

Marlene Lamdin Carvalho

**Análise e Previsão uma Série Temporal de  
velocidade do vento em Campo Grande/MS  
usando modelos de Holt-Winters e  
Box-Jenkins**

Niterói - RJ, Brasil

07 de julho de 2022

**Marlene Lamdin Carvalho**

**Análise e Previsão uma Série  
Temporal de velocidade do vento em  
Campo Grande/MS usando modelos  
de Holt-Winters e Box-Jenkins**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Moisés Lima de Menezes

Niterói - RJ, Brasil

07 de julho de 2022

**Marlene Lamdin Carvalho**

**Análise e Previsão uma Série Temporal de  
velocidade do vento em Campo Grande/MS  
usando modelos de Holt-Winters e  
Box-Jenkins**

Monografia de Projeto Final de Graduação sob o título “*Análise e Previsão uma Série Temporal de velocidade do vento em Campo Grande/MS usando modelos de Holt-Winters e Box-Jenkins*”, defendida por Marlene Lamdin Carvalho e aprovada em 07 de julho de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Moisés Lima de Menezes**  
Departamento de Estatística – UFF

---

**Profa. Dra. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

---

**Profa. Dra. Mariana Albi De Oliveira Souza**  
Departamento de Estatística – UFF

Niterói, 07 de julho de 2022

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

C331a Carvalho, Marlene Lamdin  
Análise e Previsão uma Série Temporal de velocidade do  
vento em Campo Grande/MS usando modelos de Holt-Winters e Box-  
Jenkins / Marlene Lamdin Carvalho ; Moisés Lima de Menezes,  
orientador. Niterói, 2022.  
47 f. : il.

Trabalho de Conclusão de Curso (Graduação em  
Estatística)-Universidade Federal Fluminense, Instituto de  
Matemática e Estatística, Niterói, 2022.

1. Série temporal. 2. Modelagem. 3. Previsão. 4.  
Velocidade do vento. 5. Produção intelectual. I. Menezes,  
Moisés Lima de, orientador. II. Universidade Federal  
Fluminense. Instituto de Matemática e Estatística. III.  
Título.

CDD -

# Resumo

Este trabalho propôs duas classes de modelos de séries temporais para análise e previsão de velocidade do vento em uma estação anemométrica localizada no município de Campo Grande, em Mato Grosso do Sul. Para estas análises foram usados modelos de amortecimento exponencial de Holt-Winters e modelos da classe ARIMA de Box-Jenkins. Para tanto, testes de normalidade, estacionariedade e análises das funções de autocorrelação e autocorrelação parcial foram feitos para identificação dos parâmetros dos modelos, além de criteriosas análises de significâncias dos parâmetros estimados e estudos das estatísticas de aderência e de análise dos resíduos. Para previsão, foram usadas amostras treino e teste. Por fim, o modelo mais adequado para esta série temporal obtido foi o modelo de Holt-Winters sem tendência e com sazonalidade aditiva.

Palavras-chave: Modelos de séries temporais. Velocidade do vento. Box-Jenkins. Holt-Winters. Previsão.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 10
1.1	Motivação . . . . .	p. 10
1.2	Revisão Bibliográfica . . . . .	p. 10
1.3	Proposta . . . . .	p. 11
1.4	Objetivos . . . . .	p. 12
1.5	Organização . . . . .	p. 12
<b>2</b>	<b>Materiais e Métodos</b>	p. 13
2.1	Conjunto de dados . . . . .	p. 13
2.2	Séries Temporais . . . . .	p. 14
2.2.1	Componentes de uma Série Temporal . . . . .	p. 14
2.2.2	Processos Estocásticos . . . . .	p. 15
2.2.3	Estacionariedade . . . . .	p. 15
2.2.4	Ruído Branco . . . . .	p. 16
2.2.5	Operador de Defasagem . . . . .	p. 16
2.2.6	Operador de Diferenças . . . . .	p. 16
2.2.7	FAC . . . . .	p. 16
2.2.8	FACP . . . . .	p. 17
2.2.9	Teste de Normalidade . . . . .	p. 19

2.2.9.1	Teste de normalidade de Shapiro-Wilk . . . . .	p. 19
2.2.10	Teste de Raiz Unitária . . . . .	p. 20
2.2.10.1	Teste de Dickey-Fuller . . . . .	p. 20
2.2.10.2	Teste de Dickey-Fuller Aumentado . . . . .	p. 21
2.3	Modelos de Holt-Winters . . . . .	p. 21
2.3.1	Modelo Sazonal Aditivo . . . . .	p. 22
2.3.2	Modelo Sazonal Multiplicativo . . . . .	p. 22
2.4	Modelos de Box-Jenkins . . . . .	p. 23
2.4.1	Modelo $AR(p)$ : . . . . .	p. 23
2.4.2	Modelo $MA(q)$ : . . . . .	p. 24
2.4.3	Modelo $ARMA(p, q)$ : . . . . .	p. 24
2.4.4	Definição dos parâmetros $p$ e $q$ : . . . . .	p. 24
2.4.5	Modelo $ARIMA(p, d, q)$ : . . . . .	p. 26
2.4.6	Modelo $SARIMA(p, d, q) \times (P, D, Q)_S$ . . . . .	p. 26
2.4.7	Estimação dos parâmetros dos modelos de Box-Jenkins . . . . .	p. 27
2.5	Diagnóstico do Modelo . . . . .	p. 28
2.5.1	Medidas de aderência . . . . .	p. 28
2.5.1.1	MAPE (Erro Percentual Absoluto Médio) . . . . .	p. 28
2.5.1.2	MAD (Desvio Médio Absoluto) . . . . .	p. 28
2.5.1.3	Coeficiente de determinação ( $R^2$ ) . . . . .	p. 29
2.5.1.4	Critério de informação bayesiano (BIC) . . . . .	p. 29
2.5.1.5	RMSE (Raiz Quadrada do Erro Quadrático Médio) . . . . .	p. 29
2.5.1.6	GMRAE (Média Geométrica do Erro Relativo Absoluto) . . . . .	p. 30
2.5.2	Análise dos Resíduos . . . . .	p. 30
2.6	Fluxograma da Proposta . . . . .	p. 31
<b>3</b>	<b>Análise dos Resultados</b>	<b>p. 33</b>

3.1	Análise descritiva . . . . .	p. 33
3.2	Modelagem de Holt-Winters . . . . .	p. 35
3.3	Modelagem de Box-Jenkins . . . . .	p. 36
3.3.1	Teste de Normalidade: . . . . .	p. 37
3.3.2	Teste de Raiz Unitária: . . . . .	p. 37
3.3.3	Análise dos Correlogramas: . . . . .	p. 38
3.4	Melhores modelos . . . . .	p. 41
<b>4</b>	<b>Conclusões</b>	p. 45
	<b>Referências</b>	p. 46



# Lista de Figuras

1	Média Mensal de Velocidade do Vento - Estação Campo Grande/MS - jan/2006 a out/2020. . . . .	p. 14
2	Exemplo de correlograma da FAC. . . . .	p. 17
3	Exemplo de correlograma da FACP. . . . .	p. 18
4	FAC e FACP de um modelo AR(1). . . . .	p. 25
5	FAC e FACP de um modelo MA(1). . . . .	p. 25
6	FAC e FACP de um modelo ARMA(1,1). . . . .	p. 26
7	Exemplo de correlograma da FAC dos resíduos. . . . .	p. 30
8	Fluxograma da proposta. . . . .	p. 32
9	Box-plot da média mensal de velocidade do vento por mês. . . . .	p. 34
10	Gráfico da série com transformação logarítmica. . . . .	p. 37
11	Correlograma do log série de velocidade do vento em média mensal. . . . .	p. 38
12	Correlograma da série após diferença sazonal. . . . .	p. 39
13	FAC dos resíduos do modelo de Holt-Winters Sem tendência e com sa- zonalidade aditiva. . . . .	p. 42
14	Série original (preta, linha mais fina) x série ajustada (vermelha) e pre- visão (linha vermelha após final da linha preta) usando o modelo escolhido.	p. 43

# Lista de Tabelas

1	Comportamento das funções de autocorrelação (FAC) e autocorrelação parcial(FACP) nos processos AR, MA e ARMA. . . . .	p. 25
2	Algumas medidas descritivas da Série de Velocidades do Vento da Estação Anemométrica de Campo Grande/MS - jan/2006 a out/2020. . . . .	p. 34
3	Estimação dos parâmetros dos modelos de Holt-Winters ajustados. . . . .	p. 35
4	Medidas de aderência para os modelos de Holt-Winters com a série original (Análise na amostra treino). . . . .	p. 36
5	Medidas de aderência para os modelos de Holt-Winters com a série original (Análise na amostra teste). . . . .	p. 36
6	Parâmetros estimados nos modelos de Box-Jenkins, seus coeficientes e suas significâncias (Amostra treino). . . . .	p. 40
7	Medidas de aderência para os modelos de Box-Jenkins com a série original com transformação logarítmica (Análise na amostra treino). . . . .	p. 41
8	Medidas de aderência para os modelos de Box-Jenkins com a série original com transformação logarítmica (Análise na amostra teste). . . . .	p. 41
9	Medidas de aderência para os melhores modelos de Holt-Winters e Box-Jenkins . . . . .	p. 42
10	Valores previstos pelo modelo de Holt-Winters Sem tendencia e com sazonalidade aditiva . . . . .	p. 44

# 1 Introdução

## 1.1 Motivação

Segundo a Câmara de Comercialização de Energia Elétrica (CCEE) as capacidades de geração de energia das fontes hidráulica, térmica, solar fotovoltaica e eólica são de 64%, 22%, 11% e 3% respectivamente. As fontes de energia solar fotovoltaica e eólica são fontes limpas, porém não são muito exploradas. Juntas, suas capacidades atuais são de apenas 14% (CCEE, 2021).

Durante o verão os níveis dos reservatórios ficam mais baixos, o que compromete a geração de energia da fonte hidráulica e ocasiona o aumento da utilização de energia gerada pela fonte térmica. Além de aumentar o custo da produção de energia essa fonte polui o meio ambiente, indo contra o Objetivo de Desenvolvimento Sustentável 7 (ODS 7), Energia Limpa e Acessível (NASCIMENTO, 2017; ONU, 2022).

O americano Charles Brush (1849–1929) desenvolveu, com intuito de gerar energia, a primeira turbina eólica, mas foi o dinamarquês Poul la Cour (1846–1908) o precursor dos modernos aerogeradores (DALMAZ *et. al*, 2008). A energia eólica é oriunda da transformação da energia cinética do vento em energia mecânica e posteriormente em energia elétrica (PICOLO *et. al*, 2014).

Os modelos de séries temporais são ferramentas estatísticas ideais para gerar previsões a fim de se fazerem planejamentos estratégicos, com o objetivo de reduzir os custos e evitar momentos de crise energética como os apagões.

## 1.2 Revisão Bibliográfica

Camelo *et. al* (2017) utiliza os modelos de séries temporais Auto Regressivo Integrado de Médias Móveis (ARIMA), Holt-Winters (HW) e Redes Neurais Artificiais (RNA) na previsão de geração de energia eólica. Comparando os modelos com o objetivo de iden-

tificar o que modela melhor a série de velocidade do vento, foi observado que para as regiões de Fortaleza e São Luís o modelo RNA aderiu melhor, apresentando menor erro percentual, e para Parnaíba o modelo ARIMA gerou os menores erros.

Kavasseri e Seetharaman (2008) apontam a importância de obter previsões acuradas de velocidade do vento para o dia seguinte, o modelo proposto é o f-ARIMA onde as previsões dos dados são estimadas em conjunto com a curva de potência de uma turbina em funcionamento.

Yatiyana *et. al* (2017) utilizam o modelo Auto Regressivo Integrado de Médias Móveis para estimar os modelos de previsão para a geração de energia. Neste trabalho são analisadas para a construção dos modelos a velocidade e a direção do vento, com o propósito de melhorar as previsões.

Pereira *et. al* (2017) estimam a velocidade do vento na localidade de Palmeira dos Índios (AL) a partir da metodologia de Box-Jenkins, onde foi adotada para analisar e realizar as previsões da série temporal de velocidade do vento o modelo Auto-Regressivo Integrado de Média Móvel (ARIMA).

## 1.3 Proposta

Este trabalho propõe a modelagem de uma série temporal mensal de velocidade do vento para gerar previsões com horizonte de até 1 ano, tendo o intuito de avaliar a possibilidade de implementação de um parque eólico em Campo Grande, estado de Mato Grosso do Sul. Serão usados os modelos de Holt-Winters e Box-Jenkins onde serão feitos testes de raiz unitária, testes de normalidade, análises de resíduos e das estatísticas de aderência MAPE, BIC, RMSE, GMRAE e  $R^2$ . Os dados serão divididos em amostra treino e amostra teste para avaliar a capacidade preditiva dos modelos a fim de selecionar o modelo mais adequado para previsão. Para isso, as 12 últimas observações serão desconsideradas para a modelagem e serão utilizadas como parâmetro para verificar a capacidade preditiva, uma vez que a previsão gerada poderá ser comparada e medida com dados originais.

Para a realização dos testes acima citados, serão utilizados os programas R (R core team, 2021) e Gretl (The gretl team, 2021). Análise e modelagens serão feitas pelo programa Forecast Pro for Windows (FPW, 1997). Gráficos e tabelas serão gerados com informações a partir dos programas R e Microsoft Excel (Microsoft Corporation, 2018). Análise dos resíduos serão feitos pelo Gretl e FPW e as estatísticas de aderência serão

feitas com o auxílio do FPW.

## 1.4 Objetivos

Este trabalho tem como objetivo geral avaliar a possibilidade de implementação de um parque eólico na estação anemométrica de Campo Grande, no Mato Grosso do Sul utilizando modelos de séries temporais. Os objetivos específicos são:

- avaliar a capacidade preditiva dos modelos utilizados,
- verificar qual dos modelos estudados é o mais adequado para este conjunto de dados e
- gerar previsões para o horizonte de 1 ano.

## 1.5 Organização

Este trabalho está organizado em 4 Capítulos. No Capítulo 1, é feita uma contextualização introdutória do trabalho, apresentadas as propostas e os objetivos. No Capítulo 2, serão apresentados os materiais e métodos utilizados nas análises. No Capítulo 3, serão analisados os resultados e no Capítulo 4, serão apresentadas as conclusões desta pesquisa.

## 2 Materiais e Métodos

Este Capítulo apresenta os elementos necessários para se alcançar os objetivos deste trabalho. O embasamento teórico e o conjunto de dados são apresentados aqui.

### 2.1 Conjunto de dados

Para este trabalho estão sendo considerados os dados de velocidade do vento de uma estação anemométrica localizada na cidade de Campo Grande, no estado de Mato Grosso do Sul.

Os dados foram obtidos do Instituto Nacional de pesquisa espacial (INPE), disponíveis no site do instituto na aba de dados históricos (<http://sinda.crn.inpe.br/PCD/SITE/novo/site/index.php>). Estão sendo utilizados os dados da Estação Anemométrica 31950, localizada no município de Campo Grande, MS, com localização geográfica em Latitude: -20.502, Longitude: -54.619, à 50 metros de altitude acima do nível do mar.

Os dados de médias mensais foram obtidos calculando a média das observações coletadas em cada mês. Portanto, os dados são de médias mensais de velocidade do vento em metros por segundo (m/s), variando de janeiro de 2006 a outubro de 2020. A Figura 1 apresenta o comportamento da série temporal neste intervalo de tempo.

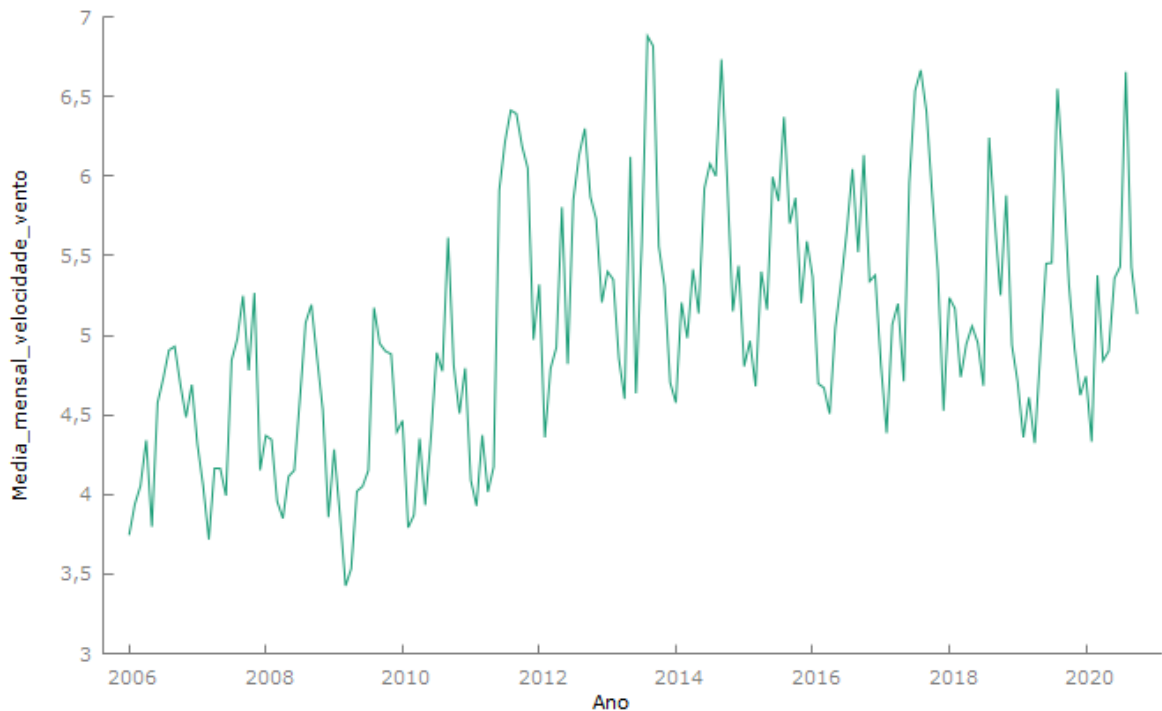


Figura 1: Média Mensal de Velocidade do Vento - Estação Campo Grande/MS - jan/2006 a out/2020.

Fonte: Elaboração própria

## 2.2 Séries Temporais

Uma série temporal é um conjunto de observações medidas ao longo do tempo em períodos regulares (horas, dias, meses, etc) (MORETTIN e TOLOI, 2018).

### 2.2.1 Componentes de uma Série Temporal

Uma série temporal pode ser decomposta em 4 componentes principais:

- **Tendência Linear:** Mostra a evolução da série ao longo prazo desconsiderando qualquer efeito de curto prazo sendo causado por flutuações ou pontos irregulares (outliers) da série.
- **Sazonalidade:** É representada pelo comportamento repetitivo da série no mesmo intervalo de tempo. Geralmente são picos ou vales bem definidos naquele período. Por exemplo, muitas séries temporais sofrem o efeito da sazonalidade devido a fatores climáticos (Exemplo: verão x inverno).

- **Fator Cíclico:** Diferentemente da sazonalidade, este fator se repete num período maior que um ano. Embora seja menos comum de aparecer nas séries temporais, ele costuma aparecer em séries de PIB e entre outras séries econômicas de longo prazo.
- **Irregular:** É o componente não explicado pelas componentes descritas acima. Pode ser aleatório ou não aleatório, mas pontual devido a falha humana ou evento extraordinário.

Um exemplo de série temporal linear com componente aleatório é apresentado em (2.1).

$$y_t = a + bx_t + \varepsilon_t, \quad (2.1)$$

onde que  $a + bx_t$  é a componente determinística e  $\varepsilon_t$  é a componente estocástica.

## 2.2.2 Processos Estocásticos

Seja  $\tau$  um conjunto arbitrário. Um processo estocástico é uma família  $Z = Z(t)$ ,  $t \in \tau$ , tal que, para cada  $t \in \tau$ ,  $Z(t)$  é uma variável aleatória. É possível definir um processo estocástico como uma família de variáveis aleatórias (v.a.) ao longo do tempo e uma série temporal como uma realização do processo estocástico. (MORETTIN e TOLOI, 2018).

## 2.2.3 Estacionariedade

Um processo estocástico é dito estacionário se ele mantém suas propriedades estatísticas ao longo do tempo. Sobre a forma de estacionariedade, dizemos que os processos podem ser:

**Fortemente estacionários ou estritamente estacionários:** se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,

$$F(Z_1, \dots, Z_n; t_1 + \tau, \dots, t_n + \tau) = F(Z_1, \dots, Z_n; t_1, \dots, t_n), \quad (2.2)$$

para quaisquer  $t_1, \dots, t_n, \tau$  de  $\tau$ . (MORETTIN e TOLOI, 2018).

**Fracamente estacionários ou estacionários de segunda ordem:** se, e somente se,

- (i)  $E(Z(t)) = \mu(t) = \mu$ , constante, para todo  $t \in \tau$  ;
- (ii)  $E(Z^2(t)) < \infty$ , para todo  $t \in \tau$  ;



(iii)  $Cov(Z(t_1), Z(t_2))$  é uma função de  $|t_1 - t_2|$ .  
(MORETTIN e TOLOI, 2018).

### 2.2.4 Ruído Branco

Dizemos que  $\varepsilon_t$ ,  $t \in \mathbb{Z}$  é um ruído branco se as variáveis aleatórias  $\varepsilon_t$  são não correlacionadas, isto é,  $Cov(\varepsilon_t, \varepsilon_s) = 0$ ,  $t \neq s$ , com média zero e variância finita e constante.

$$\varepsilon_t \sim RB(0, \sigma^2) \quad (2.3)$$

(MORETTIN e TOLOI, 2018).

### 2.2.5 Operador de Defasagem

O operador de defasagem (translação para o passado)  $B$  é definido por:

$$B^k Z_t = Z_{(t-k)}. \quad (2.4)$$

(MORETTIN e TOLOI, 2018).

### 2.2.6 Operador de Diferenças

O operador de diferenças  $\Delta$  é definido por:

$$\Delta^d Z_t = \Delta^{d-1} Z_t - \Delta^{d-1} Z_{(t-1)}. \quad (2.5)$$

(MORETTIN e TOLOI, 2018).

### 2.2.7 FAC

A função de autocorrelação (FAC) mede a autocorrelação contra a defasagem, sendo estimada por (2.5)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=k+1}^n (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}, \quad (2.6)$$

onde  $\bar{Z}$  é a média da série temporal  $Z_t$  e  $k$ , a defasagem da autocorrelação.

Na Figura 2 temos um exemplo de gráfico da função de autocorrelação (FAC), um correlograma.

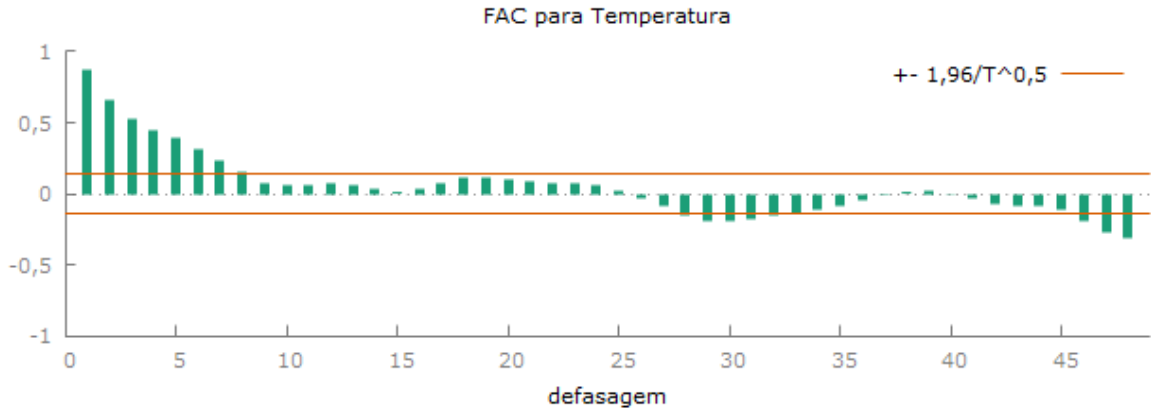


Figura 2: Exemplo de correlograma da FAC.  
Fonte: Elaboração própria

### 2.2.8 FACP

Para definir a função de autocorrelação parcial (FACP), vamos chamar de  $\phi_{ki}$  o  $i$ -ésimo coeficiente de um modelo AR( $k$ ) dessa forma  $\phi_{kk}$  é o último coeficiente. A partir de  $\rho_i = \phi_{k1}\rho_{i-1} + \phi_{k2}\rho_{i-2} + \dots + \phi_{kk}\rho_{i-k}$ ,  $i = 1, \dots, k$ , obtemos as equações de Yule-Walker:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{k-2} \\ \vdots & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.7)$$

Resolvendo estas equações sucessivamente para  $k = 1, 2, 3, \dots$  obtemos

$$\phi_{11} = \rho_1, \quad (2.8)$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad (2.9)$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \quad (2.10)$$

em geral,

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|}, \quad (2.11)$$

onde  $P_k$  é a matriz de autocorrelações e  $P_k^*$  é a matriz  $P_k$  com a última coluna substituída pelo vetor de autocorrelações.

A quantidade  $\phi_{kk}$ , é chamada função de autocorrelação parcial (MORETTIN e TOLLOI, 2018).

A Figura 3 contém um exemplo de gráfico da função de autocorrelação parcial (FACP), um correlograma.

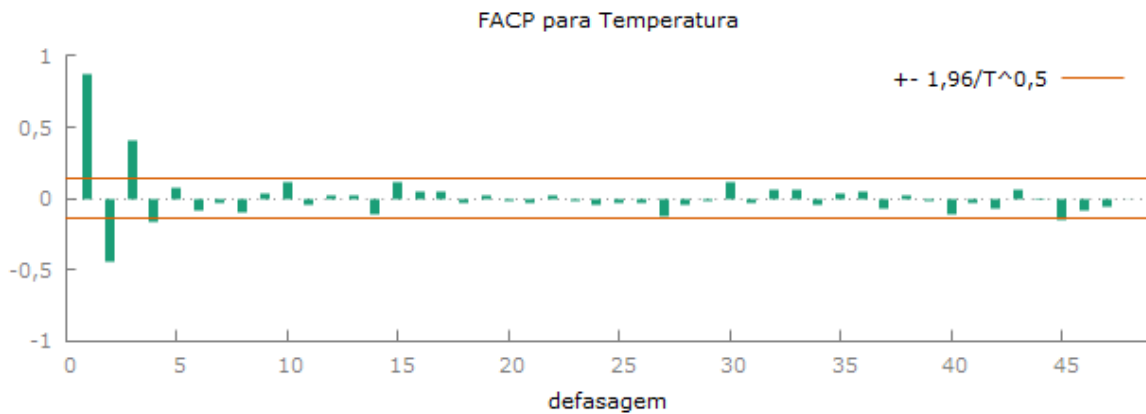


Figura 3: Exemplo de correlograma da FACP.

Fonte: Elaboração própria

## 2.2.9 Teste de Normalidade

Esse teste é utilizado para avaliar se a variável aleatória (v.a.) possui distribuição normal com parâmetros  $\mu$  e  $\sigma^2$ .

O teste de normalidade é utilizado para avaliar se os dados seguem uma distribuição normal. Ele considera as seguintes hipóteses:

$H_0$ : Os dados são provenientes de uma distribuição normal; e

$H_1$ : Os dados não são provenientes de uma distribuição normal.

### 2.2.9.1 Teste de normalidade de Shapiro-Wilk

O teste de normalidade de Shapiro-Wilk, possui a seguinte estatística de teste:

$$W = \frac{(\sum_{t=1}^n a_t - Z_{(t)})^2}{\sum_{t=1}^n (Z_t - \bar{Z})^2}, \quad (2.12)$$

onde

$$Z_{(t)} \quad (2.13)$$

é a  $i$ -ésima estatística de ordem, ou seja, o  $i$ -ésimo menor número da amostra,

$$\bar{Z} = \frac{(Z_1 + \dots + Z_n)}{n} \quad (2.14)$$

é a média da amostra,

$$a_t \quad (2.15)$$

são coeficientes dados por:

$$(a_1, \dots, a_n) = \frac{m^\top V^{-1}}{C}, \quad (2.16)$$

onde  $C$  é uma norma do vetor:

$$C = \|V^{-1}m\| = (m^\top V^{-1}V^{-1}m)^{1/2} \quad (2.17)$$

e o vetor  $m$ ,

$$m = (m_1, \dots, m_n)^\top \quad (2.18)$$

é feito dos valores esperados das estatísticas de ordem de variáveis aleatórias independentes e identicamente distribuídas (iid), onde a partir de uma distribuição normal padrão é gerada sua amostra. Finalmente,  $V$  é a matriz de covariância dessas estatísticas de ordem normal (Shapiro e Wilk, 1965).

### 2.2.10 Teste de Raiz Unitária

Considere o modelo a seguir,

$$Z_t = \rho Z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (2.19)$$

Se na equação 2.19,  $\rho = 1$  temos que  $Z_t$  depende de  $Z_{t-1}$  o que indica alta correlação entre eles, portanto  $Z_t$  é um processo não estacionário. Se  $\rho = 1$  realmente, a variável  $Z_t$  possui raiz unitária.

Se isolarmos  $\varepsilon_t$  e considerarmos  $\rho = 1$  na equação 2.20 teremos:

$$Z_t - 1Z_{t-1} = \varepsilon_t \quad (2.20)$$

tomando  $Z_t - 1Z_{t-1}$  por  $\Delta Z_t$  temos

$$\Delta Z_t = \varepsilon_t \quad (2.21)$$

Pela equação 2.21 temos que a série  $\Delta Z_t$  é um ruído branco, e sabe-se que uma série ruído branco é estacionária. Portanto, a série  $Z_t$  não é estacionária mas a série de primeira diferença é estacionária. Então, as hipóteses serão:

$H_0$ :  $\rho = 1$ , a série  $Z_t$  possui raiz unitária ( $Z_t$  não é estacionária, mas  $\Delta Z_t$  é estacionária);

e

$H_1$ :  $\rho \neq 1$ , a série  $Z_t$  não possui raiz unitária.

#### 2.2.10.1 Teste de Dickey-Fuller

O Teste de Raiz Unitária e Dickey-Fuller (DF) é utilizado para testar estacionariedade dos dados. Partindo de  $\Delta Z_t = Z_t - Z_{t-1}$  e utilizando  $Z_t = \rho Z_{t-1} + \varepsilon_t$  temos:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.22)$$

$$\Delta Z_t = (\rho Z_{t-1} + \varepsilon_t) - Z_{t-1} \quad (2.23)$$

$$\Delta Z_t = (\rho Z_{t-1} - Z_{t-1}) + \varepsilon_t \quad (2.24)$$

$$\Delta Z_t = (\rho - 1)Z_{t-1} + \varepsilon_t \quad (2.25)$$

Vamos chamar  $(\rho - 1)$  de  $\beta$ , obtendo assim

$$\Delta Z_t = \beta Z_{t-1} + \varepsilon_t \quad (2.26)$$

$\beta$  é o operador de primeira diferença. Se  $\beta = 0$  então  $\Delta Z_t = \varepsilon_t$  e a primeira diferença da série  $Z_t$  é uma série estacionária pois  $\varepsilon_t \sim RB(0, \sigma^2)$ . Ao incluir uma constante no modelo 2.26, ficamos com o seguinte modelo:

$$\Delta Z_t = \beta_0 + \beta_1 Z_{t-1} + \varepsilon_t \quad (2.27)$$

Foi demonstrado por Dickey e Fuller que se  $\beta_1 = 0$  o valor estimado t do coeficiente  $Z_{t-1}$  da equação 2.27 segue a estatística  $\tau = \frac{\widehat{\beta}_1}{\sqrt{\text{var}(\widehat{\beta}_1)}}$ , segundo as simulações de Monte Carlo. Essa estatística é analisada a partir da  $\Delta Z_t$ . Os valores críticos de  $\tau$  tabulados para os níveis de significância 0,01, 0,05 e 0,10 são dados respectivamente, por -3,43, -2,86 e -2,57. Como o teste de Dickey-Fuller é monocaudal, a decisão de teste é dada comparando os valores absolutos da estatística  $\tau$  com os valores críticos tabulados de DF. Logo, se  $|\tau| \geq$  valor crítico DF  $\Rightarrow$  Rejeita  $H_0$ .

Portanto as hipóteses do teste de Dickey-Fuller serão:

$H_0$ :  $\beta_1 = 0 \approx \rho = 1$ , a série  $Z_t$  possui raiz unitária ( $Z_t$  não é estacionária, mas  $\Delta Z_t$  é estacionária); e

$H_1$ :  $\beta_1 < 0 \approx \rho < 1$ , a série  $Z_t$  não possui raiz unitária.

### 2.2.10.2 Teste de Dickey-Fuller Aumentado

O Teste de Dickey-Fuller Aumentado é usado quando os  $\varepsilon_t$  são correlacionados. Nele os valores defasados de  $\Delta Z_t$  são adicionados. As hipóteses e a estatística de teste permanecem a mesma no teste de dickey-fuller. Sua especificação está apresentada na equação (2.28).

$$\Delta Z_t = \beta_0 + \beta_1 t + \beta_2 Z_{t-1} + \sum_{i=1}^n \alpha_i \Delta Z_{t-i} + \varepsilon_t \quad (2.28)$$

## 2.3 Modelos de Holt-Winters

Segundo Hyndman *et al.*(2002), o método de Holt Winters, também conhecido como suavização exponencial tripla, é utilizado para séries que apresentam nível, tendência e sazonalidade. Para cada um dos três componentes são definidos hiperparâmetros que

decaem exponencial com passar do tempo. Este método pode ser definido como sendo uma média móvel ponderada exponencialmente, onde para observações mais recentes são dados pesos maiores próximos de 1.

### 2.3.1 Modelo Sazonal Aditivo

A equação do modelo sazonal aditivo de Holt-Winters é dada por:

$$Z_t = L_t + T_t t + S_t + \varepsilon_t \quad (2.29)$$

onde,  $L_t$  é o nível,  $T_t$  a tendência,  $S_t$  o fator sazonal no instante  $t$  e  $\varepsilon_t$  é o ruído branco.

A estimação dos parâmetros nível, tendência e fator sazonal podem ser obtidos pelas equações (2.30), (2.31) e (2.32):

$$\hat{L}_t = \alpha(Z_t - S_{t-12}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.30)$$

$$\hat{T}_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.31)$$

$$\hat{S}_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-12} \quad (2.32)$$

E a equação de previsão por:

$$\hat{Z}_{t+m} = (L_t + T_t + m) + S_{t-12+m} \quad (2.33)$$

onde,  $\alpha$ ,  $\beta$  e  $\gamma$  são parâmetros de amortecimento do nível da tendência e da sazonalidade, respectivamente.  $\hat{Z}_{t+m}$  é a previsão  $m$  passos a frente.

### 2.3.2 Modelo Sazonal Multiplicativo

A equação do modelo sazonal multiplicativo de Holt-Winters é dada por:

$$Z_t = (L_t + T_t t)S_t + \varepsilon_t \quad (2.34)$$

onde,  $L_t$  é o nível,  $T_t$  a tendência,  $S_t$  o fator sazonal no instante  $t$  e  $\varepsilon_t$  é o ruído branco .

A estimação dos parâmetros nível, tendência e fator sazonal podem ser obtidos pelas equações (2.35), (2.36) e (2.37):

$$\hat{L}_t = \alpha \left( \frac{Z_t}{S_{t-12}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.35)$$

$$\hat{T}_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.36)$$

$$\hat{S}_t = \gamma \left( \frac{Z_t}{L_t} \right) + (1 - \gamma)S_{t-12} \quad (2.37)$$

E a equação de previsão por:

$$\hat{Z}_{t+m} = (L_t + T_t + m)S_{t-12+m} \quad (2.38)$$

onde,  $\alpha$ ,  $\beta$  e  $\gamma$  são parâmetros de amortecimento do nível, da tendência e da sazonalidade, respectivamente.  $\hat{Z}_{t+m}$  é a previsão  $m$  passos a frente.

## 2.4 Modelos de Box-Jenkins

O modelo da família Box-Jenkins conhecido por  $ARIMA(p, d, q)$  modelo autorregressivo integrado de médias móveis desenvolvido por Box e Jenkins (1970), ganhou notoriedade no meio acadêmico, bem como, na indústria por ter tido resultados satisfatórios de acurácia na previsão de uma série temporal (HYNDMAN E ATHANASOPOULOS, 2013). Este modelo matemático capta o próprio comportamento da série através da autocorrelação entre as próprias observações e estima previsões futuras considerando que a série seja estacionária. Este modelo pode ser dividido em três componentes a autorregressiva, a ordem de integração e as médias móveis.

### 2.4.1 Modelo $AR(p)$ :

A componente autorregressiva,  $AR(p)$ , que utiliza a própria série temporal defasada para estimar a próxima observação. A equação do modelo autorregressivo é apresentada em (2.39):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + \varepsilon_t, \quad (2.39)$$

onde  $p$  é a ordem autorregressiva,  $\phi_i$  são parâmetros do modelo para  $i = 1, 2, 3, \dots, p$  e  $\varepsilon_t$  é o ruído branco.



### 2.4.2 Modelo $MA(q)$ :

É o último componente do modelo  $ARIMA(p, d, q)$  são as médias móveis  $MA(q)$ . Enquanto o  $AR(p)$  são combinações lineares das observações defasadas, por sua vez, as médias móveis utilizam os erros de previsões denominados por  $\theta_i$  para fazer previsão estatística. A equação do modelo de média móvel é apresentado em (2.40):

$$Z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}, \quad (2.40)$$

onde  $q$  é a ordem de média móvel,  $\theta_i$  são parâmetros do modelo para  $i = 1, 2, 3, \dots, q$  e  $\varepsilon_t$  é o ruído branco.

### 2.4.3 Modelo $ARMA(p, q)$ :

Dessa forma, a equação de previsão do modelo geral Box-Jenkins, também conhecido por  $ARMA(p, q)$ , é definida em (2.41):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.41)$$

onde  $p$  é a ordem autorregressiva,  $\phi_i$  são parâmetros do modelo para  $i = 1, 2, \dots, p$ ,  $q$  é a ordem de média móvel,  $\theta_i$  são parâmetros do modelo para  $i = 1, 2, \dots, q$  e  $\varepsilon_t$  é o ruído branco.

### 2.4.4 Definição dos parâmetros $p$ e $q$ :

A definição dos parâmetros a serem estimados nos modelos é feita com base nas análises dos correlogramas, como segue:

Para identificação do parâmetro  $q$ , o comportamento FAC é verificado e para identificar o parâmetro  $p$  o comportamento FACP é verificado. Quando a FAC decai exponencialmente e há um corte brusco na FACP, então trata-se de um modelo  $AR(p)$ , se há um corte brusco na defasagem  $q$  na FAC e um decaimento exponencial na FACP, então trata-se de um modelo  $MA(q)$ . Se a fAC decai exponencialmente após a defasagem  $q$  e a fACP decai exponencialmente após a defasagem  $p$ , então trata-se de um modelo  $ARMA(p, q)$ . A Tabela 1 apresenta o resumo do comportamento das funções FAC e FACP dos modelos  $AR$ ,  $MA$  e  $ARMA$ .

Tabela 1: Comportamento das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) nos processos AR, MA e ARMA.

	FAC	FACP
AR	Decai exponencialmente	Corte brusco após defasagem p
MA	Corte brusco após a defasagem q	Decai exponencialmente
ARMA	Decai exponencialmente após a defasagem q	Decai exponencialmente após a defasagem p

Fonte: Elaboração própria

A Figura 4 contém um exemplo de correlogramas de FAC e FACP de uma série que se adequa a um modelo AR(1), pois sua FAC decai exponencialmente e sua FACP apresenta um corte brusco na defasagem 1.

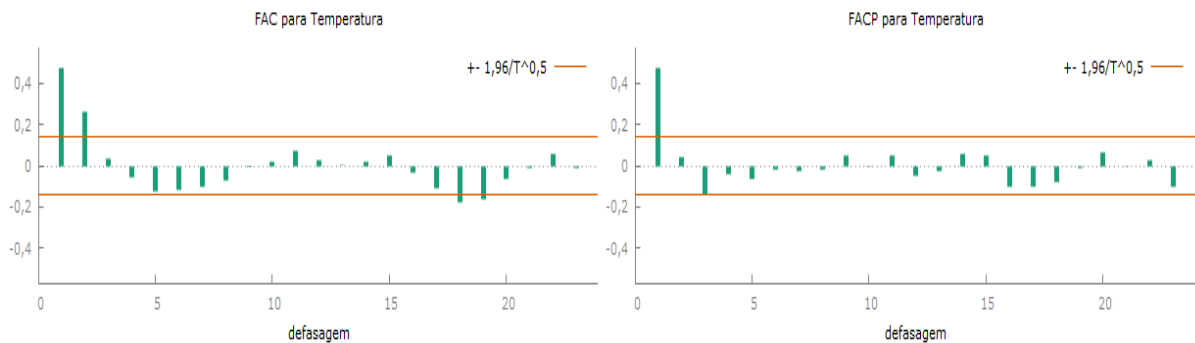


Figura 4: FAC e FACP de um modelo AR(1).

Fonte: Elaboração própria

A Figura 5 contém um exemplo de correlogramas de FAC e FACP de uma série que se adequa a um modelo MA(1), pois apresenta um corte brusco na FAC após a defasagem 1 e decai exponencialmente na FACP.

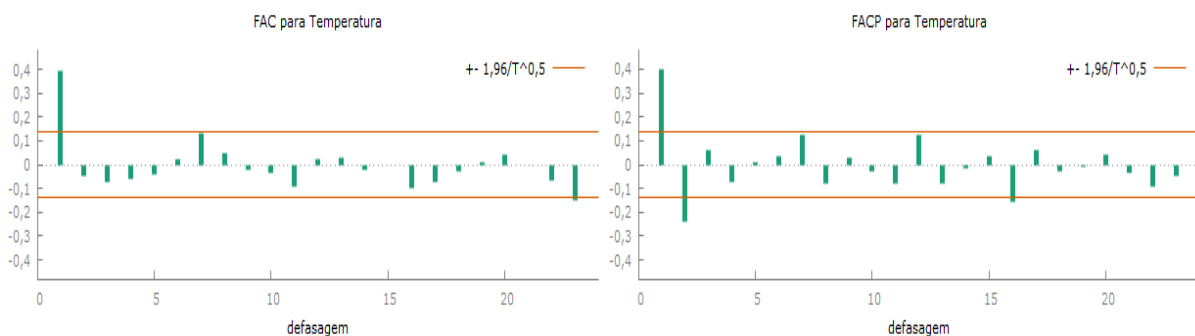


Figura 5: FAC e FACP de um modo MA(1).

Fonte: Elaboração própria

A Figura 6 contém um exemplo de correlogramas de FAC e FACP de uma série

que se adequa a um modelo ARMA(1,1), pois tanto a FAC quanto a FACP decaem exponencialmente após a defasagem 1.

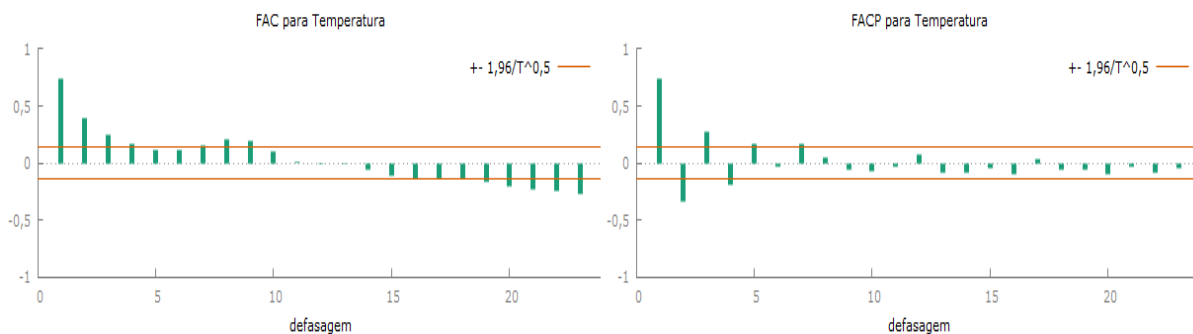


Figura 6: FAC e FACP de um modelo ARMA(1,1).

Fonte: Elaboração própria

### 2.4.5 Modelo $ARIMA(p, d, q)$ :

A outra componente denominada por ordem de integração assume o parâmetro  $d$  que representa o número de diferenças das observações da série até que a mesma seja considerada estacionária. Este parâmetro costuma variar de zero até duas diferenças.

Para uma série não estacionária que se torna estacionária após diferenciação temos o modelo  $ARIMA(p, d, q)$ , modelo autorregressivo integrado de médias móveis (2.42):

$$\Delta^d Z_t = \phi_1 \Delta^d Z_{t-1} + \phi_2 \Delta^d Z_{t-2} + \dots + \phi_p \Delta^d Z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \quad (2.42)$$

onde  $d$  é a ordem de integração,  $\Delta$  o operador de diferenças,  $p$  é a ordem autorregressiva,  $\phi_i$  são parâmetros do modelo para  $i = 1, 2, \dots, p$ ,  $q$  é a ordem de média móvel,  $\theta_i$  são parâmetros do modelo para  $i = 1, 2, \dots, q$  e  $\varepsilon_t$  é o ruído branco.

### 2.4.6 Modelo $SARIMA(p, d, q) \times (P, D, Q)_S$

O modelo ARIMA sazonal  $SARIMA(p, d, q) \times (P, D, Q)_S$  é dado por (2.43):

$$\phi(B)\Phi(B^S)(1 - B^S)^D(1 - B)^d Z_t = \theta(B)\Theta(B^S)\varepsilon_t, \quad (2.43)$$

onde

$$\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \phi_3(B^3) - \dots - \phi_p(B^p), \quad (2.44)$$

é o polinômio autoregressivo de ordem  $p$ ,

$$\Phi(B^S) = 1 - \Phi_1(B^S) - \Phi_2(B^{2S}) - \Phi_3(B^{3S}) - \dots - \Phi_P(B^{PS}), \quad (2.45)$$

é o polinômio autoregressivo sazonal de ordem  $PS$ ,

$$(1 - B^S)^D = \Delta_S^D, \quad (2.46)$$

é a diferença sazonal e  $D$  é o número de diferenças sazonais,

$$(1 - B)^d = \Delta^d, \quad (2.47)$$

é a diferença simples e  $d$  é o número de diferenças simples,

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - \theta_3(B^3) - \dots - \theta_q(B^q), \quad (2.48)$$

é o polinômio de média móvel de ordem  $q$ ,

$$\Theta(B^S) = 1 - \Theta_1(B^S) - \Theta_2(B^{2S}) - \dots - \Theta_Q(B^{QS}), \quad (2.49)$$

é o polinômio de média móvel sazonal de ordem  $QS$  e  $\varepsilon_t$  é o ruído branco.

### 2.4.7 Estimação dos parâmetros dos modelos de Box-Jenkins

Após a identificação de um modelo para a série temporal, deve-se estimar seus parâmetros. Considerando um modelo  $ARIMA(p, d, q)$  e colocando seus  $p + q + 1$  parâmetros no vetor  $\xi = (\phi, \theta, \sigma_\varepsilon^2)$ , onde  $\phi = (\phi_1, \phi_2, \dots, \phi_p)$  e  $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ .

Para determinar o estimador de máximo verossimilhança (EMV) trabalha-se com a suposição que o processo  $\varepsilon_t$  é normal, ou seja, para cada  $t$ ,  $\varepsilon_t \sim N(0, \sigma^2)$ . Nestas condições os EMV serão aproximadamente estimadores de mínimos quadrados (EMQ).

Sob a suposição de normalidade dos  $\varepsilon_t \sim N(0, \sigma^2)$ , a função densidade conjunta de  $\varepsilon_1, \dots, \varepsilon_n$  é dada por:

$$f(\varepsilon_1, \dots, \varepsilon_n) = (2\pi)^{-n/2} (\sigma_\varepsilon)^{-n} \exp \left\{ - \sum_{t=1}^n \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2} \right\}. \quad (2.50)$$

Para estimação dos parâmetros  $\phi$  e  $\theta$  do modelo  $ARIMA(p, d, q)$  é utilizado o método da máxima verossimilhança, cuja função de verossimilhança é dada por:

$$L(\phi, \theta | \sigma_\varepsilon) = (\sigma_\varepsilon^2 2\pi)^{-\frac{1}{2}} \exp \left( - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \varepsilon_t^2 \right). \quad (2.51)$$

Os estimadores de  $\xi = (\phi, \theta, \sigma_\varepsilon^2)$  são os valores que maximizam  $L(\phi, \theta | \sigma_\varepsilon)$ .

## 2.5 Diagnóstico do Modelo

Com os modelos já ajustados, é importante testar se eles são adequados, ou seja, se eles representam os dados. Para tal, serão avaliadas as medidas de aderência e analisados os resíduos.

### 2.5.1 Medidas de aderência

As medidas de aderência são utilizadas para comparar a previsão estatística gerada pelo modelo com valores reais da amostra teste que não foram utilizados na estimação dos parâmetros. Para o atual trabalho serão apresentadas as métricas: Erro Percentual Absoluto Médio (MAPE), o desvio médio absoluto (MAD), o coeficiente de determinação do modelo ( $R^2$ ), o critério de informação bayesiano (BIC), a raiz quadrada do erro quadrático médio (RMSE) e a média geométrica do erro relativo absoluto (GMRAE) que dizem respeito a qualidade do modelo estatístico para previsão.

Todas as medidas de aderências são calculadas dentro da amostra treino de tamanho  $n$ . Para as medidas na amostra teste, será considerado  $t - n = 12$ .

#### 2.5.1.1 MAPE (Erro Percentual Absoluto Médio)

A estatística MAPE mede em percentual do quanto o modelo estatístico está desviando do real valor, conforme é apresentado em (2.53):

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Z_t - \hat{Z}_t}{Z_t} \right|}{n} \times 100, \quad (2.52)$$

onde  $n$  é o tamanho da amostra,  $Z_t$  representa o valor real no instante  $t$ ,  $\hat{Z}_t$  representa o valor previsto no instante  $t$ .

#### 2.5.1.2 MAD (Desvio Médio Absoluto)

A estatística MAD está definida abaixo na equação (2.54):

$$MAD = \sum_{t=1}^n \frac{|Z_t - \hat{Z}_t|}{n}, \quad (2.53)$$

onde  $Z_t$  é o valor observado no instante  $t$  e  $\hat{Z}_t$  é o valor ajustado no mesmo instante.

### 2.5.1.3 Coeficiente de determinação ( $R^2$ )

O coeficiente de determinação ( $R^2$ ) é uma medida de ajustamento do modelo estatístico, isto é, o quanto as variáveis preditoras explicam o evento estimado (variável resposta). Essa medida assume valores entre 0 e 1 e é representada na equação (2.55):

$$R^2 = 1 - \left( \frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{\sum_{t=1}^n (Z_t - \bar{Z})^2} \right), \quad (2.54)$$

onde  $\bar{Z}$  representa a média das observações.

### 2.5.1.4 Critério de informação bayesiano (BIC)

O critério de informação bayesiano (BIC) utiliza o princípio da parsimônia penalizando modelos com muitas variáveis. Valores menores de BIC são preferíveis. O cálculo é feito através da fórmula (2.56)

$$BIC = -2 \log L_p + [(p + 1) + 1] \log n, \quad (2.55)$$

onde  $L_p$  é o máximo da função de verossimilhança,  $p$  número de parâmetros e  $n$  tamanho da amostra

### 2.5.1.5 RMSE (Raiz Quadrada do Erro Quadrático Médio)

O RMSE é uma forma de analisar o erro na escala do evento estimado (diferentemente do MAPE). Primeiro, eleva-se ao quadrado a diferença entre o previsto e o real para não cancelar erros positivos com negativos, após calcula-se a média, e por fim retira-se a raiz quadrada a fim de retornar a escala original conforme é visto em (2.57):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Z_t - \hat{Z}_t)^2}{n}} \quad (2.56)$$

### 2.5.1.6 GMRAE (Média Geométrica do Erro Relativo Absoluto)

O GMRAE compara o erro do modelo selecionado com o da previsão. Deseja-se que o GMRAE seja menor que 1. O cálculo é feito através da fórmula (2.58):

$$GMRAE = \left( \prod_{t=1}^m \left| \frac{Z_t - \hat{Z}_t}{Z_t - \hat{Z}_t^*} \right| \right)^{\frac{1}{m}}, \quad (2.57)$$

onde  $m$  é o período sazonal e  $\hat{Z}_t^* = \begin{cases} Z_t - 1 & \text{não sazonal,} \\ Z_{t-m} & \text{sazonal} \end{cases}$

## 2.5.2 Análise dos Resíduos

A análise dos resíduos pode ser feita usando a FAC ou a FACP para verificar a independência dos dados. Na ausência de correlação desses erros podemos supor que os dados são independentes e identicamente distribuídos (iid) com media zero e variância constante, tendo assim um bom modelo.

Para essa verificação será avaliada a FAC dos resíduos, que devera estar em torno de zero, com nenhuma ou poucas autocorrelações fora do intervalo de confiança. A Figura 7 apresenta um exemplo de correlograma da FAC dos resíduos.

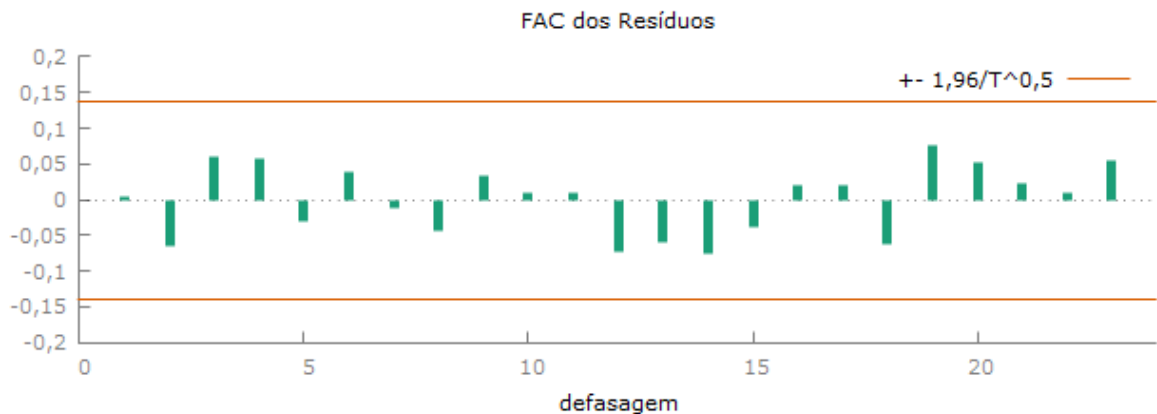


Figura 7: Exemplo de correlograma da FAC dos resíduos.

Fonte: Elaboração própria

## 2.6 Fluxograma da Proposta

A proposta deste trabalho consiste em gerar modelos de séries temporais para escolha do melhor com base nas estatísticas de aderência e da análise dos resíduos para previsão de velocidade do vento.

Na ocasião, são gerados modelos de Holt-Winters com e sem tendência linear e sazonalidade e modelos de Box-Jenkins a partir das análises de FAC e FACP, além dos testes de Normalidade e estacionariedade e da análise da capacidade preditiva dos modelos.

Ao final, o melhor modelo de Holt-Winters e o melhor modelo de Box-Jenkins são testados e o modelo campeão é o escolhido.

A previsão é feita com a subdivisão dos dados da série em amostra de treino e amostra de teste. Na amostra de treino, o modelo é treinado com base nos dados amostrais e na amostra de teste, o modelo ajustado é testado quanto a sua capacidade preditiva usando parte da própria série como amostra para calcular as estatísticas de aderência. Por fim, a previsão é feita com o horizonte de 1 ano.

A Figura 8 apresenta um fluxograma sugestivo do caminho trilhado neste TCC.



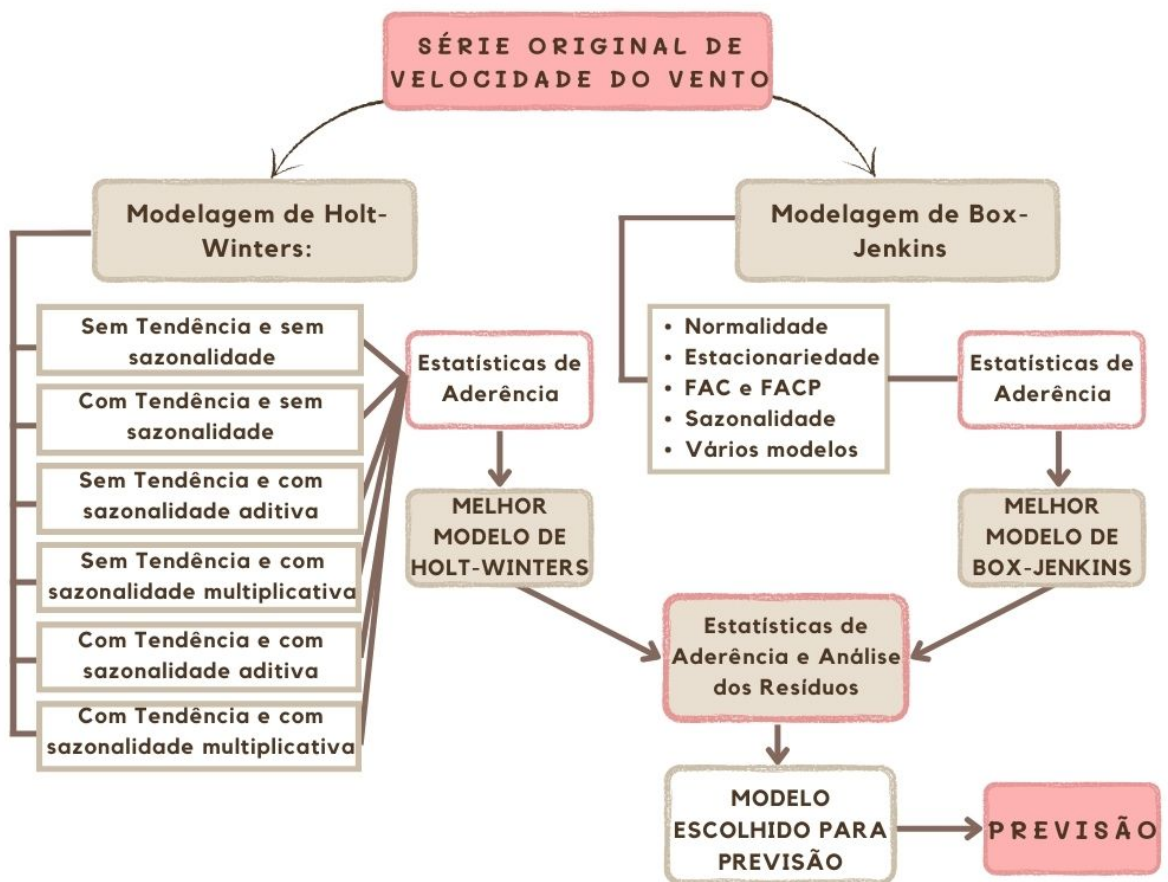


Figura 8: Fluxograma da proposta.

## 3 Análise dos Resultados

Para as análises realizadas a seguir serão utilizados dados de janeiro de 2006 a outubro de 2020. Separando um período de 12 meses, ao final, para análise na amostra teste. Sendo assim, para a análise na amostra treino será considerado o período de 166 meses referente a janeiro de 2006 a outubro de 2019 e para a análise na amostra teste será considerado o período de novembro de 2019 a outubro de 2020.

As modelagens de Holt-Winters foram feitas utilizando o programa FPW bem como todas as estatísticas de aderência e a análise dos resíduos. Nas modelagens de Box-Jenkins, os testes de normalidade e estacionariedade juntamente com as análises dos correlogramas foram feitos a partir do programa gretl e para suas modelagens, estatísticas de aderência e análise dos resíduos foi utilizado o FPW.

### 3.1 Análise descritiva

A velocidade do vento que otimiza a geração de energia está dentro do intervalo de 3m/s a 15m/s. A Velocidade do vento inferior a 3m/s é insuficiente para gerar energia e acima de 15m/s, a turbina é travada automaticamente para prevenir que aconteçam danos por excesso de vento (PICOLO et. al, 2014). A Tabela 2 apresenta algumas medidas descritivas da série de velocidades do vento da estação anemométrica de Campo Grande/MS onde a velocidade do vento mínima medida é 3,426 e a máxima é 6,879, estando incluso, portanto, no intervalo que otimiza a geração de energia.

Tabela 2: Algumas medidas descritivas da Série de Velocidades do Vento da Estação Anemométrica de Campo Grande/MS - jan/2006 a out/2020.

Estatística	Data	Vel. vento mensal (m/s)
Mínimo	01/2006	3,426
Quartil 1	09/2009	4,507
Mediana	05/2013	4,942
Média	05/2013	5,037
Quartil 3	01/2017	5,453
Máximo	10/2020	6,879

Fonte: Elaboração própria

A Figura 9 apresenta o comportamento das médias mensais de velocidade do vento analisadas mês a mês. É possível verificar que nos meses de agosto e setembro há uma velocidade do vento maior em relação aos outros meses, isso acontece de forma sazonal já que esse gráfico representa a média mensal verificada por mês de um conjunto de dados gerais. Essa situação juntamente com a permanência das velocidades do vento dentro do intervalo ideal de geração de energia eólica, indica a possibilidade de implementação de um parque eólico na região.

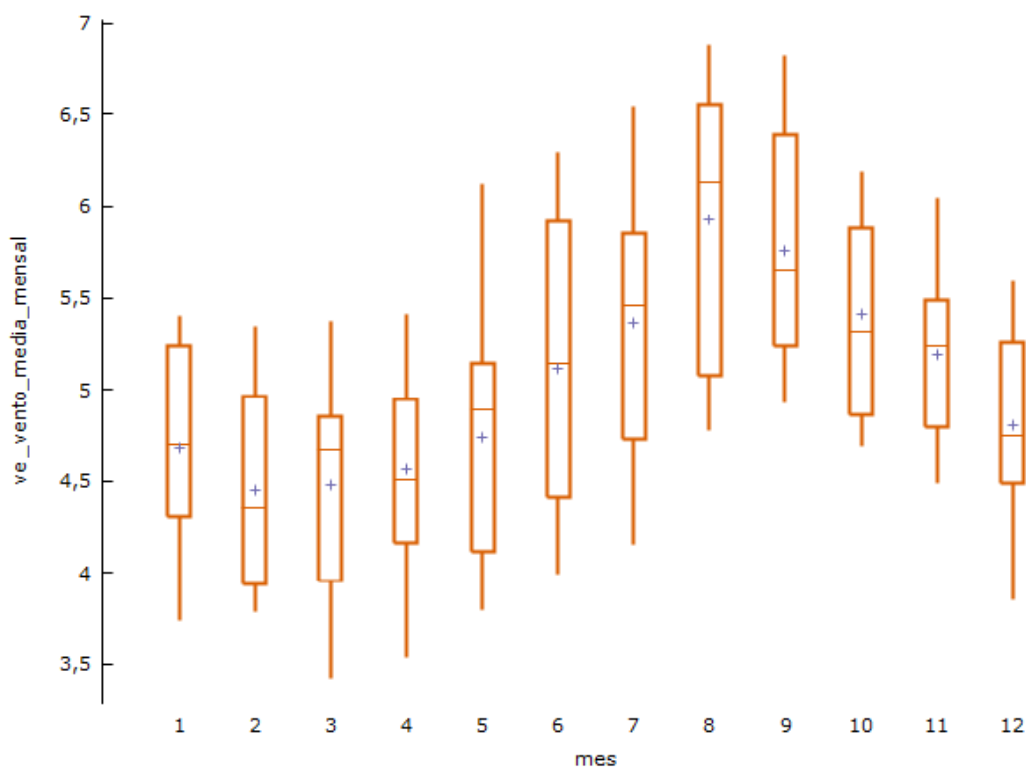


Figura 9: Box-plot da média mensal de velocidade do vento por mês.

Fonte: Elaboração própria

Pode ser visto na Figura 9 que há um crescimento das velocidades do vento de março a agosto e que de setembro a fevereiro o comportamento é de decréscimo o que corrobora com a existência de sazonalidade na série.

## 3.2 Modelagem de Holt-Winters

Na modelagem de Holt-Winters são ajustados modelos sem tendência e sem sazonalidade, com tendência e sem sazonalidade, sem tendência e com sazonalidade aditiva, sem tendência e com sazonalidade multiplicativa, com tendência e com sazonalidade aditiva, com tendência e com sazonalidade multiplicativa. A Tabela 3 apresenta os ajustes destes modelos com as estimativas do nível, da tendência e do fator sazonal.

Tabela 3: Estimação dos parâmetros dos modelos de Holt-Winters ajustados.

Modelos	Nível	Tendência	Sazonalidade
Sem tendência e sem sazonalidade	5,4763	-	-
Com tendência e sem sazonalidade	5,4755	0,000002006	-
Sem tendência e com sazonalidade aditiva	5,2102	-	0,14757
Sem tendência e com sazonalidade multiplicativa	5,1818	-	0,14793
Com tendência e com sazonalidade aditiva	5,2165	0,001400000	0,14788
Com tendência e com sazonalidade multiplicativa	5,1893	0,001360000	0,14465

Fonte: Elaboração própria

A Tabela 4 apresenta as medidas de aderência para os modelos de Holt-Winters com a série original de velocidade do vento, onde pode ser visto um empate entre os modelos: Sem tendência e com sazonalidade aditiva e Sem tendência e com sazonalidade multiplicativa. Estes dois modelos estão empatados com menores estatísticas de aderência e maior  $R^2$  ajustado.

Tabela 4: Medidas de aderência para os modelos de Holt-Winters com a série original (Análise na amostra treino).

Modelos	MAPE	RMSE	BIC	R <sup>2</sup>
Sem tendência e sem sazonalidade	0,085	0,543	0,552	0,498
Com tendência e sem sazonalidade	0,085	0,543	0,560	0,498
<b>Sem tendência e com sazonalidade aditiva</b>	<b>0,061</b>	<b>0,401</b>	<b>0,414</b>	<b>0,726</b>
<b>Sem tendência e com sazonalidade multiplicativa</b>	<b>0,061</b>	<b>0,401</b>	<b>0,414</b>	<b>0,726</b>
Com tendência e com sazonalidade aditiva	0,061	0,401	0,420	0,726
Com tendência e com sazonalidade multiplicativa	0,061	0,401	0,420	0,726

Fonte: Elaboração própria

A solução para o desempate está na Tabela 5, pois apenas o modelo Sem tendência e com sazonalidade aditiva está minimizando as estatísticas de aderência, na amostra teste, que por limitação do programa apresenta apenas as estatísticas MAPE, MAD, GMRAE, a amostra teste é a amostra onde o modelo é testado para previsão. Portanto, o nosso modelo de Holt-Winters campeão é o modelo Sem tendência e com sazonalidade aditiva.

Tabela 5: Medidas de aderência para os modelos de Holt-Winters com a série original (Análise na amostra teste).

Modelos	MAPE	MAD	GMRAE
Sem tendência e sem sazonalidade	0,094	0,504	0,854
Com tendência e sem sazonalidade	0,094	0,504	0,854
<b>Sem tendência e com sazonalidade aditiva</b>	<b>0,058</b>	<b>0,309</b>	<b>0,508</b>
Sem tendência e com sazonalidade multiplicativa	0,059	0,315	0,519
Com tendência e com sazonalidade aditiva	0,059	0,312	0,550
Com tendência e com sazonalidade multiplicativa	0,060	0,319	0,561

Fonte: Elaboração própria

### 3.3 Modelagem de Box-Jenkins

Para a modelagem de Box-Jenkins é necessário considerar algumas condições iniciais, como normalidade e estacionariedade. Pode ser visto na Figura 1 que a série temporal de média mensal da velocidade do vento do município de Campo Grande/MS aparenta ter um comportamento estacionário e com sazonalidade. Para verificar estas suposições, testes de normalidade e de raiz unitária serão feitos antes das análises dos correlogramas

para definição dos parâmetros a serem estimados nos ajustes dos modelos.

### 3.3.1 Teste de Normalidade:

Seguindo as hipóteses nula e alternativa:

$H_0$ : Os dados são provenientes de uma distribuição normal; e

$H_1$ : Os dados não são provenientes de uma distribuição normal.

Com o  $p$  – valor = 0,0341, resultante do teste de normalidade de Shapiro-Wilk e considerando o nível de significância de 5% rejeitamos a hipótese nula de os dados seguirem distribuição normal. Ou seja, os dados não são normais e será necessário realizar uma transformação logarítmica nesses dados.

A Figura 10 apresenta os dados da série de velocidade do vento em média mensal após a transformação logarítmica.

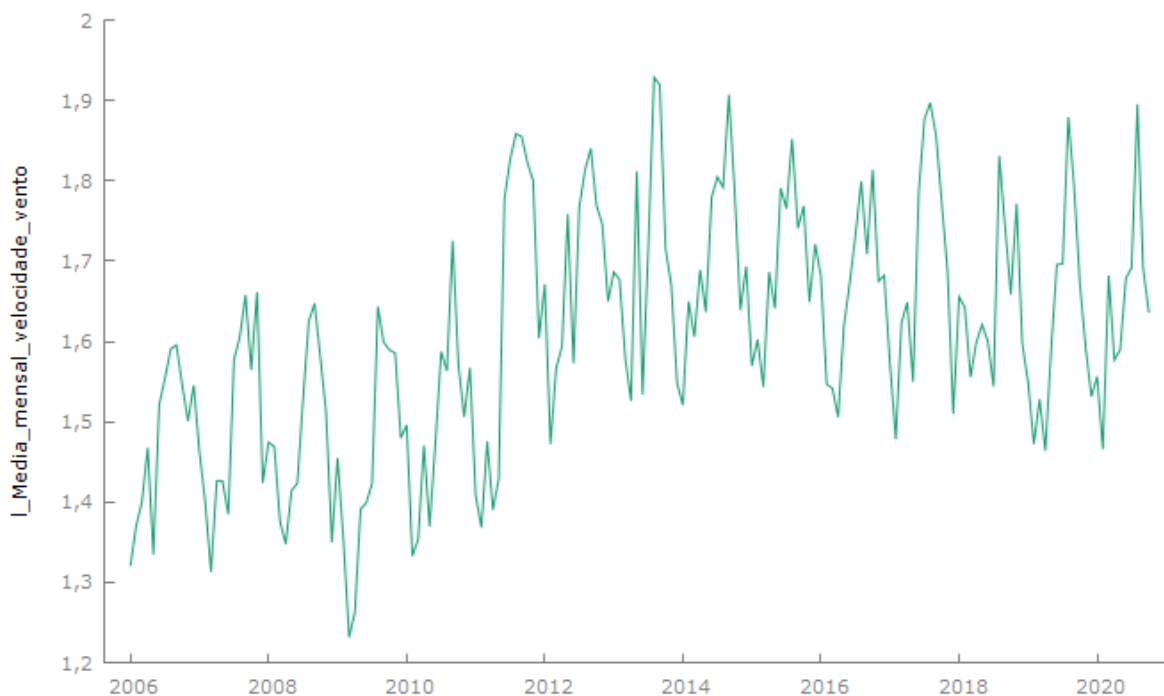


Figura 10: Gráfico da série com transformação logarítmica.

Fonte: Elaboração própria

### 3.3.2 Teste de Raiz Unitária:

Seguindo as hipóteses nula e alternativa:

$H_0$ :  $\beta_1 = 0 \approx \rho = 1$ , a série  $Z_t$  possui raiz unitária ( $Z_t$  não é estacionária, mas  $\Delta Z_t$  é

estacionária); e

$H_1: \beta_1 < 0 \approx \rho < 1$ , a série  $Z_t$  não possui raiz unitária.

Com  $p\text{-valor} = 0,0160$  como resultado do teste de Dickey-Fuller Aumentado e considerando o nível de significância de 5%, rejeitamos a hipótese nula de que a série não é estacionária e que possui raiz unitária; ou seja, a série média mensal de velocidade do vento é estacionária.

### 3.3.3 Análise dos Correlogramas:

A Figura 11 apresentada a seguir, contém os correlogramas da FAC e da FACP do log da série de velocidade do vento em média mensal.

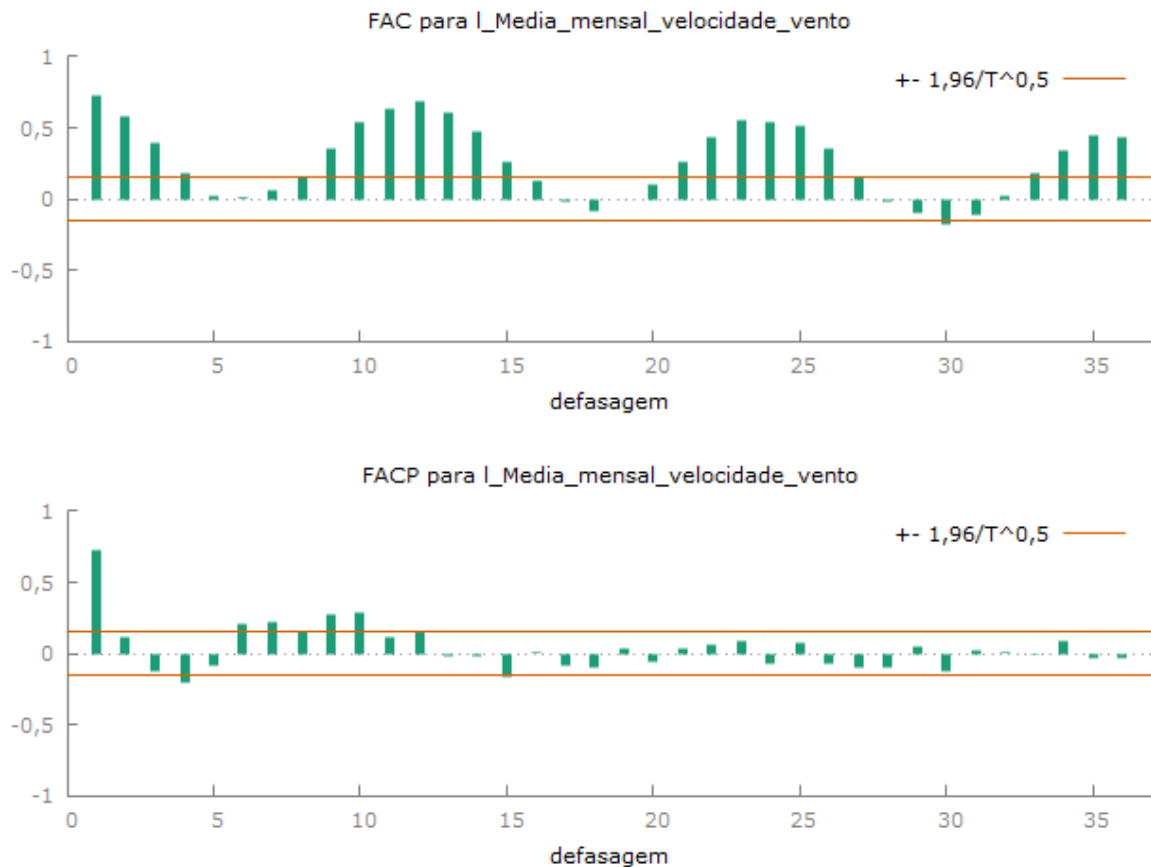


Figura 11: Correlograma do log série de velocidade do vento em média mensal.

Fonte: Elaboração própria

É possível observar que a FAC apresenta um decaimento rápido nos lags iniciais e a FACP apresenta um corte brusco no lag 1. Este comportamento é típico de um modelo AR(1). Ao observar a parte sazonal (de 12 em 12), pode-se perceber que nos lags 12, 24 e 36 há um decaimento lento indicando a necessidade de se fazer uma diferença sazonal.

A Figura 12 apresenta o correlograma da série de velocidade do vento em média mensal após a diferença sazonal.

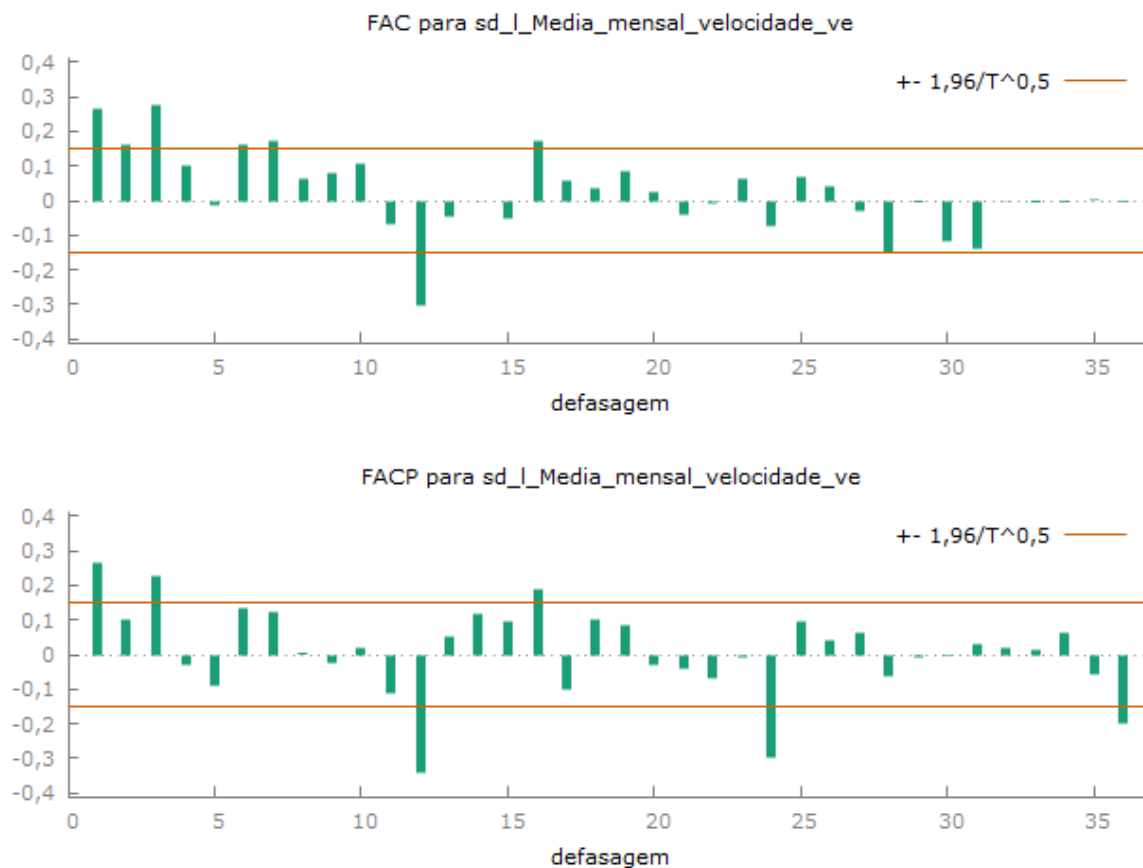


Figura 12: Correlograma da série após diferença sazonal.

Fonte: Elaboração própria

Após a diferença sazonal, pode-se perceber no correlograma da Figura 7 que a FAC apresenta um corte brusco no lag 12 e a FACP apresenta um decaimento nos lags 12, 24 e 36, podendo indicar um comportamento típico de um modelo MA(1) sazonal.

Porém, tal comportamento também pode ser entendido como um corte brusco no lag 12 na FAC e 3 cortes bruscos nos lags 12, 24 e 36 na FACP, o que configura um modelo ARMA(3,1) sazonal.

Com isso, associando a parte simples, podemos ter os modelos:

$$SARIMA(1, 0, 0) \times (0, 1, 1)_{12}$$

$$SARIMA(1, 0, 0) \times (3, 1, 1)_{12}$$

Em seguida, são sugeridos na Tabela 6 modelos aproximados a estes dois, onde são apresentados os seus parâmetros estimados, coeficientes e significâncias, pois mesmo após as análises e os testes é possível haver modelos mais adequados que estes. O nível de



significância utilizado na estimação é de 95%.

Tabela 6: Parâmetros estimados nos modelos de Box-Jenkins, seus coeficientes e suas significâncias (Amostra treino).

Modelos	Parâmetros estimados	Coefficientes	Significância
SARIMA(1,0,1) × (2, 1, 1) <sub>12</sub>	a[1]	1,0000	1,0000
	b[1]	0,7552	1,0000
	A[12]	-0,0617	0,5714
	A[24]	-0,2237	0,9939
	B[12]	0,8846	1,0000
	SARIMA(1,0,0) × (0, 1, 1) <sub>12</sub>	a[1]	0,6425
B[12]		0,8840	1,0000
SARIMA(1,0,0) × (1, 1, 1) <sub>12</sub>	a[1]	0,5787	1,0000
	A[12]	0,1464	0,9063
	B[12]	0,8802	1,0000
SARIMA(1,0,0) × (2, 1, 1) <sub>12</sub>	a[1]	0,6907	1,0000
	A[12]	0,0654	0,5788
	A[24]	-0,2342	0,9959
	B[12]	0,8621	1,0000
SARIMA(1,0,0) × (3, 1, 1) <sub>12</sub>	a[1]	0,7207	1,0000
	A[12]	0,0351	0,3052
	A[24]	-0,2407	0,9972
	B[12]	-0,0914	0,6922
	B[24]	0,8670	1,0000
	SARIMA(2,0,0) × (3, 1, 1) <sub>12</sub>	a[1]	0,4344
a[2]		0,4820	1,0000
A[12]		-0,1862	0,9785
A[24]		-0,3457	1,0000
A[36]		-0,2396	0,9967
B[12]		0,9063	1,0000
SARIMA(2,0,1) × (3, 1, 1) <sub>12</sub>		a[1]	0,9619
	a[2]	0,0351	0,2475
	B[1]	0,7181	1,0000
	A[12]	-0,1344	0,8985
	A[24]	-0,2710	0,9995
	A[36]	-0,1722	0,9634
	B[12]	0,8963	1,0000
	SARIMA(3,0,1) × (3, 1, 1) <sub>12</sub>	a[1]	1,0321
a[2]		0,0790	0,4919
a[3]		-0,1142	0,6684
b[1]		0,7839	1,0000
A[12]		-0,1550	0,9380
A[24]		-0,2673	0,9993
A[36]		-0,1924	0,9788
B[12]		0,8863	1,0000

Fonte: Elaboração própria

Destes modelos apresentados na Tabela 6, apenas os modelos  $SARIMA(1, 0, 0) \times (0, 1, 1)_{12}$  e  $SARIMA(2, 0, 0) \times (3, 1, 1)_{12}$  não possuem parâmetros com significância menor que 95%, os demais modelos foram eliminados por seus parâmetros terem significância menor que 95%. Tendo isso em foco, e avaliando a Tabela 7 que contém as medidas de aderência para os modelos de Box-Jenkins usando a amostra treino, o modelo  $SARIMA(2, 0, 0) \times (3, 1, 1)_{12}$  aparenta ser o melhor.

Tabela 7: Medidas de aderência para os modelos de Box-Jenkins com a série original com transformação logarítmica (Análise na amostra treino).

Modelos	MAPE	RMSE	BIC	R <sup>2</sup>
$SARIMA(1,0,0) \times (0, 1, 1)_{12}$	0,066	0,449	0,451	0,668
<b><math>SARIMA(2,0,0) \times (3,1,1)_{12}</math></b>	<b>0,058</b>	<b>0,391</b>	<b>0,413</b>	<b>0,754</b>

Fonte: Elaboração própria

Embora na Tabela 7 o modelo  $SARIMA(2, 0, 0) \times (3, 1, 1)_{12}$  tenha aparentado ser o melhor, ao avaliar as medidas de aderência fora da amostra da Tabela 8, obtemos que o melhor modelo é o  $SARIMA(1, 0, 0) \times (0, 1, 1)_{12}$ .

Tabela 8: Medidas de aderência para os modelos de Box-Jenkins com a série original com transformação logarítmica (Análise na amostra teste).

Modelos	MAPE	MAD	GMRAE
<b><math>SARIMA(1,0,0) \times (0,1,1)_{12}</math></b>	<b>0,058</b>	<b>0,312</b>	<b>0,452</b>
$SARIMA(2,0,0) \times (3, 1, 1)_{12}$	0,082	0,437	0,719

Fonte: Elaboração própria

### 3.4 Melhores modelos

Os melhores modelos de Holt-Winter e Box-Jenkins juntamente com suas medidas de aderência, estão apresentados abaixo na Tabela 9. Ao serem comparados, o modelo de Holt-Winters Sem tendência e com sazonalidade aditiva se destaca como o melhor modelo, pois este minimiza as estatísticas de aderência MAPE, RMSE e BIC e possui o maior  $R^2$  ajustado.

Tabela 9: Medidas de aderência para os melhores modelos de Holt-Winters e Box-Jenkins

Modelos	MAPE	RMSE	BIC	R <sup>2</sup>
<b>Sem tendência e com sazonalidade aditiva</b>	<b>0,061</b>	<b>0,401</b>	<b>0,414</b>	<b>0,726</b>
SARIMA(1,0,0)×(0,1,1) <sub>12</sub>	0,066	0,449	0,451	0,668

Fonte: Elaboração própria

Para analisar os resíduos do modelo de Holt-Winters Sem tendência e com sazonalidade aditiva, vamos avaliar a função de autocorrelação (FAC) dos resíduos, para que o modelo seja adequado os resíduos não devem ser correlacionados, é adequado que os valores da FAC não estejam fora do intervalo de confiança. A Figura 13 apresenta a FAC dos resíduos desse modelo.

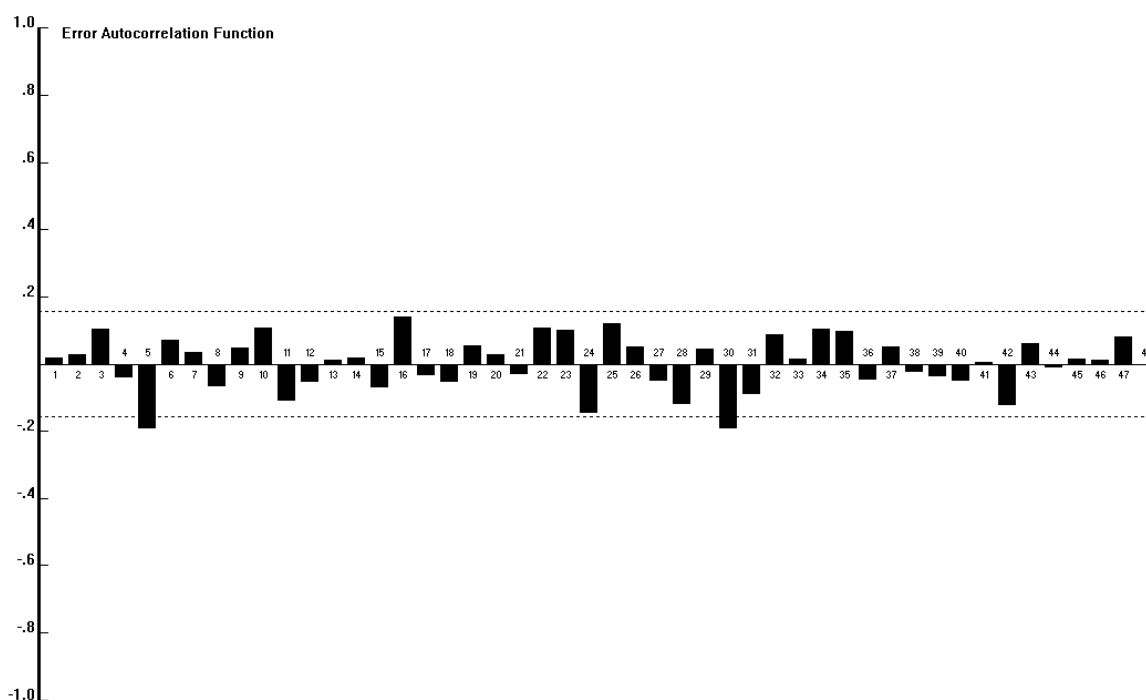


Figura 13: FAC dos resíduos do modelo de Holt-Winters Sem tendência e com sazonalidade aditiva.

Fonte: Elaboração própria

Como visto na Figura 13, a maioria dos valores da função de autocorrelação dos resíduos estão dentro do intervalo de confiança, o que indica que o modelo é adequado.

A seguir, na Figura 14, temos a série e previsão com horizonte de 1 ano do modelo de Holt-Winters Sem tendência e com sazonalidade aditiva. Ao observar o gráfico vemos que o modelo segue o comportamento da série original e ao avaliar a função de autocorrelação dos erros na Figura 8 concluiu-se que eles eram não correlacionados, indicando que o modelo se ajustou bem.

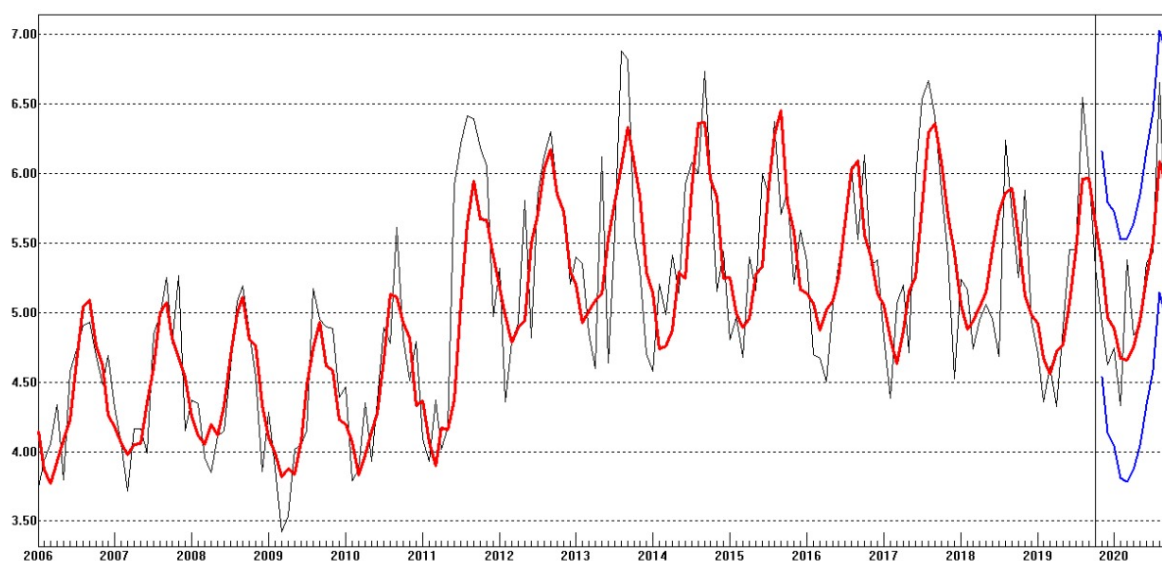


Figura 14: Série original (preta, linha mais fina) x série ajustada (vermelha) e previsão (linha vermelha após final da linha preta) usando o modelo escolhido.

Fonte: Elaboração própria

A Tabela 10 apresenta os valores observados da amostra teste e os valores previstos pelo modelo de Holt-Winters Sem tendência e com sazonalidade aditiva com horizonte de 1 ano após a amostra teste. Ao comparar os valores observados na amostra teste com os valores previstos para este mesmo período, vemos que os valores são próximos indicando um bom ajuste do modelo, o que é corroborado pelas estatísticas de aderência feitas na amostra teste, as quais o modelo ajustado utilizado foi o que minimizou estas medidas, que tem como base a diferença entre os valores ajustados e observados. Tal comparação pode ser vista na Tabela 10. Os valores previstos estão entre 3m/s e 15m/s, indicando a possibilidade de implementação de uma usina eólica em Campo Grande/MS.

Tabela 10: Valores previstos pelo modelo de Holt-Winters Sem tendencia e com sazonalidade aditiva

Data	Valores observados	Valores previstos	IC
11/2019	4,918	5,346	[4,535; 6,157]
12/2019	4,624	4,964	[4,137; 5,790]
01/2020	4,742	4,882	[4,040; 5,724]
02/2020	4,332	4,669	[3,812; 5,526]
03/2020	5,377	4,658	[3,786; 5,529]
04/2020	4,840	4,754	[3,868; 5,639]
05/2020	4,901	4,945	[4,046; 5,845]
06/2020	5,358	5,251	[4,337; 6,165]
07/2020	5,430	5,513	[4,585; 6,440]
08/2020	6,654	6,089	[5,148; 7,030]
09/2020	5,425	5,913	[4,958; 6,867]
10/2020	5,133	5,540	[4,572; 6,507]
11/2020	-	5,346	[4,365; 6,327]
12/2020	-	4,964	[3,970; 5,957]
01/2021	-	4,882	[3,876; 5,888]
02/2021	-	4,669	[3,650; 5,688]
03/2021	-	4,658	[3,627; 5,689]
04/2021	-	4,754	[3,711; 5,797]
05/2021	-	4,945	[3,890; 6,001]
06/2021	-	5,251	[4,184; 6,318]
07/2021	-	5,513	[4,434; 6,592]
08/2021	-	6,089	[4,998; 7,179]
09/2021	-	5,913	[4,811; 7,015]
10/2021	-	5,540	[4,426; 6,653]

Fonte: Elaboração própria

Conforme pode ser observado na Tabela 10 os valores previstos na amostra teste estão aderindo aos valores observados entre novembro de 2019 a outubro de 2020. Isso contribui para que acreditemos que as previsões sejam boas. Além dos valores previstos também são apresentados intervalos de confiança ao nível de 95% com base nos erros de previsão, os valores dentro destes intervalos ainda estão dentro dos limites de 3m/s e 15m/s necessários para geração de energia eólica.

## 4 Conclusões

A proposta desse trabalho era a modelagem de uma série temporal mensal de velocidade do vento para gerar previsões com horizonte de até 1 ano. Seu objetivo geral era avaliar a possibilidade de implementação de um parque eólico em Campo Grande, estado de Mato Grosso do Sul.

Com os dados de velocidade do vento da estação anemométrica localizada na cidade de Campo Grande, no estado de Mato Grosso do Sul, foram estimados modelos de Holt-Winters e Box-Jenkins e comparados a partir de suas medidas de aderência calculadas nas amostras treino e teste, avaliando as capacidade preditivas dos modelos e verificado qual dos modelos estudados era o mais adequado para este conjunto de dados.

Após realizar o diagnóstico do modelo e obter resultado positivo, pode-se concluir que o modelo de Holt-Winters sem tendência e com sazonalidade aditiva é o modelo mais adequado para esta série de velocidade do vento em Campo Grande/MS.

As estimativas de valores futuros encontradas mostram que a implementação de uma usina eólica no local da estação anemométrica de Campo Grande/MS é viável, uma vez que os valores previstos estão dentro dos limites padrões de velocidade do vento para a geração de energia eólica. Ou seja, de 3m/s a 15m/s.

Além da verificação da possibilidade de implementação de uma usina eólica através das estimativas pontuais, os valores estimados no intervalo de confiança ao nível de 95% também se encontram dentro do limite de velocidade do vento necessário para geração de energia, isso fortalece a conclusão de que o estabelecimento do parque eólico é viável.

Em trabalhos futuros, pode ser avaliada a possibilidade de implementação de energia fotovoltaica na mesma região via previsão de incidência solar com modelos Holt-Winters e Box-Jenkins.

# Referências

Dados horários da operação. CCEE. Infomercado - dados horários da operação. São Paulo, 2021. Disponível em: <https://www.ccee.org.br/web/guest/dados-e-analises/dados-geracao>. Acesso em: 28/11/2021.

Nascimento, R. L. Energia Solar no Brasil: situação e perspectivas. *Consultoria Legislativa, Estudo técnico*, Março/2017.

Camelo, H. N.; Lucio, P. S.; Leal Junior, J. V.. Modelagem da velocidade do vento usando metodologias ARIMA, Holt-Winters e RNA na previsão de geração eólica no nordeste brasileiro. *Revista Brasileira de Climatologia*, [S.l.], v. 21, nov. 2017. ISSN 2237-8642. doi:<http://dx.doi.org/10.5380/abclima.v21i0.48565>.

Kavasseri, R. G., Seetharaman, K.. Day-ahead wind speed forecasting using f-ARIMA models, *Renewable Energy*, 34 (5), pp. 1388 - 1393, 2009, ISSN 0960-1481, doi: <https://doi.org/10.1016/j.renene.2008.09.006>.

Yatiana, E., Rajakamura, S., and Ghosh, A.. Wind speed and direction forecasting for wind power generation using ARIMA model, Australasian Universities Power Engineering Conference (AUPEC), 2017, pp. 1-6, doi: 10.1109/AUPEC.2017.8282494.

Pereira, L. T.; Perceus, M.; Pereira, E. T. Modelagem matemática e previsão da velocidade do vento em Palmeira dos Índios/AL com a utilização dos modelos Box Jenkins. *REVEXT - Revista de Extensão da Universidade Estadual de Alagoas - UNEAL*, [S. l.], v. 2, n. 1, p. 17–32, 2017.

Picolo, A. P., Rühler, A. J. e Rampinelli, G. A.. Uma abordagem sobre a energia eólica como alternativa de ensino de tópicos de física clássica. *Revista Brasileira de Ensino de Física*, 36 (4), 2014.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D. e Grose, S. A. *A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods*. International Journal of Forecasting, vol.18, n. 3, p. 439–454, 2002.

Hyndman, R. J. e Athanasopoulos, G.. Forecasting Principles and Practice. Otexts. 2013.

Morettin P. A. e Toloí C. M. C. *Análise de séries temporais Modelos lineares univariados*, 2018.

Dalmaz, A. Passos, J. C. Colle, S.. Energia eólica para geração de eletricidade ea importância da previsão. Revista ABCM–Engenharia, v. 13, n. 1, p. 27-32, 2008.

Box, G.; Jenkins, G.. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. 1970.

Nações Unidas Brasil. Brasília, DF, Brasil, 2021. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 30/01/2022.

Dickey, D.A. Fuller, W.A.. *Distribution of the Estimators for Autoregressive Time Series with a unit root*. Journal of the American Statistical Association, vol,n. 74, p. 427-431, 1979.

Shapiro SS, Wilk MB. *An Analysis of Variance Test for Normality (Complete Samples)*. Biometrika. 52(3/4):591-611, 1965. Available from: <http://www.jstor.org/stable/2333709>.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

The gretl team. gretl [Internet]. 2021. Available from: <http://gretl.sourceforge.net/>

Forecast Pro for Windows. Version 3.00D Extended Edition Business Forecast Systems, Inc. Belmont, Massachusetts, USA, 1997.

Microsoft Corporation, 2018. Microsoft Excel, Available at: <https://office.microsoft.com/excel>.