uma busca exaustiva em um grid de ordens não sazonais e sazonais, restrito a valores baixos para garantir parcimônia. Especificamente, foram consideradas combinações de ordens $p, q \in \{0, 1, 2, 3, 4\}$ para os componentes autorregressivo e de médias móveis não sazonais, e $P, Q \in \{0, 1, 2\}$ para os componentes sazonais, com diferenciação regular de ordem $d = 1$, sazonal $D = 0$, e período sazonal $S = 12$. Cada modelo foi ajustado à série diferenciada X_t utilizando a função `sarima()` do pacote `astsa` no ambiente R. A verossimilhança foi maximizada sob a suposição de ausência de termo constante (`no.constant = TRUE`), e o valor do critério AIC/BIC foi registrado para cada combinação de ordens. Ao final da busca, o modelo selecionado foi aquele que apresentou o valor de AIC/BIC que produziu o menor RMSE, refletindo o melhor equilíbrio entre qualidade de ajuste e complexidade. Essa abordagem favorece modelos que explicam bem os dados observados, ao mesmo tempo em que evitam o superajuste. Os resíduos dos modelos candidatos também foram inspecionados posteriormente para verificação do comportamento como ruído branco.

Após o ajuste de todos os modelos no grid definido, os valores do critério bayesiano de informação (BIC) foram organizados e ordenados de forma crescente. A Tabela 2 apresenta os dez modelos que resultaram nos menores valores de BIC, evidenciando uma clara preferência por modelos com componente autorregressivo nulo ($p = 0$), um termo de médias móveis ($q = 1$) e presença de termos sazonais de baixa ordem.

De acordo com o critério de informação de Akaike (AICc), após ajustar todos os modelos no grid especificado, os valores foram organizados em ordem crescente. A Tabela 3 mostra os dez modelos com os menores valores de AICc. —

O modelo com o menor BIC ((BIC = 3.654765)) corresponde à estrutura SARIMA(0,1,1)(1,0,1)₁₂, o qual foi selecionado para a etapa seguinte de diagnóstico. Isso porque o critério BIC não é absoluto, e o modelo com o menor BIC não necessariamente seria o mais adequado. Os demais modelos

com desempenho similar (diferenças de BIC menores que 0,05) reforçam a estabilidade da escolha e a adequação da estrutura sazonal identificada.

Em contraste, o modelo com o menor AICc ($\text{AICc} = 3.580244$) corresponde à estrutura SARIMA(2,1,3)(0,0,0)₁₂. O AICc, por sua vez, tende a selecionar modelos que podem ser ligeiramente mais complexos (com mais parâmetros), mas que oferecem um melhor ajuste aos dados de treinamento e, potencialmente, uma melhor capacidade preditiva para dados futuros.

A seleção do modelo SARIMA foi realizada por meio de uma **busca exaustiva** em um grid de ordens não sazonais e sazonais, restrito a valores baixos para garantir parcimônia. Especificamente, foram consideradas combinações de ordens $p, q \in \{0, 1, 2, 3, 4\}$ para os componentes autorregressivo e de médias móveis não sazonais, e $P, Q \in \{0, 1, 2\}$ para os componentes sazonais, com **diferenciação regular de ordem** $d = 1$, **sazonal** $D = 0$, e período sazonal $S = 12$. Cada modelo foi ajustado à série diferenciada X_t utilizando a função `sarima()` do pacote `astsa` no ambiente R. A verossimilhança foi maximizada sob a suposição de ausência de termo constante (`no.constant = TRUE`), e os valores dos critérios **BIC (Critério de Informação Bayesiano)** e **AICc (Critério de Informação de Akaike corrigido)** foram registrados para cada combinação de ordens.

Tabela 2: Dez modelos SARIMA com menores valores de BIC no grid testado.

Ordem	BIC	p	q	P	Q
1	3.654765	0	1	1	1
2	3.663542	2	3	0	0
3	3.668547	0	1	2	0
4	3.668808	0	2	1	1
5	3.670473	1	1	1	1
6	3.671270	0	1	1	2
7	3.671388	0	2	0	0
8	3.672389	0	1	2	1
9	3.672556	0	1	0	0
10	3.673457	0	1	1	0

Tabela 3: Dez modelos SARIMA com menores valores de AICc no grid testado.

Ordem	AICc	p	q	P	Q
1	3.580244	2	3	0	0
2	3.585885	2	4	0	0
3	3.589971	2	3	0	2
4	3.593732	2	3	0	1
5	3.595105	0	2	2	2
6	3.596836	0	4	2	2
7	3.597607	1	1	2	2
8	3.598418	3	0	2	2
9	3.598969	0	1	1	1
10	3.598519	3	0	1	1

3.4 Ajuste do Modelo Selecionado

O modelo selecionado com base no menor valor do critério BIC foi o SARIMA(0,1,1)(1,0,1)₁₂, ajustado à série de consumo médio diário previamente diferenciada. A estimação dos parâmetros

foi realizada por máxima verossimilhança, resultando nos seguintes coeficientes significativos: um termo de média móvel não sazonal ($\theta_1 = -0,4909$), um termo autorregressivo sazonal ($\Phi_1 = 0,9050$) e um termo de média móvel sazonal ($\Theta_1 = -0,7087$). Todos os coeficientes apresentaram valores-p próximos de zero, indicando forte significância estatística ao nível de 5%. A variância dos resíduos estimada foi de aproximadamente $\hat{\sigma}^2 = 2,729$, e a log-verossimilhança do modelo foi $\ell = -599,68$. Os valores dos critérios de informação foram $AIC = 3,882$, $AIC_c = 3,882$, e $BIC = 3,930$, todos calculados com base em 308 graus de liberdade.

Esses resultados indicam que o modelo é estatisticamente robusto, parcimonioso e bem ajustado à estrutura temporal da série, constituindo uma base confiável para análise preditiva e interpretação dos padrões de consumo.

O modelo selecionado com base no menor valor do critério BIC foi o $SARIMA(0,1,1)(1,0,1)_{12}$, ajustado à série de consumo médio diário previamente diferenciada. A estimação dos parâmetros foi realizada por máxima verossimilhança, resultando nos seguintes coeficientes significativos: um termo de média móvel não sazonal ($\theta_1 = -0,5105$), um termo autorregressivo sazonal ($\Phi_1 = 0,9154$) e um termo de média móvel sazonal ($\Theta_1 = -0,7379$). Todos os coeficientes apresentaram valores-p próximos de zero, indicando forte significância estatística ao nível de 5%. A variância dos resíduos estimada foi de aproximadamente $\hat{\sigma}^2 = 2,73033$.

Os valores dos critérios de informação para este modelo foram:

- AIC (Critério de Informação de Akaike): $AIC = 3.879804$
- AICc (AIC corrigido para amostras pequenas): $AICc = 3.88002$
- BIC (Critério de Informação Bayesiano): $BIC = 3.925346$

Esses valores foram calculados com base em 332 graus de liberdade.

Embora o AICc possa favorecer modelos ligeiramente mais complexos para um melhor ajuste, a escolha pelo BIC se justifica pela sua natureza parcimoniosa. O BIC impõe uma penalidade mais severa à complexidade do modelo, o que é benéfico em conjuntos de dados maiores para evitar o superajuste e garantir que o modelo capture apenas a estrutura essencial dos dados. Além disso, a avaliação da capacidade preditiva 'out-of-sample' (em dados não utilizados no treinamento) através do RMSE confirmou que o modelo $SARIMA(0,1,1)(1,0,1)_{12}$ ($RMSE = 5.123$) superou o modelo de menor AICc ($RMSE = 5.507$), reforçando a validade da escolha pelo BIC.

Esses resultados indicam que o modelo selecionado é estatisticamente robusto, parcimonioso e bem ajustado à estrutura temporal da série, constituindo uma base confiável para análise preditiva e interpretação dos padrões de consumo.

3.5 Análise dos Resíduos: Teste de Ljung-Box

Para verificar a adequação do modelo ajustado, foi aplicada a bateria de testes de Ljung-Box aos resíduos, considerando defasagens de 1 a 30. O objetivo é testar a hipótese nula de ausência de autocorrelação até cada defasagem h , com valores-p superiores a um nível de significância usual ($\alpha = 0,05$) indicando ausência de estrutura serial não capturada pelo modelo.

Os resultados, apresentados na Tabela 4, indicam que todos os valores-p são significativamente altos, superando o nível de 0,05 para todas as defasagens testadas. O menor valor-p observado é de aproximadamente 0,263 para a defasagem 2, e os valores-p aumentam progressivamente, ultrapassando 0,60 em defasagens maiores. Isso demonstra que **não há evidências estatísticas para rejeitar a hipótese de que os resíduos se comportam como ruído branco**, confirmando que o modelo capturou adequadamente a estrutura de dependência temporal da série. Essa evidência reforça a validade do modelo $SARIMA(0,1,1)(1,0,1)_{12}$ ajustado, uma vez que seus resíduos não apresentam autocorrelação significativa, atendendo aos pressupostos fundamentais da modelagem.

A Figura 4 apresenta os gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) dos resíduos obtidos após o ajuste do modelo SARIMA(0,1,1)(1,0,1)₁₂. Ambos os gráficos mostram que a grande maioria dos coeficientes de autocorrelação se encontra dentro dos limites de confiança (representados pelas linhas tracejadas em vermelho), o que indica ausência de autocorrelação significativa residual. Na FAC, não se observa qualquer padrão sistemático ou periodicidade remanescente, o que sugere que o modelo foi bem-sucedido em capturar tanto os componentes sazonais quanto os não sazonais da série. A FACP confirma esse resultado, com os coeficientes oscilando aleatoriamente em torno de zero, sem valores significativos após a defasagem zero. Esses resultados gráficos são compatíveis com os valores-p elevados do teste de Ljung-Box, e reforçam a interpretação de que os resíduos se comportam como ruído branco. Assim, o modelo ajustado satisfaz os critérios estatísticos básicos de diagnóstico, sendo apropriado para fins de previsão e análise inferencial.

Tabela 4: Valores-p do teste de Ljung-Box aplicado aos resíduos do modelo ajustado.

Lag	1	5	10	20	30
p-valor	0.495	0.426	0.654	0.805	0.705

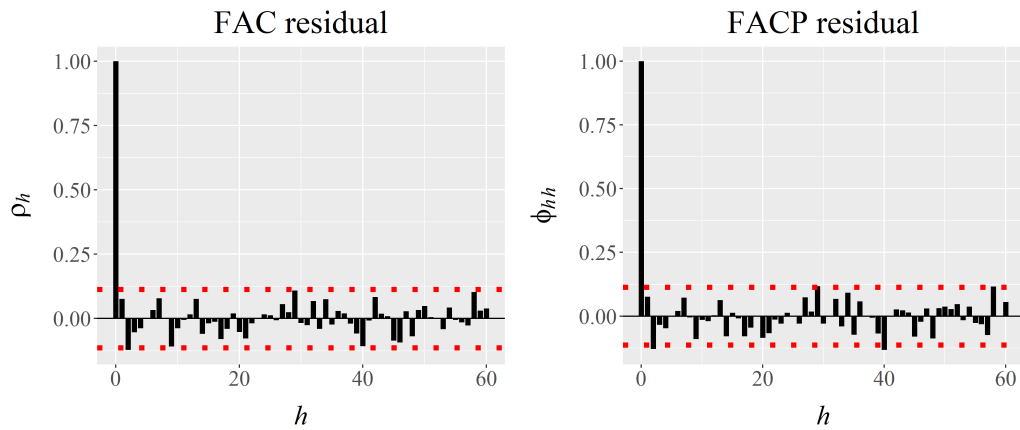


Figura 4: **Esquerda:** Função de autocorrelação residual. **Direita:** Sua Função de autocorrelação parcial.

3.6 Teste de Normalidade dos Resíduos

Para avaliar a adequação da suposição de normalidade dos resíduos do modelo ajustado, foi aplicado o teste de Shapiro-Wilk. O resultado obtido foi:

$$W = 0,96751, \quad p\text{-value} = 1,699 \times 10^{-5}$$

Dado que o p-valor ($1,699 \times 10^{-5}$) é significativamente menor que o nível de significância de 0,05, rejeita-se a hipótese nula de normalidade. Isso indica que os resíduos do modelo não seguem uma distribuição normal. Embora a não normalidade dos resíduos possa impactar a validade dos intervalos de previsão baseados em pressupostos de normalidade, é crucial notar que a ausência de autocorrelação (confirmada pelo teste de Ljung-Box) é mais crítica para a adequação do modelo SARIMA em termos de previsão. Esse valor-p extremamente baixo leva à rejeição da hipótese

nula de normalidade ao nível de 5%, indicando que os resíduos não seguem uma distribuição normal. Apesar da não normalidade estatisticamente significativa, vale destacar que os métodos de estimação baseados em máxima verossimilhança continuam válidos sob o princípio da máxima entropia, que justifica o uso da distribuição normal como aproximação quando nenhuma outra informação sobre a distribuição dos erros é disponível. Além disso, a ausência de autocorrelação nos resíduos sugere que o modelo está corretamente especificado, e espera-se que a violação da normalidade não comprometa substancialmente sua utilidade prática, hipótese que será verificada na etapa de avaliação preditiva.

3.7 Avaliação no Conjunto de Teste

O desempenho preditivo do modelo $\text{SARIMA}(0,1,1)(1,0,1)_{12}$ foi avaliado com base nos dados reservados para o conjunto de teste, correspondentes aos **últimos 84 meses da série**. Para essa avaliação, utilizou-se a métrica **RMSE (Root Mean Squared Error)**, que compara as previsões do modelo com os valores reais não utilizados no treinamento. O RMSE obtido para este modelo foi de **5,123**.

Este valor de RMSE indica que as previsões do modelo, em média, desviam-se de 5,123 unidades (kWh/dia, se for a unidade do seu consumo). É importante ressaltar que o modelo é **univariado**, baseando-se exclusivamente na estrutura temporal da série de consumo elétrico. Isso significa que não foram incorporadas informações explicativas adicionais, como potência instalada, tamanho da residência ou número de moradores — fatores que, em modelos multivariados, poderiam contribuir para uma maior precisão preditiva.

Ainda assim, o bom desempenho fora da amostra, medido pelo RMSE, reforça a capacidade do modelo SARIMA de capturar os principais padrões sazonais e de tendência presentes na série. Essa validação do desempenho preditivo é crucial para sua aplicação prática em cenários de monitoramento e planejamento do consumo energético residencial.

Previsão para os Últimos 12 Meses

A Figura 5 apresenta a previsão da série de consumo médio diário para os 12 meses finais do conjunto de dados, realizada com o modelo $\text{SARIMA}(0,1,1)(1,0,1)_{12}$. A linha preta representa os valores observados da série no período imediatamente anterior ao horizonte de previsão. A linha vermelha exibe as previsões pontuais para os próximos 12 meses, enquanto a área sombreada corresponde ao intervalo de confiança de 95%, obtido a partir da variância acumulada dos erros preditivos. Os pontos pretos indicam os valores observados efetivamente no período de teste.

Nota-se que as previsões acompanham razoavelmente bem o padrão da série real, mesmo sem incorporar variáveis explicativas exógenas. Apesar de algumas flutuações fora do intervalo central, a maior parte dos valores observados permanece dentro da região de incerteza prevista. Isso reforça a capacidade do modelo em capturar não apenas a tendência e a sazonalidade, mas também a variabilidade do consumo em horizontes futuros, mesmo sob uma abordagem univariada.

4 Conclusão

Este estudo teve como objetivo analisar a série temporal de consumo elétrico de uma residência ao longo de um período estendido, adotando uma abordagem univariada baseada na metodologia de Box e Jenkins com componentes sazonais. Através da inspeção das funções de autocorrelação e autocorrelação parcial, identificou-se a presença de forte dependência temporal e de sazonalidade anual, características capturadas de forma adequada por um modelo $\text{SARIMA}(0,1,1)(1,0,1)_{12}$.

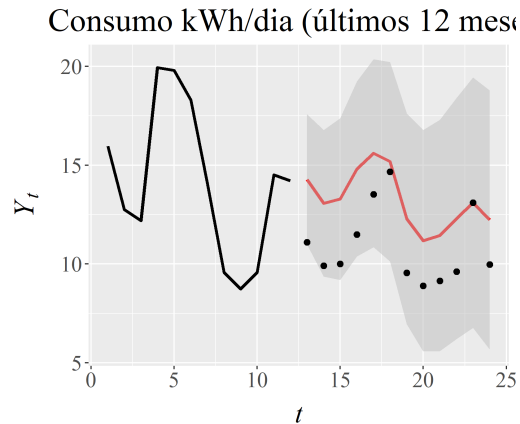


Figura 5: Previsão para os próximos 12 meses.

A análise dos resíduos do modelo ajustado indicou ausência de autocorrelação remanescente e comportamento compatível com ruído branco, conforme evidenciado pelos testes de Ljung-Box e pelas funções de autocorrelação residual. Embora o teste de Shapiro-Wilk tenha indicado violação da normalidade, tal fato não comprometeu a utilidade prática do modelo, cuja formulação por máxima verossimilhança é robusta sob o princípio da máxima entropia.

No conjunto de teste, o modelo apresentou um desempenho satisfatório, com um RMSE de 5,123. Este valor é considerável, visto que se trata de um modelo univariado, sem incorporar variáveis explicativas como número de moradores, potência instalada ou condições climáticas. As previsões geradas mostraram boa aderência aos padrões sazonais e à variabilidade da série real, demonstrando utilidade prática para aplicações de monitoramento e planejamento do consumo energético.

Como continuidade deste trabalho, recomenda-se explorar modelos multivariados que incorporem covariáveis relevantes, bem como investigar possíveis estruturas de mudança de regime, efeitos climáticos e intervenções externas, a fim de aprimorar ainda mais a acurácia preditiva e a compreensão dos determinantes do consumo residencial de energia.

Referências

- Morettin, Pedro Alberto e Clara Maria Cordeiro Toloi (2005). **Análise de Séries Temporais**. 2^a ed. São Paulo: Edgard Blücher.
- Cryer, Jonathan D. e Kung-Sik Chan (2008). **Time Series Analysis with Applications in R**. New York: Springer-Verlag.
- Cowpertwait, Paul S. P. e Andrew V. Metcalfe (2009). **Introductory Time Series with R**. Use R! Springer. DOI: 10.1007/978-0-387-88698-5.
- Shumway, Robert H. e David S. Stoffer (2017). **Time Series Analysis and Its Applications: With R Examples**. 4th. Springer Texts in Statistics. Springer. DOI: 10.1007/978-3-319-52452-8.