

Modelagem e previsão de modelos de séries temporais do consumo de energia elétrica na Região Nordeste do Brasil

Eduardo G. Araújo^{1†}, Silvio F. A. Xavier Júnior², Nyedja F. M. Barbosa³, Tiago A. Oliveira⁴

¹Discente do Programa de Pós-Graduação em Biometria e Estatística Aplicada - PPGBEA, Departamento de Estatística e Informática - DEINFO, Universidade Federal Rural de Pernambuco - UFRPE

²Docente do Departamento de Estatística, Universidade Estadual da Paraíba - UEPB. E-mail: silvio@servidor.uepb.edu.br

³Docente do Departamento de Estatística, Universidade Estadual da Paraíba - UEPB. E-mail: nyedja@servidor.uepb.edu.br

⁴Docente do Departamento de Estatística, Universidade Estadual da Paraíba - UEPB. E-mail: tadolive@servidor.uepb.edu.br

Resumo: A energia elétrica é uma das formas de energia mais utilizadas, sendo de grande importância para o crescimento econômico mundial. Neste contexto, analisou-se a série temporal do consumo de energia elétrica no Nordeste brasileiro, no período de janeiro de 1997 a maio de 2021 e estimou-se a previsão para todo o ano de 2021, baseando-se em metodologias utilizadas na análise de séries temporais. Os métodos aplicados para estas previsões foram os algoritmos de Holt-Winters, nas formas aditiva e multiplicativa, e também o método de modelagem de Box-Jenkins, nas formas SARIMA e SARIMAX. Para o caso do modelo SARIMAX, aplicou-se uma variável dummy, com valores 1 para os meses a partir de Junho de 2001, adotando como referência o racionamento de energia do governo Fernando Henrique Cardoso, e 0 para os meses anteriores. A escolha do modelo para ajuste da série temporal foi feita com base nos critérios AIC (Critério de Informação de Akaike), BIC (Critério de Informação Bayesiano) e EQMP (Erro Quadrático Médio de Previsão). Os cálculos foram feitos no software estatístico R. Por fim, o modelo SARIMA(4,1,1)(2,0,0)₁₂ foi o que forneceu a melhor previsão para o consumo de energia elétrica no Nordeste brasileiro.

Palavras-chave: Previsões; Holt-Winters; Box-Jenkins.

Abstract: Electric energy is one of the most used forms of energy, which is important for world economic growth. In this context, we analyzed the time series of electricity consumption in the Brazilian Northeast from January 1997 to May 2021. The forecast for the year 2021 was estimated based on methodologies used in the time series. The methods applied for these predictions were the Holt-Winters algorithms in the additive and multiplicative forms and the Box-Jenkins modeling methodology (we used the SARIMA and SARIMAX models). For the case of the SARIMAX model, a dummy variable was applied, with values of 1 for the months from June 2001 onwards, using the energy rationing of the Fernando Henrique Cardoso government as a reference and 0 for the previous months. The choice of model for time series adjustment was based on the criteria AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and MSPE (Mean Prediction Forecast Error). The calculations were performed using the R statistical software. Finally, the model SARIMA (4,1,1)(2,0,0)₁₂ provided the best forecast for electricity consumption in the Brazilian Northeast.

Keywords: Forecasts; Holt-Winters; Box-Jenkins.

[†]*Autor correspondente: eduardogomesdearaujo.1995@gmail.com.

Introdução

A importância do setor elétrico é cada vez mais evidente, pois contribui para o desenvolvimento das diferentes regiões brasileiras, proporcionando grande comodidade à população. No Brasil, a energia elétrica é dada por uma matriz composta por energias renováveis e não-renováveis, sendo formada por hidráulica, solar, eólica e térmicas. Grande parte da energia elétrica gerada é proveniente das usinas hidrelétricas, e apenas uma pequena parte das demais (ENERGÉTICA, 2019).

Segundo Schmidt e Lima (2004), a demanda por energia elétrica vem aumentando exponencialmente nas últimas décadas, devido pelo bom desempenho e expansão dos setores comercial, residencial e industrial de nosso país. Devido esse aumento, foi necessário toda uma reestruturação do sistema para que se tivesse um setor elétrico com fornecimento de energia de qualidade e minimização de custo (CAMPOS, 2008).

Sendo assim, procedimentos e métodos estatísticos são valiosos na análise do consumo de energia elétrica, com o objetivo de identificar e discutir padrões de comportamento ao longo dos anos, bem como analisar as relações presentes nas observações (TIDRE; BIASE; SILVA, 2013).

De acordo com Morettin e Toloí (2006), uma série temporal é qualquer coleção de observações ordenadas no tempo, na qual a sequência dos dados é fundamental. Neste sentido, o estudo do comportamento de uma variável ao longo do tempo por meio de uma série temporal se torna importante.

Para o presente estudo foi-se utilizado dados oriundos do Instituto de Pesquisa Econômica Aplicada (IPEADATA, 2021), referente a região Nordeste do Brasil no período de janeiro de 1979 a maio de 2021. Autores como Tidre, Biase e Silva (2013), que realizaram previsões para a região Norte, utilizando diversos modelos e Campos (2008), que realizou previsões para a série do consumo de energia elétrica em Minas Gerais; obtiveram ótimas performances na previsão do consumo de energia, concluindo que tais modelos podem auxiliar nas tomadas de decisões no setor elétrico.

Dessa forma, esse trabalho tem como objetivo aplicar modelos de séries temporais ao consumo de energia a fim de encontrar o modelo que melhor represente seu caráter preditivo. Através do Alisamento Exponencial de Holt-Winters, aditivo e multiplicativo, busca-se previsões 8 passos à frente, ou seja, prever como o consumo de energia se comportará até o final do ano de 2021. E também a modelagem Box-Jenkins por meio de previsões nos modelos SARIMA e SARIMAX, na qual busca-se um melhor modelo através dos dados e utilizando um modelo base (JUNIOR et al., 2018), como comparação para obter uma melhor previsão para o consumo de energia.

Metodologia

Conceitos básicos

Uma série temporal $Y_t = \{t_1, t_2, t_3, \dots, t_n\}$ é definida como uma coleção de observações sequenciais em um determinado tempo t , ou seja, são dados observados em diferentes instantes do tempo. Esta pode ser *discreta*, quando acontece em um tempo específico, como $(Y_t = t_1, \dots, t_n)$, ou *contínua*, quando as observações ocorrem continuamente no tempo $(Y_t = t : t_1 < t < t_2)$ (MORETTIN; TOLOI, 2006).

De acordo com Morettin e Toloí (2006), para compreensão do comportamento das séries temporais é necessário ter conhecimento dos componentes característicos, sendo eles:

- **Tendência:** comportamento que a série apresenta a longo prazo, incluindo crescimento e/ou decréscimo, com vários possíveis padrões;
- **Sazonalidade:** comportamento que tende a se repetir a cada período de tempo;
- **Nível:** oscilação dos dados que ocorrem repetidamente em torno da tendência;

- **Erro aleatório:** flutuações não identificadas.

Outra característica importante em séries temporais, remete-se a sua estacionariedade ou não estacionariedade. Tecnicamente existem duas formas de estacionariedade, sendo elas: estritamente (fortemente) estacionária e fracamente estacionária (ou de segunda ordem). Segundo Morettin e Toloi (2006), tem-se as seguintes definições:

- Estacionariedade Forte: Um processo estocástico $Z = \{Z_t, t \in T\}$ diz-se estritamente estacionário se todas as distribuições finito-dimensionais, ou seja, $F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n\}$, permanecem invariantes no tempo.

Assim, temos que sua média e variância são constantes nas translações do tempo, portanto:

$$E(t) = \mu \text{ e } Var(t) = \sigma^2, \forall t \in T. \quad (1)$$

- Estacionariedade Fraca: Um processo estocástico $Z = \{Z_t, t \in T\}$ diz-se fracamente estacionário se e somente se:

1. $E[Z_t] = \mu$, constante $\forall t \in T$;
2. $E^2[Z_t] < \infty, t \in T$;
3. $\gamma_{t,s} = Cov[Z_t, Z_s]$ é uma função de $t - s$, chamada de defasagem.

Modelos de Suavização Exponencial

Modelos de suavização é uma grande classe de métodos de previsão que se baseiam na ideia de que observações passadas contém informações sobre o padrão da série temporal, ou seja, os valores mais relevantes a serem estimados são os termos mais recentes de tendência. Considerando-se Z_1, Z_2, \dots, Z_t uma série temporal, suas previsões serão construídas para Z_{t+1}, Z_{t+2}, \dots . Sendo assim, através dos erros encontrados em previsões passadas, são obtidas as previsões dos modelos de suavização exponencial.

Alisamento Exponencial Simples (AES)

O método de Alisamento Exponencial Simples (AES) é utilizado para séries temporais onde não existe a presença de sazonalidade, nem tendência sistemática (EHLERS, 2007), ou seja, são séries localmente constantes, sendo compostas apenas por seu nível e ruído aleatório (MORETTIN; TOLOI, 2006). Assim, a série é decomposta da seguinte forma:

$$Z_t = N_t + \epsilon_t, \text{ com } t = 1, 2, \dots, n. \quad (2)$$

em que $\epsilon_t \sim (0, \sigma_t^2)$ e N_t representam o nível da série no instante t .

Alisamento Exponencial de Holt

É apropriado para situações em que a série apresenta tendência e não apresenta sazonalidade, pois previsões futuras pelo AES induziriam a um erro sistemático de previsão devido ao atraso no acompanhamento da tendência (MORETTIN; TOLOI, 2006). O método de Holt utiliza dois coeficientes diferentes de alisamento, sendo eles o nível (também presente no AES) e a tendência linear (crescente ou decrescente) (MONTGOMERY; JENNINGS; KUKAHCI, 2015).

Esses coeficientes são estimados da seguinte maneira:

$$N_t = \alpha Z_t + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (3)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (4)$$

com $0 < \alpha < 1$ e $0 < \beta < 1$, onde α e β são as constantes de suavização.

Alisamento Exponencial de Holt-Winters (HW)

Utilizado quando a série apresenta tendência e sazonalidade, o modelo de Holt-Winters é uma expansão do método Holt, desenvolvido por Winters, e se divide em dois grupos: modelos aditivos e multiplicativos. A utilização destes algoritmos baseia-se em três equações com constantes de suavização diferentes, sendo associada a cada uma das componentes do padrão da série: nível, tendências e sazonalidade (MORETTIN; TOLOI, 2006).

- Série sazonal multiplicativa: Considerando-se uma série sazonal com período s . A variante mais usual do método HW considera o fator sazonal F_t como sendo multiplicativo, enquanto a tendência permanece aditiva (MORETTIN; TOLOI, 2006). A forma de recorrência é dada pelas três equações de suavização:

$$N_t = \frac{\alpha Z_t}{F_{t-s}} + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (6)$$

$$F_t = \frac{\gamma Z_t}{N_t} + (1 - \gamma)F_{t-s}, \quad (7)$$

com $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \gamma < 1$, onde α , β e γ são constantes de suavização do algoritmo.

- Série sazonal aditiva: Após modificar o procedimento multiplicativo em que o fator sazonal passa a ser aditivo, suas atualizações dos dados são dadas pelas seguintes expressões (MORETTIN; TOLOI, 2006):

$$N_t = \alpha(Z_t - F_{ts}) + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (8)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (9)$$

$$F_t = \gamma(Z_t - N_t) + (1 - \gamma)F_{t-s} \quad (10)$$

com $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \gamma < 1$, onde α , β e γ são as constantes de suavização do algoritmo.

Modelagem Box-Jenkins

Uma abordagem bastante utilizada para a construção de modelos paramétricos para séries temporais univariadas é conhecida como metodologia Box e Jenkins. Tal metodologia consiste em propor e ajustar modelos lineares estacionários (ou não-estacionários) a uma série de tempo observada (MORETTIN; TOLOI, 2006).

De acordo com Morettin e Toloí (2006), a metodologia Box e Jenkins é baseada em um ciclo com as seguintes etapas:

1. Especificação: Uma classe geral de modelos é considerada;
2. Identificação: Um modelo é identificado com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
3. Estimação: Estimam-se os parâmetros do modelo;

4. Verificação ou diagnóstico: Verifica-se se o modelo escolhido ajusta-se adequadamente aos dados, pela análise de resíduo.

Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Muitos modelos poderão ser estimados em uma série, portanto, se a finalidade da estimação é a previsão, busca-se o modelo que tiver o menor erro quadrático médio de previsão (MORETTIN; TOLOI, 2006).

Modelo autorregressivo integrado de médias móveis (ARIMA)

O modelo autorregressivo integrado de médias móveis (ARIMA) é uma generalização do modelo ARMA. Quando as séries não apresentam estacionariedade, a série pode se tornar estacionária ao aplicarmos d diferenças aos dados, na qual no máximo 2 diferenças podem ser aplicadas (BOX et al., 2016).

O modelo é denotado por $ARIMA(p, d, q)$, sendo p o operador autorregressivo, d o operador diferença e q o operador de médias móveis, da qual a equação é dada por:

$$W_t = \phi_1 W_{t-1} + \dots + \phi_p W_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (11)$$

em que $W_t = \nabla^d Z_t$.

E pode ainda ser escrita como:

$$\phi(B)((1 - B)^d Z_t - \alpha) = \theta(B)\epsilon_t \quad (12)$$

Modelo autorregressivo de médias móveis sazonal (SARIMA)

Sendo uma extensão do modelo $ARIMA(p, q, d)$, o modelo SARIMA leva em consideração a sazonalidade estocástica dos dados (MORETTIN; TOLOI, 2006). Este modelo possui notação $SARIMA(p, d, q) \times (P, D, Q)$, na qual:

- **P** é o número de coeficientes sazonais autorregressivos referentes a estacionariedade;
- **D** o número de diferenças sazonais para tornar a série estacionária de acordo com o número de períodos s ; e
- **Q** como sendo o número de coeficientes sazonais de médias móveis referentes a invertibilidade.

Dessa forma, o modelo é dado por:

$$\phi(B)\Phi(B)[(1 - B)^d(1 - B^s)^D - \alpha]Z_t = \theta(B)\Theta(B)\epsilon_t \quad (13)$$

Segundo Ehlers (2007), muitas séries temporais contêm uma componente periódica sazonal que se repete a cada s observações ($s > 1$). Portanto, ao trabalhar com dados mensais ($s = 12$), particularmente espera-se que Z_t dependa de Z_{t-12} e talvez de Z_{t-24} , além de Z_{t-1} , Z_{t-2} , ...

Modelo autoregressivo de médias móveis sazonal com variáveis exógenas (SARIMAX)

Surgido de uma adaptação feita a partir do modelo SARIMA, o SARIMAX é um modelo de regressão linear denotado por $SARIMAX(p, d, q, s) \times (P, D, Q)$ e a equação do modelo é dada por:

$$\phi(B)\Phi(B)(1 - B)^D Z_t = \alpha + X\beta + \theta(B)\Theta(B^S)\epsilon_t, \quad (14)$$

em que X é uma matriz $n \times k$ de regressores não aleatórios e β um vetor de parâmetros $k \times 1$.

Função de autocorrelação (FAC)

De acordo com Montgomery, Jennings e Kulahci (2015), se uma série temporal é estacionária, de modo que W_t e W_{t+k} sejam duas observações, então sua distribuição de probabilidade conjunta é a mesma para qualquer período de tempo t e $t+k$ cujo são separados por um mesmo intervalo (lag) k .

Definimos a função de autocorrelação amostral (FAC), r_k , no lag k por:

$$r_k = \frac{\sum_{t=k+1}^n (W_t - \bar{W})(W_{t-k} - \bar{W})}{\sum_{t=1}^n (W_t - \bar{W})^2}, \quad k = 1, 2, \dots \quad (15)$$

e onde n é o número de observações.

Sabendo que para a realização de uma análise satisfatória, deve-se saber qual a distribuição de probabilidade de r_k . Assim, se n for grande suficiente, então $r_k \sim N(0, \frac{1}{n})$. Desta forma, encontrada a distribuição de r_k , pode-se utilizar de intervalos de confiança e testar hipóteses para verificação da hipótese nula da correlação.

Para testar as hipóteses em que os primeiros coeficientes de autocorrelação k são conjuntamente iguais a zero, utiliza-se a estatística de Ljung-Box:

$$Q_K = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k}, \quad (16)$$

na qual r_k^2 é a autocorrelação dos resíduos na defasagem k , m é o número de lags sendo testados e n o número de lags testados. O teste irá rejeitar a hipótese nula de que os resíduos não são correlacionados caso $Q > \chi_{(1-\alpha; q)}^2$, sendo $\chi_{(1-\alpha; q)}^2$ uma distribuição qui-quadrado com q graus de liberdade e nível de significância α (SALES, 2017).

Função de autocorrelação parcial (FACP)

A função de autocorrelação parcial (FACP) é denotada por ϕ_{kk} e mensura a correlação entre W_t e W_{t-k} após remover o efeito das variáveis intervenientes $W_{t-1}, W_{t-2}, \dots, W_{t-k+1}$ (SHUMWAY; STOFFER, 2017).

A FACP, assim como a FAC, também é usada para a identificação de modelos para determinada série temporal, auxiliando na identificação da estacionariedade da série.

$$\left\{ \begin{array}{l} W_t = \phi_{11}W_{t-1} + \epsilon_t \\ W_t = \phi_{11}W_{t-1} + \phi_{22}W_{t-2} + \epsilon_t \\ \vdots \\ W_t = \phi_{k1}W_{t-1} + \phi_{k2}W_{t-2} + \dots + \phi_{kk}W_{t-k} + \epsilon_t \end{array} \right. \quad (17)$$

De forma similar, a função de autocorrelação parcial no caso do n ser suficiente grande, tem-se que $\phi_{kk} \sim N(0, \frac{1}{n})$, para $k > p$.

Etapas do modelo SARIMA

De acordo com Walter et al. (2013), a metodologia Box-Jenkins é definida por um ciclo composto de três etapas iterativas: identificação do modelo, estimação dos parâmetros e verificação.

Identificação

Inicialmente, na fase de identificação, busca-se encontrar os valores apropriados de p , d e q . Porém, essa fase é bem complexa, já que tem-se uma dificuldade no momento de determinar esses parâmetros, pois ao determinar de forma errônea, os resultados serão comprometidos.

A identificação do modelo que melhor descreve a série temporal ocorre por intermédio da observação das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) (LIMA; CASTRO; CARTAXO, 2019).

Por ser mais simples, inicia-se determinando o parâmetro de integração (parâmetro d), pois pode-se aplicar algum teste de estacionariedade com propósito de encontrar a ordem de integração.

Teste de Estacionariedade de Dickey-Fuller

Considerando o modelo $Z_t = \rho Z_t + \epsilon_t$, onde ϵ_t é o ruído branco. Caso $\rho = 1$, a série Z_t possui uma raiz unitária e não é estacionária, entretanto, se $|\rho| < 1$ a série Z_t é estacionária, não possuindo raiz unitária. Segundo Dickey e Fuller (1979), o processo gerador da série Z_t é expresso por um destes modelos:

1. $\Delta Z_t = \alpha + \beta_t + \lambda_3 Z_{t-1} + \epsilon_t$
2. $\Delta Z_t = \alpha + \lambda_2 Z_{t-1} + \epsilon_t$
3. $\Delta Z_t = \lambda_1 Z_{t-1} + \epsilon_t$

em que, $\lambda_i = \rho - 1$, $\forall_i = 1, 2, 3$ e α e β sendo constantes a serem estimadas.

As hipóteses a serem testadas são apresentadas de acordo com a Tabela 1, em que $H_0 : \rho = 1$ equivale a $H_0 : \lambda = 0$.

Portanto, se ao menos umas das hipóteses apresentadas na Tabela 1 não for rejeitada, a série possui pelo menos uma raiz unitária, sendo assim não estacionária.

Tabela 1: Testes da raiz unitária de Dickey-Fuller

Modelos	H_0	Regras de decisão
1	$\lambda_3 = 0$ $(\alpha, \beta, \lambda_3) = (0, 0, 0)$ $(\alpha, \beta, \lambda_3) = (\alpha, 0, 0)$	$\tau_3 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada $\delta_2 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada $\delta_3 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
2	$\lambda_2 = 0$ $(\alpha, \lambda_2) = (0, 0)$	$\tau_2 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada $\delta_1 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
3	$\lambda_1 = 0$	$\tau_1 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada

Fonte: Adaptada de Dickey e Fuller (1979).

Análise Gráfica

Tendo a série estacionária definida, assim também como seu parâmetro d , o próximo passo é a determinação dos parâmetros p e q . Para isto, segundo Morettin e Tolo (2006), esta escolha é feita com base na análise gráfica das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). A Tabela 2 sintetiza as FAC's e FACP's para os modelos ARIMA.

Tabela 2: Comportamento da FAC e FACP para modelos ARIMA.

	FAC	FACP
AR(p)	Decaimento p/ zero (assintoticamente)	Decaimento p/ zero (abruptamente)
MA(q)	Decaimento p/ zero (abruptamente)	Decaimento p/ zero (assintoticamente)
ARMA(p,q)	Decaimento p/ zero (assintoticamente)	Decaimento p/ zero (assintoticamente)

Fonte: Adaptada de Shumway e Stoffer (2017).

Estimação

Após a estimação do modelo provisório para a série, o próximo passo é estimar seus parâmetros dos termos autorregressivos e dos termos de média móvel incluídos no modelo, podendo ser usado o procedimento iterativo de Mínimos Quadrados Condicionais ou o Método de Máxima Verossimilhança (MORETTIN; TOLOI, 2006).

Neste estudo será utilizado o método de Máxima Verossimilhança por ele maximizar a verossimilhança de um conjunto de observações, e ainda também devido a aplicação do software usado apresentar cálculos bem implementados das estimativas dos parâmetros.

A função adaptada de máxima verossimilhança é dada por Brockwell et al. (2016).

$$L(\phi_i, \theta_j, \sigma_\xi^2/w) = (2\pi\sigma_\xi^2)^{-\frac{n}{2}} \left| \sum \right|^{-\frac{n}{2}} e^{-\frac{1}{2}x' \Sigma^{-1} w}{\sigma_\xi^2} \quad (18)$$

Ao maximizar esta função obtemos as estimativas dos parâmetros ϕ_i , θ_j e σ_ξ^2 .

Validação

Com o modelo estimado o próximo passo é a verificação de sua adequação através do diagnóstico. De acordo com Lima, Castro, Cartaxo (2019), esta fase divide-se em duas etapas: análise dos resíduos e avaliação do modelo.

A análise de resíduos (ϵ_t) corrobora para verificar se os resíduos do modelo apresentam comportamento de um ruído branco, caso contrário, o processo deve-se iniciar novamente com a identificação de um novo modelo, seguido das mesmas etapas até que seja encontrado um modelo apropriado. Assim, os coeficientes de autocorrelação residual devem ser estatisticamente iguais a zero e, para testar esta condição, utiliza-se do teste de Ljung-Box.

Alguns modelos podem apresentar poucos parâmetros (modelos parcimoniosos) que possivelmente são correlacionados (ou não-significativos). Não existe uma solução específica para saber quantos parâmetros são necessários para o modelo, mas pode-se recorrer a *softwares* estatísticos que estimam utilizando de várias combinações de p e q , descartando-se as não-significativas. Para isso existem os seguintes critérios:

- **AIC** - Critério de Informação de Akaike

$$AIC = 2 \log \hat{L} + 2(p + q) \quad (19)$$

em que \hat{L} é o valor maximizado da verossimilhança.

- **BIC** - Critério de Informação Bayesiano

$$BIC = -2 \log \hat{L} + (p + q) \log M \quad (20)$$

Dessa forma, com a seleção e validação do modelo, a previsão dos valores futuros da série temporal modelada pode ser encontrada.

Avaliação de Previsão

Seja uma série temporal apresentada por T observações, dela são retiradas as n últimas observações, gerando n previsões para essa série, a partir de:

$$\epsilon_j = Z_j + \hat{Z}_j \quad (21)$$

em que, \hat{Z}_j é a previsão de Z_j .

A avaliação da capacidade de previsão do modelo estimado pode ser avaliado pelos erros apresentados a seguir, em que $k = T - n$.

1. **MPE** - Erro Percentual Médio

$$MPE = \left(\frac{1}{n} \sum_{j=k+1}^M \frac{\epsilon_j}{Z_j} \right) \times 100 \quad (22)$$

2. **MSE** - Erro Quadrático Médio

$$MSE = \frac{1}{n} \sum_{j=k+1}^M (\epsilon_j)^2 \quad (23)$$

3. **MAE** - Erro absoluto Médio

$$MAE = \frac{1}{n} \sum_{j=k+1}^M |\epsilon_j| \quad (24)$$

4. **TE** - Erro Total

$$TE = \sum_{j=k+1}^M \epsilon_j \quad (25)$$

5. **TAE** - Erro Absoluto Total

$$TAE = \sum_{j=k+1}^M |\epsilon_j| \quad (26)$$

6. **TPE** - Erro Percentual Total

$$TPE = \left(\frac{\sum_{j=k+1}^M \epsilon_j}{\sum_{j=k+1}^M Z_j} \right) \times 100 \quad (27)$$

Resultados

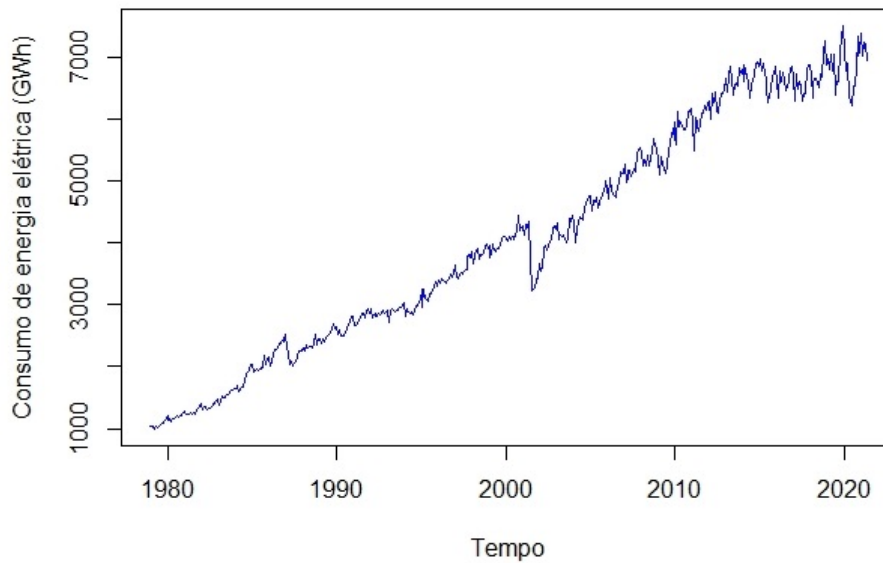
Nesta seção explicar-se-á as aplicações dos modelos para os dados do consumo de energia elétrica na região Nordeste do Brasil, modelo Holt-Winters e metodologia de Box-Jenkins. Os dados são referentes ao consumo de energia elétrica na região Nordeste do Brasil por unidade Gigawatts-hora (GWh), compreendidos entre janeiro de 1979 a maio de 2021. E disponíveis no site do Instituto de Pesquisa Econômica Aplicada (IPEA) <http://www.ipeadata.gov.br/>, totalizando 509 observações. Para avaliar o consumo de energia elétrica, por meio de modelos de séries temporais, foi utilizado o software R (R CORE TEAM, 2021).

Análise Exploratória de Dados

Na Figura 1 é apresentado o gráfico da série original do consumo de energia elétrica (em GWh) da região Nordeste do Brasil. Pode-se observar que, ao longo dos anos, o consumo de energia vem aumentando consideravelmente, havendo pequenas oscilações durante o período em estudo.

A transformação logarítmica foi necessária para estabilização da variância, preservando as propriedades dos dados da série. Pode-se observar o gráfico da série transformada por meio da Figura 2.

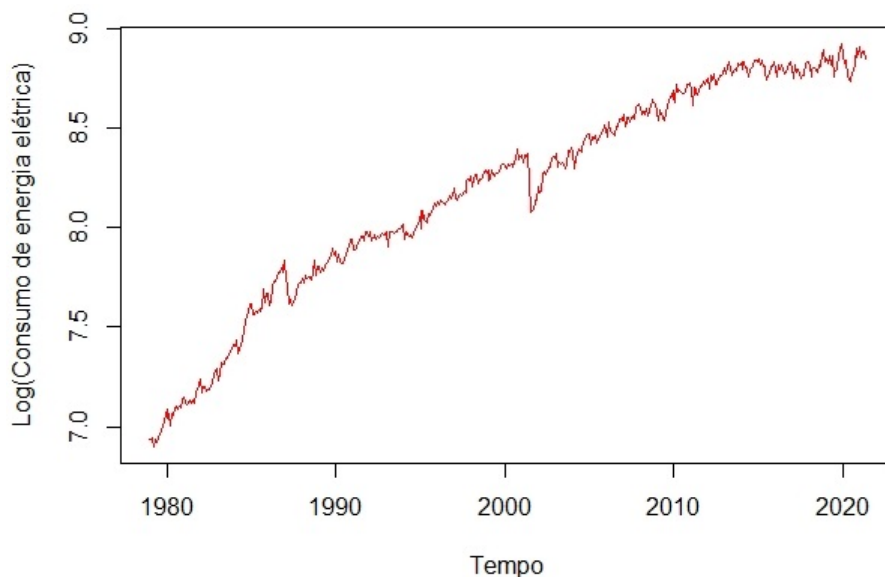
Figura 1: Série original do consumo de energia elétrica na região Nordeste no período de janeiro de 1979 a maio de 2021.



Fonte: Dos autores.

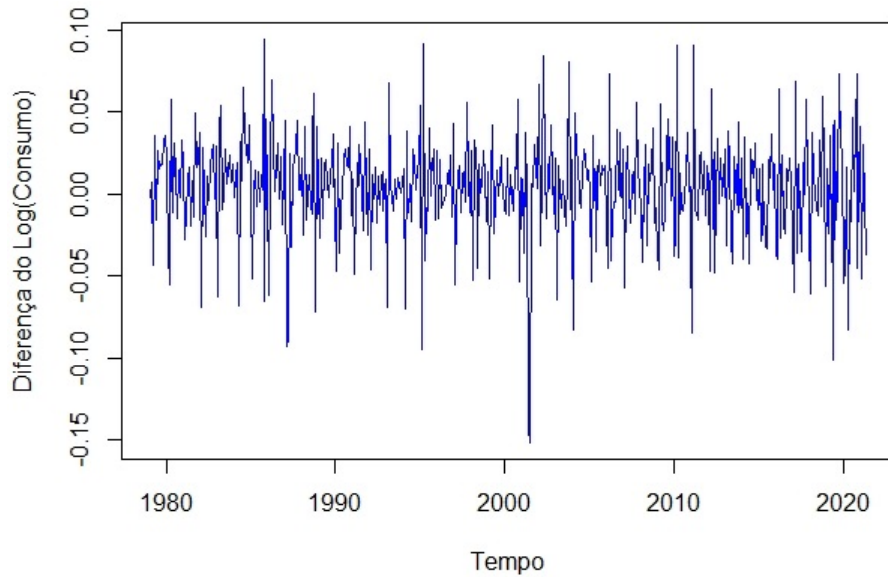
A tendência caracteriza um estágio não estacionário e, como o ajuste dos modelos pressupõe estacionariedade, a série foi diferenciada uma vez para eliminar esta componente, na qual comprova-se a estacionariedade pelo teste de Dickey-Fuller, que apresentou um $p\text{-valor} = 0,01$. O gráfico da série diferenciada é representado na Figura 3.

Figura 2: Série logaritmizada do consumo de energia elétrica na região Nordeste no período de janeiro de 1979 a maio de 2021.



Fonte: Dos autores.

Figura 3: Série diferenciada logaritmizada do consumo de energia elétrica na região Nordeste no período de janeiro de 1979 a maio de 2021.

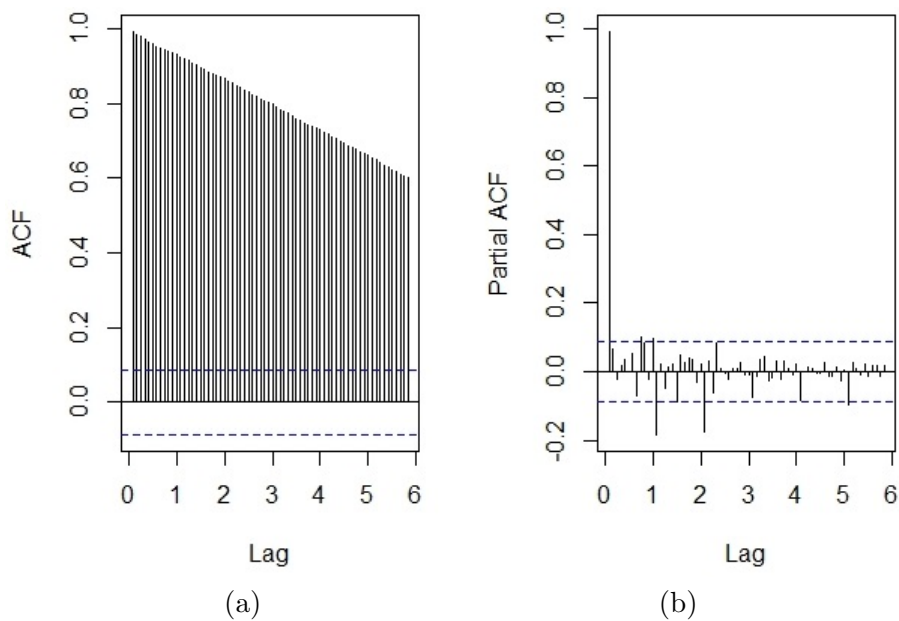


Fonte: Dos autores

A média amostral da série é 4070 (GWh) com desvio-padrão de 1880,093 (GWh). O maior valor observado é de 7515 (GWh), correspondente ao mês de dezembro de 2019, e o menor é de 992 (GWh), referente ao mês de abril de 1979.

A Figura 4 apresenta as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série. Na FAC observa-se um decaimento lento para zero evidenciando a não-estacionariedade da série. Já na FACP notam-se picos que ultrapassam o limite, indicando que existe uma forte correlação entre os meses.

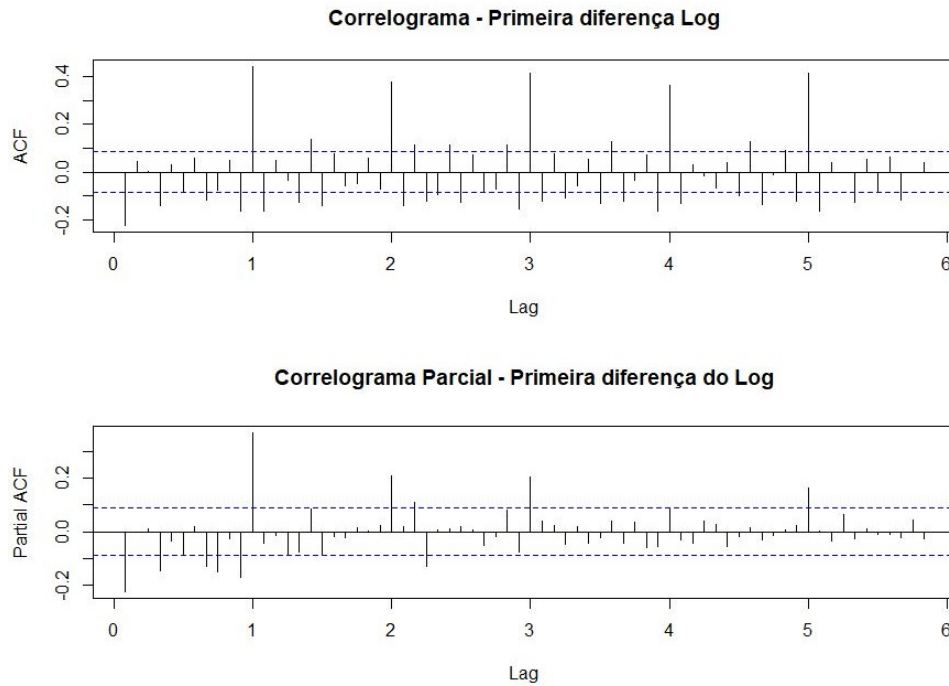
Figura 4: Funções de autocorrelação amostral (a) e autocorrelação parcial (b) da série.



Fonte: Dos autores

Para examinar a estacionariedade da série, nos correlogramas, aplicou-se a primeira diferença ($d = 1$) nos dados logaritmizados. Na Figura 5 observa-se que os gráficos FAC e FACP apresentam picos que ultrapassam o limite de confiança, indicando uma forte sazonalidade.

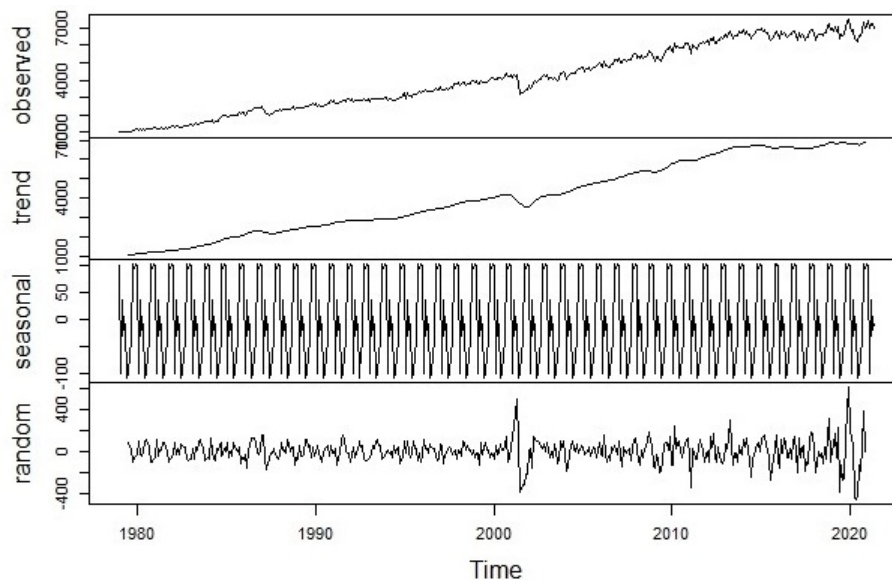
Figura 5: Funções de autocorrelação amostral e autocorrelação parcial da primeira diferença dos dados logaritmizados.



Fonte: Dos autores.

A Figura 6 apresenta a decomposição da série em componentes não observáveis da tendência, sazonalidade e aleatoriedade dos resíduos. Constata-se a tendência e a sazonalidade, ambas estocásticas.

Figura 6: Decomposição da série temporal do consumo de energia em componentes de sazonalidade, da tendência e aleatoriedade dos resíduos.



Fonte: Dos autores.

Modelagem

Na modelagem prossegue as aplicações dos modelos para os dados do consumo de energia elétrica na Região Nordeste do Brasil, possuindo valores referentes ao período de janeiro de 1979 a maio de 2021. Desta forma, serão consideradas previsões 1 e 8 passos à frente, sendo utilizado o modelo Holt-Winters, na forma do modelo aditivo e multiplicativo, e a metodologia de Box-Jenkins, nas formas SARIMA e SARIMAX.

Alisamento Exponencial de Holt-Winters

Como foi observada a presença de sazonalidade na série, então opta-se pelo uso dos algoritmos de suavização aditivo e multiplicativo de Holt-Winters para produzir previsões da dinâmica do consumo de energia. Com intuito de manter a escala original da série, foi modelada a série logaritimizada e suas previsões foram exponenciadas.

Um ponto importante é a escolha dos parâmetros de suavização que minimizam o valor do MSE (Erro Quadrático Médio) um passo à frente, critério utilizado para escolher os valores das constantes de suavização. Estes valores podem ser vistos na Tabela 3.

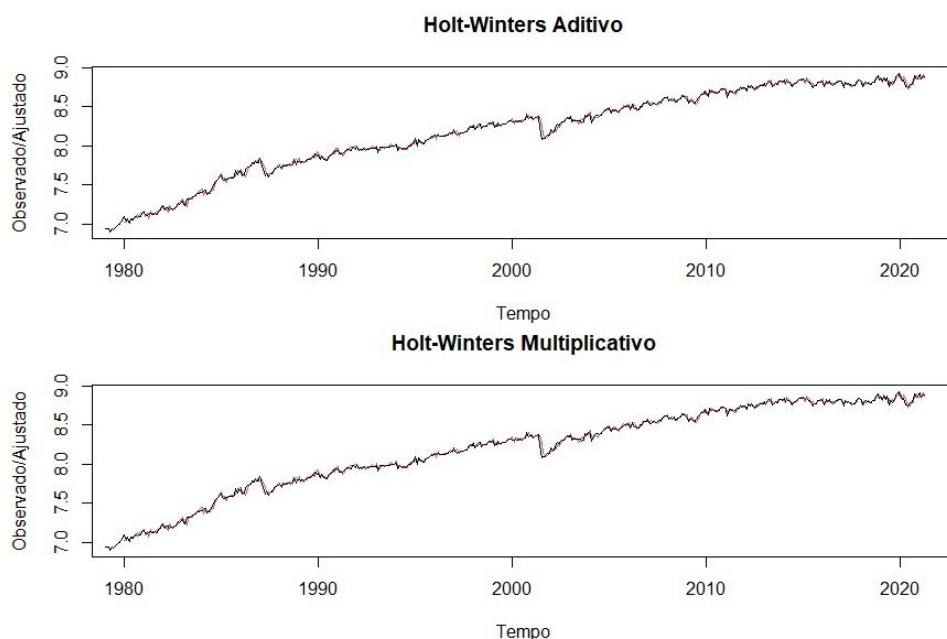
Tabela 3: Valores otimizados das constantes de suavização para Holt-Winters aditivo e multiplicativo.

Método	α	β	γ
HW aditivo	0.7778099	0.01057658	0.6198174
HW multiplicativo	0.7625259	0.01053796	0.6381249

Fonte: Dos autores.

Mediante a predição dos algoritmos aditivo e multiplicativo de Holt-Winters, um passo à frente, foram encontrados os valores de previsão de 7341,149 e de erro de 5,50 % para o aditivo e de 7352,415 com erro de 5,66 % para o multiplicativo. Os resultados das previsões foram superestimados. Através da Figura 7 pode-se observar que os dados previstos ficaram bem ajustados em relação aos dados observados.

Figura 7: Gráfico da série (linha preta) e seu ajuste de Holt-Winters (linha vermelha) aditivo e multiplicativo (1 passo à frente).



Fonte: Dos autores.

Posteriormente, ao verificar-se que o método de alisamento exponencial de Holt-Winters obteve um comportamento satisfatório, realizou-se uma previsão da série com 8 passos à frente, ou seja, até dezembro de 2021.

Tabela 4: Previsões adquiridas pelos algoritmos Holt-Winters Aditivo e Multiplicativo.

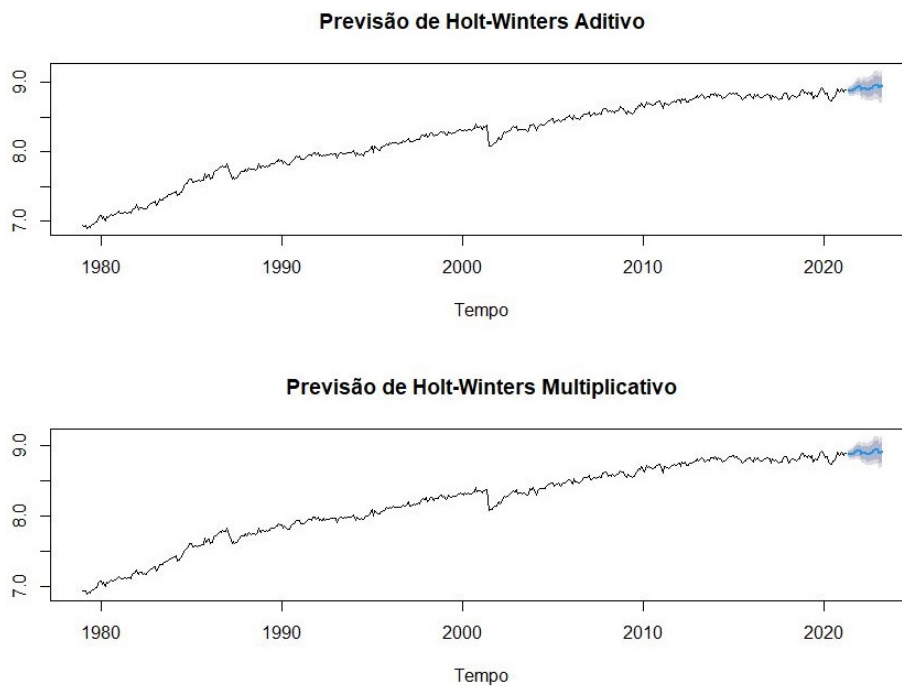
Mês/2021	HW aditivo	HW multiplicativo
Maio	7341,149	7352,415
Junho	7120,173	7132,324
Julho	7193,400	7214,184
Agosto	7246,825	7261,357
Setembro	7325,359	7341,844
Outubro	7647,918	7671,390
Novembro	7578,685	7594,833
Dezembro	7571,678	7588,017

Fonte: Dos autores.

Na Tabela 4 percebe-se que em junho e julho as previsões do consumo de energia elétrica apresentaram os menores valores, enquanto os três últimos meses apresentam os maiores. Observa-se também que o método de alisamento Holt-Winters multiplicativo apresentou valores superiores ao aditivo.

Na Figura 8 pode-se observar a série e suas previsões com o intervalo de confiança. As previsões são mostradas como uma linha azul, com os intervalos de previsão de 80% como uma área sombreada em cinza e os intervalos de previsão de 95% como uma área sombreada em cinza claro.

Figura 8: Gráfico de previsão de Holt-Winters aditivo e multiplicativo da série (8 passos à frente).



Fonte: Dos autores.

De acordo com Hyndman e Khandakar (2008), na função *ets* do pacote *forecast*, no software R (R CORE TEAM, 2021), algumas medidas de adequação são obtidas fornecendo teste de

precisão da previsão ajustada com base nos dados da série. Na Tabela 5 observa-se que o teste de previsão resultou em erros baixos, indicando uma boa capacidade de previsão para o modelo.

Tabela 5: Valores dos erros de previsão.

ME	RMSE	MAE	MPE	MAPE	MASE
-0.001490057	0.02461783	0.01746582	-0.01891604	0.2152695	0.2759566

Fonte: Dos autores.

Modelagem Box-Jenkins

Nesta etapa utilizou-se a modelagem Box-Jenkins, nos modelos SARIMA e SARIMAX, com intuito de gerar uma melhor previsão para o consumo de energia elétrica na região Nordeste do Brasil de maio/2021 a dezembro/2021. De acordo com Hyndman e Khandakar (2008), para encontrar um modelo SARIMA mais adequado e com melhor capacidade preditiva, utiliza-se a função *auto.arima*, contida no pacote *forecast*, no software R. Assim, foi encontrado o modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$, no qual se encontra uma previsão de 7195,104 (GWh) com erro relativo de aproximadamente 3,40%. Tem-se um valor predito acima do esperado, entretanto aproximado.

Com base em pesquisas anteriores para o mesmo tipo de modelagem, foi encontrada uma análise referente à previsão do consumo de eletricidade no Nordeste brasileiro, em que foi utilizado um modelo $SARIMA(1, 1, 0)(0, 0, 1)_{12}$, proposto por Junior et al. (2018). E este modelo forneceu uma previsão de 7152,963 (GWh), com um erro de previsão de 2,80%, sendo menor que o encontrado com a função *auto.arima*.

Sabe-se que existia uma grande probabilidade de ocorrer uma crise de falta de energia elétrica a partir de 1997, devido o setor elétrico ter sofrido uma forte e rápida reforma, pois apresentava desequilíbrio fiscal. Para evitar a crise, foi implementado, entre 1º de junho de 2001 a 1º de março de 2002, o racionamento de energia, com a meta de economizar 20% de energia elétrica (GOLDENBERG; PRADO, 2003).

Desta forma, aplicou-se uma variável *dummy*, com valores de 1 a partir de junho de 2001 e 0 para os meses anteriores, a fim de prever diferenciando os acontecimentos marcantes no setor elétrico brasileiro.

Estudando novamente o modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$ e o modelo base $SARIMA(1, 1, 0)(0, 0, 1)_{12}$, com o acréscimo da variável *dummy*, que resultará em modelos SARIMAX, calculou-se novamente as previsões.

Obteve-se uma previsão de 7194,673 (GWh), com erro relativo de aproximadamente 3,40% para o modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$, e uma previsão de 7152,915 (GWh), com erro relativo aproximadamente de 2,80% para o modelo base. Esses erros foram relativamente menores, mas bem aproximados aos erros dos modelos sem a variável *dummy*.

Como critério de escolha do melhor modelo, calculou-se o AIC (Critério de Informação de Akaike) e o BIC (Critério de Informação Bayesiano), que baseiam-se na soma dos quadrados dos erros (SQE).

Na Tabela 6 observa-se os valores referentes aos critérios AIC e BIC, assim como os erros relativos dos modelos:

Tabela 6: Valores referentes ao critério de informação dos modelos da série e erros relativos, 1 passo à frente.

	Modelo	AIC	BIC	Erro de Previsão (%)
1	$SARIMA(4, 1, 1)(2, 0, 0)_{12}$	-2198,445	-2164,617	3,407643
2	$SARIMA(1, 1, 0)(0, 0, 1)_{12}$ (base)	-2123,036	-2110,35	2,801994
3	$SARIMAX(4, 1, 1)(2, 0, 0)_{12}$	-2196,577	-2158,521	3,401458
4	$SARIMAX(1, 1, 0)(0, 0, 1)_{12}$ (base)	-2121,134	-2104,22	2,801311

Fonte: Dos autores.

Como a escolha do melhor modelo com base nos critérios de informação é o que apresenta os menores valores do AIC e BIC, nota-se que o modelo 1 apresenta um melhor ajuste, pois possui os menores valores para os critérios de informação. Porém o modelo 4 é o que apresenta o menor erro de previsão 1 passo à frente.

Calculados os critérios de escolha das modelagens, a análise foi refeita com objetivo de prever com 8 passos à frente, ou seja, de Maio/2021 até Dezembro/2021. As previsões podem ser vistas nas Tabela 7 e Tabela 8.

Tabela 7: Previsões dos modelos SARIMA e SARIMAX, 8 passos à frente, de Maio/2021 a Agosto/2021.

	Modelos	Maio/2021	Junho/2021	Julho/2021	Agosto/2021
	$SARIMA(4, 1, 1)(2, 0, 0)_{12}$	7195,104	6941,282	7148,305	7071,944
	$SARIMA(1, 1, 0)(0, 0, 1)_{12}$ (base)	7152,963	7195,296	7269,978	7293,656
	$SARIMAX(4, 1, 1)(2, 0, 0)_{12}$	7194,673	6940,606	7148,034	7071,208
	$SARIMAX(1, 1, 0)(0, 0, 1)_{12}$ (base)	7152,915	7195,286	7269,976	7293,665

Fonte: Dos autores.

Tabela 8: Previsões dos modelos SARIMA e SARIMAX, 8 passos à frente, de Setembro/2021 a Dezembro/2021.

	Modelos	Setembro/2021	Outubro/2021	Novembro/2021	Dezembro/2021
	$SARIMA(4, 1, 1)(2, 0, 0)_{12}$	7164,936	7485,181	7409,351	7482,147
	$SARIMA(1, 1, 0)(0, 0, 1)_{12}$ (base)	7391,589	7523,018	7395,543	7384,699
	$SARIMAX(4, 1, 1)(2, 0, 0)_{12}$	7164,084	7484,397	7408,227	7480,923
	$SARIMAX(1, 1, 0)(0, 0, 1)_{12}$ (base)	7391,622	7523,074	7395,564	7384,704

Fonte: Dos autores.

Tabela 9: Erros relativos de previsão dos modelos SARIMA e SARIMAX, 8 passos à frente, de Maio/2021 a Agosto/2021.

	Modelos	Maio/2021	Junho/2021	Julho/2021	Agosto/2021
	$SARIMA(4, 1, 1)(2, 0, 0)_{12}$	3,4076	-0,2403	2,7350	1,6376
	$SARIMA(1, 1, 0)(0, 0, 1)_{12}$ (base)	2,8020	3,4104	4,4837	4,8240
	$SARIMAX(4, 1, 1)(2, 0, 0)_{12}$	3,4014	-0,2500	2,7317	1,6270
	$SARIMAX(1, 1, 0)(0, 0, 1)_{12}$ (base)	2,8013	3,4102	4,4837	4,8242

Fonte: Dos autores.

Ao observar os erros de previsão nas Tabela 9 e Tabela 10, os modelos $SARIMA(4, 1, 1)(2, 0, 0)_{12}$ e $SARIMAX(4, 1, 1)(2, 0, 0)_{12}$ foram os que apresentaram os menores erros ao longo do tempo. Desta forma, escolheu-se modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$ como melhor modelo de previsão, pois apresentou também os melhores critérios de informação AIC e BIC.

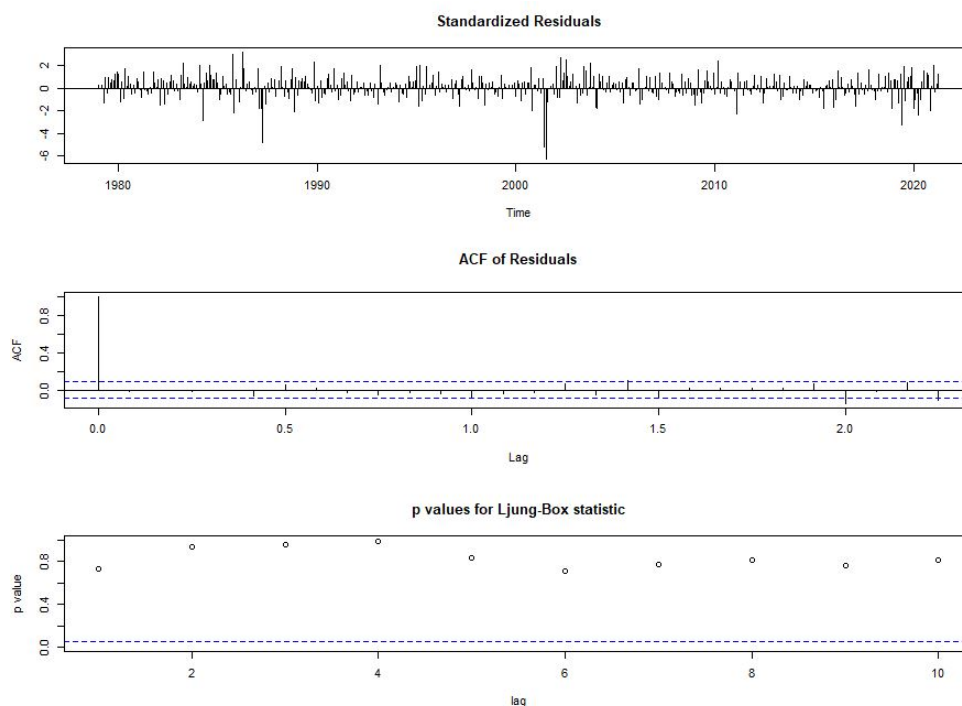
Tabela 10: Erros relativos de previsão dos modelos SARIMA e SARIMAX, 8 passos à frente, de Setembro/2021 a Dezembro/2021.

Modelos	Setembro/2021	Outubro/2021	Novembro/2021	Dezembro/2021
$SARIMA(4, 1, 1)(2, 0, 0)_{12}$	2,9741	7,5766	6,4868	7,5330
$SARIMA(1, 1, 0)(0, 0, 1)_{12}$ (base)	6,2315	8,1204	6,2883	6,1325
$SARIMAX(4, 1, 1)(2, 0, 0)_{12}$	2,9618	7,5653	6,4706	7,5154
$SARIMAX(1, 1, 0)(0, 0, 1)_{12}$ (base)	6,2320	8,1212	6,2886	6,1326

Fonte: Dos autores.

Na Figura 9 pode-se observar o gráfico de diagnóstico com os resíduos, a função de autocorrelação dos resíduos e os p-valores do teste de Ljung-Box para o modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$.

Figura 9: Gráfico do resíduo e sua função de autocorrelação, além dos p-valores do teste de Ljung-Box do modelo $SARIMA(4, 1, 1)(2, 0, 0)_{12}$.



Fonte: Dos autores.

O modelo apresenta resíduos estacionários, com média zero e variância constante. Em relação ao ACF, tem-se que todos os *lags* (exceto 0, que sempre é igual a 1) possuem valores dentro do intervalo de confiança.

Já os p-valores do teste de Ljung-Box estão todos acima de 5%, indicando que não se rejeita a hipótese nula de que as n primeiras auto-correlações dos erros são iguais a zero. Assim, o modelo proposto mostrou-se adequado.

Conclusão

Neste trabalho avaliou-se o consumo de energia na região Nordeste do Brasil através de modelagens, focando-se nas suas respectivas previsões. Para tais previsões, foram utilizados os algoritmos de Holt-Winters nas formas aditiva e multiplicativa e os modelos SARIMA e SARIMAX, ambos pela metodologia de Box-Jenkins. No modelo SARIMAX foi adicionada uma variável *dummy* assumindo o valor 1 a partir de junho de 2001 e 0 para os meses antecedentes, devido à crise de energia elétrica.

De forma geral, os modelos apresentaram previsões semelhantes, mas o modelo melhor ajustado foi o $SARIMA(4, 1, 1)(2, 0, 0)_{12}$, tendo como base os critérios de informação AIC e BIC. Por outro lado, o modelo base na forma SARIMAX, proposto por Junior et al. (2018), apresentou menor erro 1 passo à frente e os modelos de Holt-Winters aditivo e multiplicativo os maiores. Entretanto, os modelos $SARIMA(4, 1, 1)(2, 0, 0)_{12}$ e $SARIMAX(4, 1, 1)(2, 0, 0)_{12}$ apresentaram melhores erros ao longo do tempo. Desta forma, foi visto que a crise pouco influenciou o consumo de energia elétrica no Nordeste, pois a diferença entre os erros dos dois modelos foi mínima.

Como perspectiva futura pode-se expandir este estudo acerca do consumo de energia elétrica para as demais regiões do Brasil, possibilitando estabelecer um comparativo que permita analisar o comportamento do setor elétrico. Ademais, seria possível observar onde se situa o maior e/ou menor consumo de energia elétrica nas regiões brasileiras.

Referências

BOX, G. E. P. et al. Time series analysis. Forecasting and control. 5th ed. 5th ed..ed. [S.l.]: Hoboken, NJ: John Wiley & Sons, 2016. 712 p. ISSN 1940-6347. ISBN 978-1-118-67502-1/hbk.

BROCKWELL, P. J. et al. Introduction to time series and forecasting. [S.l.]: Springer, 2016.

CAMPOS, R. J. Previsão de séries temporais com aplicações a séries de consumo de energia elétrica. Universidade Federal de Minas Gerais, 2008.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, Taylor & Francis, v. 74, n. 366a, p. 427-431, 1979.

EHLERS, R. S. Análise de séries temporais. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná, 2007. Disponível em: <http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>. Acesso em: 16 de Abril de 2021.

ENERGÉTICA, E. D. P. Matriz Energética e Elétrica1. 2019. Disponível em: <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>. Acesso em: 13 de Agosto de 2021.

GOLDENBERG, J.; PRADO, L. T. S. Reforma e crise do setor elétrico no período fhc. Tempo social, SciELO Brasil, v. 15, p. 219-235, 2003.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for r. Journal of statistical software, v. 27, n. 1, p. 1-22, 2008.

IPEADATA. Banco de Dados: Energia elétrica referente ao consumo na região Nordeste (NE). 2021. Disponível em: <http://www.ipeadata.gov.br/Default.aspx>. Acesso em: 19 de Julho de 2021.

JUNIOR, L. M. C. et al. Previsão do consumo de eletricidade no nordeste brasileiro. Engevista, v. 20, n. 3, p. 408-423, 2018.

LIMA, J. E. C.; CASTRO, L. F. de; CARTAXO, G. A. A. Aplicação do modelo sarima na previsão de demanda no setor calçadista/application of the sarima model in the forecast for demand in the footwear sector. ID on line REVISTA DE PSICOLOGIA, v. 13, n. 46, p.

892?913, 2019.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction to time series analysis and forecasting. [S.l.]: John Wiley & Sons, 2015.

MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: Análise de séries temporais. [S.l.: s.n.], 2006. p. 538-538.

SALES, L. O. A. Testando a hipótese de passeio aleatório no mercado de ações brasileiro. Tese (Doutorado), 2017.

SCHMIDT, C. A. J.; LIMA, M. A. A demanda por energia elétrica no brasil. Revista brasileira de economia, SciELO Brasil, v. 58, p. 68?98, 2004.

SHUMWAY, R.; STOFFER, D. Time series analysis using the R Statistical Package. [S.l.]: free dog publishing, 2017.

R CORE TEAM. R: A language and environment for statistical computing (R Version 4.0. 3, R Foundation for Statistical Computing, Vienna, Austria, 2020). 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 12 de Agosto de 2021.

TIDRE, P. V. V.; BIASE, N. G. G.; SILVA, M. I. de S. Utilização dos modelos de séries temporais na previsão do consumo mensal de energia elétrica da região norte do brasil. Matemática e Estatística em Foco, v. 1, n. 1, p. 57-66, 2013.

WALTER, O. M. F. C. et al. Aplicação de um modelo sarima na previsão de vendas de motocicletas. Exacta, Universidade Nove de Julho, v. 11, n. 1, p. 77-88, 2013.