



*Trabalho de Conclusão de Curso*

# Uso de Séries Temporais para Previsão de Irradiação Solar

de Gregory Albertt Santos Carvalho

orientado por

Prof. Dr. Ícaro Bezerra Queiroz de Araújo

Universidade Federal de Alagoas  
Instituto de Computação  
Maceió, Alagoas  
15 de Setembro de 2020



**Catlogação na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

C331u Carvalho, Gregory Albertt Santos.  
Uso de séries temporais para previsão de irradiação solar / Gregory  
Albertt Santos Carvalho. – 2020.  
69 f. : il., figs. e tabs. color.

Orientador: Ícaro Bezerra Queiroz de Araújo.  
Monografia (Trabalho de Conclusão de Curso em Engenharia da  
Computação) – Universidade Federal de Alagoas. Instituto de Computação.  
Maceió, 2020.

Bibliografia: f. 60-63.  
Apêndices: f. 64-69.

1. Séries temporais. 2. Modelo SARIMA. 3. Irradiância solar. 4. Box-  
Jenkins. I. Título.

CDU: 551.521.31

UNIVERSIDADE FEDERAL DE ALAGOAS  
Instituto de Computação

## USO DE SÉRIES TEMPORAIS PARA PREVISÃO DE IRRADIAÇÃO SOLAR

Trabalho de Conclusão de Curso submetido ao Instituto de Computação da Universidade Federal de Alagoas como requisito parcial para a obtenção do grau de Engenheiro de Computação.

Aprovado em 15 de Setembro de 2020:

---

Ícaro Bezerra Queiroz de Araújo,  
Prof. Dr., Orientador

---

João Raphael Souza Martins,  
Prof. Dr., UFAL

---

Winnie de Lima Torres,  
Ma.

# Dedicatória

À minha família.

# Agradecimentos

À minha família, base da minha vida. Em especial aos meus pais, Alberto e Adriana, ídolos e espelhos, por lutarem pela minha educação e me concederem o privilégio de ter bons estudos e todas as condições possíveis para chegar até aqui.

Ao meus tios e avós, em especial a meu padrinho e avô Antônio Cajueiro, *in memoriam*, que com certeza ficaria muito feliz por mim e aos meus tios Genivaldo e Neilde, que me deram suporte sem igual ao me acolherem em seu lar e me darem todo o carinho como se dá a um filho. Eterna gratidão a todos.

Ao meu orientador, prof Dr. Ícaro, pelo companheirismo, conselhos e todo ensinamento repassado. Aos participantes do projeto em parceria com a Eletrobrás-AL/Equatorial-AL, a partir de onde os dados puderam ser coletados.

Aos professores Prof. Dr Davi Bibiano, Prof. Dr. Thiago Cordeiro, Prof. Dr. Heitor Savino, Prof. Dr. Rodrigo Peixoto e Prof. Dr. João Raphael, por serem muito importantes em minha formação.

Aos amigos que me ajudaram desde que os conheço: Matheus, Gabriel, meu afilhado Ícaro. Aos amigos de faculdade que levarei para vida: Luis (obrigado pelas caronas), Dudu, Glauber, Biu e Rappael. Aos amigos do Ifal: Romário e Gabi.

E a todos aqueles que colaboraram direta ou indiretamente com a produção deste trabalho.

Meus sinceros agradecimentos.

*Gregory Albertt*

*“Ateu! Não é verdade. Tenho passado a vida  
a criar deuses que morrem logo, ídolos que  
depois derrubo” - Graciliano Ramos, Caetés*

# Resumo

O planejamento eficiente e o processo de tomada de decisão relacionados às usinas de energia solar fotovoltaicas exigem cada vez mais um conhecimento acurado da irradiância solar. A análise de séries temporais aplicada a dados climatológicos tem atraído um interesse especial nos últimos anos, visto que o clima interfere diretamente em muitas atividades econômicas. Este trabalho tem como objetivo desenvolver e analisar o uso de modelos de séries temporais SARIMA para previsão de irradiância solar a partir dos dados de uma estação meteorológica localizada em Maceió-AL. Com base nos dados de irradiância, foi aplicado o método de previsão Box-Jenkins para identificar um modelo matemático. Para a estimação dos parâmetros do modelo foi utilizado o método da máxima verossimilhança. Para seleção dos modelos mais adequados foram utilizadas métricas bem definidas, como o erro médio quadrático (MSE). Como consequência, foi possível realizar a previsão do dia seguinte à série temporal estudada e avaliá-la com métodos de análise de erros. Finalmente, neste trabalho foi possível descrever e implementar técnicas de predição nas séries temporais com dados coletados, possibilitando um estudo de caso real, onde foi possível concluir que os modelos construídos são satisfatórios para previsão do dia seguinte a partir das métricas e limites estabelecidos.

***Palavras-chave: Previsão; SARIMA; Box-Jenkins; Séries Temporais; Radiação Solar.***



# Abstract

Efficient planning and decision-making processes related to photovoltaic solar power plants increasingly demand an accurate knowledge of solar irradiance. The analysis of time series applied to climatological data has attracted special interest in recent years, since the climate directly interferes in many economic activities. This work aims to develop and compare models of SARIMA time series to forecast solar irradiance, based on data from a meteorological station located in Maceió-AL. Based on the irradiance data, the Box-Jenkins forecast method was applied to identify a mathematical model. To estimate the model parameters, the maximum likelihood method was used. To select the most appropriate models, well-defined metrics were used, such as the mean square error (MSE). Hence, it was possible to make the forecast for the day ahead of the studied time series and evaluate it with error analysis methods. Finally, in this work it was possible to describe and implement prediction techniques in the time series with collected data, enabling a real case study, where it was possible to conclude that the models built are satisfactory for forecasting a day ahead from the established metrics and limits

***Keywords: Forecast; SARIMA; Box-Jenkins; Time series; Solar radiation.***

# Lista de Figuras

3.1	Fluxograma simplificado do método aplicado. Fonte: autor . . . . .	27
3.2	Fluxograma expandido do método aplicado. Fonte: autor . . . . .	35
4.1	Evolução do consumo de energia global. Adaptado de: [Ritchie and Roser, 2020] . . . . .	36
4.2	Matriz energética mundial. [EPE, 2020] . . . . .	37
4.3	Investimento em energia renovável por tecnologia. Adaptado de: [Ritchie and Roser, 2020] . . . . .	37
4.4	Energia Solar no Brasil . . . . .	38
4.5	Matriz elétrica brasileira. [ABSOLAR, 2020] . . . . .	39
4.6	Componentes da Radiação Solar: difusa, direta e refletida. Fonte: [Viana et al., 2012] . . . . .	40
4.7	Célula fotovoltaica de silício monocristalino. Fonte: [Zilles et al., 2016] . . . . .	42
4.8	Simbologia de um módulo fotovoltaico. Fonte: autor . . . . .	42
4.9	Processo de agrupamento de painel. [de Leva et al., 2004] . . . . .	43
4.10	Esquemático do SFVI. [RIBSOL, 2020] . . . . .	43
4.11	Esquemático do SFVCR. [RIBSOL, 2020] . . . . .	44
4.12	Estação solarimétrica e painéis fotovoltaicos [Bastos, 2020] . . . . .	45
5.1	Séries temporais de irradiância . . . . .	46
5.2	Diferenciações a partir da série original . . . . .	47
5.3	Análise dos gráficos de ACF e PACF . . . . .	48
6.1	Previsões realizadas a partir dos modelos SARIMA que apresentaram me- lhores índices de qualidade . . . . .	56
6.2	Previsão do dia seguinte à série temporal utilizada . . . . .	57
A.1	Análise de Resíduos do modelo SARIMA(1,1,1)(1,1,1) . . . . .	64
A.2	Análise de Resíduos do modelo SARIMA(2,1,2)(1,1,1) . . . . .	65
A.3	Análise de Resíduos do modelo SARIMA(3,1,3)(1,1,1) . . . . .	65
A.4	Análise de Resíduos do modelo SARIMA(4,1,4)(1,1,1) . . . . .	66
A.5	Análise de Resíduos do modelo SARIMA(5,1,5)(1,1,1) . . . . .	66
A.6	Análise de Resíduos do modelo SARIMA(2,1,2)(2,1,2) . . . . .	67

A.7	Análise de Resíduos do modelo SARIMA(3,1,3)(2,1,2) . . . . .	67
A.8	Análise de Resíduos do modelo SARIMA(4,1,4)(2,1,2) . . . . .	68
A.9	Análise de Resíduos do modelo SARIMA(5,1,5)(2,1,2) . . . . .	68
A.10	Análise de Resíduos do modelo SARIMA(3,1,3)(3,1,3) . . . . .	69
A.11	Análise de Resíduos do modelo SARIMA(4,1,4)(3,1,3) . . . . .	69

# Lista de Tabelas

3.1	Comportamento das Funções de Autocorrelação e Autocorrelação Parcial, adaptado de [Box et al., 1994] . . . . .	29
5.1	Testes estatísticos. Fonte: o autor . . . . .	49
5.2	Modelos estimados com sazonalidade 1. Fonte: o autor . . . . .	51
5.3	Modelos estimados com sazonalidade 2. Fonte: o autor . . . . .	52
5.4	Modelos estimados com sazonalidade 3. Fonte: o autor . . . . .	53
6.1	Valores de MAPE e MSE calculados a partir dos modelos estimados . . . . .	55

# Lista de Símbolos

- B** Operador Backward Shift
- F** Operador Forward Shift
- $\nabla$  Operador Backward Difference
- $\mu$  Média
- $\sigma^2$  Variância
- $z(t)$  Série Temporal
- $\varphi(B)$  Operador autoregressivo generalizado
- $p(z)$  Distribuição de probabilidade
- $\rho_k$  Função de autocorrelação
- $\rho_j$  Função de autocorrelação parcial
- $\phi$  Parâmetro autoregressivo
- $\theta$  Parâmetro de média móvel
- $\Phi$  Parâmetro autoregressivo sazonal
- $\Theta$  Parâmetro de média móvel sazonal
- $L(\xi|z)$  Função de verossimilhança
- $l(\xi|z)$  Logarítimo da função de verossimilhança

# Lista de Abreviaturas

**IC** Instituto de Computação.

**ACF** Autocorrelation Function

**PACF** Partial Autocorrelation Function

**AR** Autoregressive

**MA** Moving-Average

**ARMA** Autoregressive Moving-Average

**ARIMA** Autoregressive Integrated Moving Average

**SARIMA** Seasonal Autoregressive Integrated Moving Average

**SFVI** Sistemas Fotovoltaicos Isolados

**SFVCR** Sistemas Fotovoltaicos Conectados à Rede

**MAPE** Mean Absolute Percentage Error

**MSE** Mean Squared Error

# Sumário

<b>1</b>	<b>Introdução</b>	<b>15</b>
1.1	Estado da arte . . . . .	16
1.2	Objetivos do trabalho . . . . .	17
1.2.1	Objetivos específicos . . . . .	17
1.2.2	Organização do texto . . . . .	17
<b>2</b>	<b>Modelos Matemáticos para Previsão</b>	<b>19</b>
2.1	Séries Temporais . . . . .	19
2.2	Modelos Matemáticos Dinâmicos Estocásticos e Determinísticos . . . . .	19
2.2.1	Modelos Estocásticos Estacionários e Não-Estacionários para previsão	20
2.2.2	Média e Variância de um Processo Estacionário . . . . .	20
2.2.3	Operadores . . . . .	20
2.2.4	Funções de Correlação . . . . .	21
2.3	Modelos Estocásticos . . . . .	22
2.3.1	Modelos Auto-Regressivos (AR) . . . . .	23
2.3.2	Modelos de Médias Móveis (MA) . . . . .	23
2.3.3	Modelos Auto-regressivos de Médias Móveis (ARMA) . . . . .	24
2.3.4	Modelos Auto-Regressivos Integrados com Médias Móveis (ARIMA)	25
2.3.5	Modelos Sazonais . . . . .	26
<b>3</b>	<b>Método para Previsão</b>	<b>27</b>
3.1	Método Box-Jenkins . . . . .	27
3.1.1	Identificação . . . . .	28
3.1.2	Estimação . . . . .	30
3.1.3	Diagnóstico . . . . .	32
3.1.4	Previsão . . . . .	33
<b>4</b>	<b>Energia Solar</b>	<b>36</b>
4.1	Energia Solar e Matriz Energética . . . . .	36
4.1.1	Radiação Solar . . . . .	39
4.1.2	Terminologias Importantes . . . . .	40

4.1.3	Sistemas Fotovoltaicos . . . . .	41
4.2	Sistema Fotovoltaico Utilizado . . . . .	44
<b>5</b>	<b>Desenvolvimento</b>	<b>46</b>
5.1	Método Box-Jenkins . . . . .	46
5.1.1	Identificação . . . . .	46
5.1.2	Estimação . . . . .	50
5.1.3	Diagnóstico . . . . .	50
<b>6</b>	<b>Resultados e Discussões</b>	<b>55</b>
6.1	Previsão . . . . .	55
<b>7</b>	<b>Conclusão</b>	<b>58</b>
	<b>Bibliografia</b>	<b>60</b>
<b>A</b>	<b>Resíduos</b>	<b>64</b>



# Capítulo 1

## Introdução

Os dados meteorológicos são relevantes para a geração de energia solar e eólica, de forma que essas informações são decisivas para se mensurar quanta energia pode ser produzida e onde a usina deve ser implantada. Além disso, estes dados podem ajudar a prever possíveis problemas durante o processo de produção, evitando assim, riscos à estrutura de produção e a rede de transmissão de energia. O planejamento eficiente e o processo de tomada de decisão relacionados às usinas de energia solar fotovoltaicas exigem cada vez mais um conhecimento acurado da irradiância solar.

Todavia, o maior problema enfrentado pela geração fotovoltaica é a sua não-continuidade [Yang et al., 2012]. Uma variação brusca de corrente fornecida pelo arranjo de painéis fotovoltaicos pode ser causada pelo sombreamento de nuvens, que resulta em uma rápida queda na irradiação solar que este arranjo recebe. Visto isso, uma previsão precisa da irradiação solar, mesmo que em um curto espaço de tempo, é essencial para a produção de energia cuja fonte é solar fotovoltaica.

No Brasil, de acordo com a Associação Brasileira de Energia Solar Fotovoltaica, do total da matriz elétrica, 1,6% da energia produzida é proveniente de sistemas solares fotovoltaicos, o equivalente a 2,9GW de potência instalada. Esse valor era de 1,19GW, que correspondia a 0,75% da matriz energética naquele ano [ABSOLAR, 2020], e uma das principais tecnologias facilitadoras para a integração da energia solar na rede é a previsão de curto prazo [Reikard et al., 2017].

Nesse contexto, a análise de séries temporais aplicada a dados climatológicos tem atraído um interesse especial nos últimos anos, pois o clima interfere diretamente em muitas atividades econômicas, determinando o sucesso ou fracasso de vários empreendimentos. As séries temporais possibilitam estudar a existência de mudanças ao longo do tempo, em uma determinada variável, sendo importante instrumento para, por meio do comportamento passado, ajustar um modelo matemático para tendências e previsões futuras [Cargnelutti Filho et al., 2010].

## 1.1 Estado da arte

Na literatura, diversos modelos são propostos para a previsão de irradiância solar, entre eles, destacam-se: redes neurais artificiais (RNA) ([Pazikadin et al., 2020]; [Voyant et al., 2017]; [Mellit and Pavan, 2010]), modelos ARIMA (*Autoregressive Integrated Moving Average*) ([Yang et al., 2012]; [Reikard, 2009]) e métodos híbridos que utilizam séries temporais e RNAs ([Ji and Chee, 2011]; [Zhang et al., 1998]; [Zheng and Zhong, 2011]). [Chu et al., 2015] utiliza dados de irradiância e imagens coletadas do céu em uma plataforma que integra técnicas de rastreamento das nuvens e algoritmos de Redes Neurais Artificiais (RNA), e segundo ele, em pequenas escalas (menos de 15min) modelos baseados em dados de imagens frequentemente superam outros métodos de previsão. [Hammer et al., 1999], por sua vez, utiliza imagens de satélites como fonte de dados para previsão e descreve a aplicação de um método estatístico para detectar movimento da estrutura das nuvens a partir de imagens de satélite. [Wang et al., 2018] também utiliza imagens do céu. Dentre os métodos citados, a análise de séries temporais é a que requer menos processamento de dados.

As técnicas baseadas na análise de séries temporais assumem que padrões passados nos dados podem ser usados para prever eventos futuros. Nos últimos anos, uma das formas mais populares de modelagem de séries temporais é a modelagem de média móvel integrada auto-regressiva (ARIMA). Seu principal objetivo é estudar as observações passadas de uma série temporal para desenvolver um modelo apropriado que possa prever valores futuros para a série [Murat et al., 2018]. Esses modelos consideram características, como: (i) sazonalidade; (ii) correlação serial ao longo do tempo e (iii) acomodação de séries não estacionárias.

A análise de modelos ARIMA mostram-se uma ferramenta viável para previsão de irradiância solar uma vez que, segundo [Reikard, 2009], pode ter uma acurácia maior que os modelos híbridos e, particularmente para uma resolução de 60 minutos tem mostrado que supera todos os outros métodos de previsão de séries temporais em quatro das seis estações de testes estudadas. E em horizontes maiores, na ordem de 4 horas ou maiores, NWP (*Numerical Weather Prediction*) possuem as predições mais precisas [Reikard et al., 2017].

E, segundo [Shams et al., 2016], os modelos ARIMA fornecem ainda um procedimento rápido de construção de um processo com uma base estatística rigorosa para a construção de modelos estocásticos adequados para dados meteorológicos e que do ponto de vista prático, estes são uma opção atraente em comparação com outros mais sofisticados.

O sucesso do ARIMA e modelos relativos na previsão de irradiância remontam à sua capacidade de reproduzir ciclos diurnos. Uma revisão prévia sobre técnicas de séries temporais mostra que modelos ARIMA são capazes de superar a maioria das alternativas nesse quesito, incluindo redes neurais e modelos de componentes não observadas, que

utiliza termos trigonométricos para capturar o comportamento cíclico [Reikard, 2009].

Para estes casos em que a série temporal em análise apresenta periodicidade, ou seja, quando padrões se repetem nos dados, utiliza-se na literatura o modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*), que possui a adição de uma componente sazonal.

[Valipour, 2015] compara os modelos ARIMA e SARIMA para previsão de escoamento de chuva nos EUA, numa série histórica de 10 a 20 anos. Em ambas o SARIMA se mostrou mais eficiente para a previsão.

E, uma vez que se aguarda um comportamento sazonal da série temporal a ser estudada (radiação solar), os modelos SARIMA serão considerados neste trabalho.

## 1.2 Objetivos do trabalho

Este trabalho tem como objetivo desenvolver e comparar modelos de séries temporais autoregressivos com média móvel integrada sazonal (SARIMA) através do método Box-Jenkins para previsão de irradiância solar, a partir dos dados de uma estação localizada em Maceió-AL.

### 1.2.1 Objetivos específicos

- Fazer um levantamento das técnicas de previsão de dados meteorológicos utilizados em sistemas de energia solar descritos na literatura;
- Fazer uma breve apresentação dos sistemas de produção de energia solar, afim de descrever em quais situações a previsão de dados climáticos podem ser úteis nesta previsão;
- Descrever e implementar técnicas de predição de dados que utilizam modelos de séries temporais;
- Comparar os modelos e métodos utilizados utilizando um estudo de caso real;

### 1.2.2 Organização do texto

O presente trabalho apresenta a seguinte organização:

Capítulo 2: Neste capítulo, as séries temporais são introduzidas, juntamente com os conceitos básicos para a sua modelagem.

Capítulo 3: O método iterativo de Box-Jenkins para previsão é apresentado e detalhado.

Capítulo 4: Um panorama geral sobre energia solar no Brasil e no mundo são analisados.

Capítulo 5: O método apresentado é agora aplicado à série temporal estudada.

Capítulo 6: Finalmente, são apresentados os resultados e discussões.

# Capítulo 2

## Modelos Matemáticos para Previsão

### 2.1 Séries Temporais

Segundo [Bayer and Souza, 2010], uma série temporal é um conjunto de observações ordenadas no tempo de qualquer fenômeno aleatório. A análise de séries temporais consiste em encontrar relações de dependência existentes temporalmente nos dados, buscando-se identificar o mecanismo gerador da série com o objetivo de extrair periodicidades relevantes nas observações, descrever seu comportamento e fazer previsões. Em outra perspectiva, segundo [Silva et al., 2008], é um conjunto de observações ordenadas em intervalos de tempo, comumente iguais que apresentam dependência serial, constituindo um dos objetivos do estudo de séries temporais analisar e modelar essa dependência. Por fim, [Murat et al., 2018] a reconhece como uma sequência ordenada de valores de uma variável em intervalos de tempo igualmente espaçados, isto é, temperaturas horárias em uma estação meteorológica.

O principal objetivo da modelagem de séries temporais é analisar cuidadosamente e processar rigorosamente as observações passadas das séries temporais para desenvolver um modelo apropriado que descreva a estrutura inerente às séries.

### 2.2 Modelos Matemáticos Dinâmicos Estocásticos e Determinísticos

Segundo [Box et al., 1994], a ideia de usar um modelo matemático para descrever o comportamento de um fenômeno físico é bem estabelecido. Conseqüentemente, é possível deduzir um modelo baseado em leis físicas, que permitem calcular o valor de uma quantia quase exata em qualquer instante de tempo, e, se esse cálculo é possível, tal modelo é denominado determinístico.

Na vida real, praticamente não há modelos determinísticos, uma vez que variáveis desconhecidas não são possíveis de determinar. A exemplo do cálculo de vendas futuras

de um produto, em que existem fatores que não se podem prever.

Existe, no entanto, a possibilidade de auferir um modelo que pode ser utilizado para calcular a probabilidade do número de vendas citado anteriormente estar entre dois limites especificados. Este modelo pode ser chamado de probabilístico ou estocástico.

### 2.2.1 Modelos Estocásticos Estacionários e Não-Estacionários para previsão

Um processo é dito estacionário se suas características se mantêm inalteradas no tempo, ou seja, se suas propriedades estatísticas não mudam com qualquer translação da origem dos tempos.  $z(t)$  e  $z(t + l)$  possuem as mesmas características independentes do  $l$ .

Entretanto, as séries temporais são frequentemente melhores representadas com modelos não-estacionários, usualmente por não possuírem média constante no tempo.

Todavia, no capítulo 3 será apresentada uma maneira de contornar esse problema através de uma transformação aplicada nos dados.

### 2.2.2 Média e Variância de um Processo Estacionário

Quando  $m = 1$ , a suposição de estacionariedade implica que a distribuição de probabilidade  $p(z_t)$  é igual para todos os tempos  $t$ 's e pode ser escrita como  $p(z)$ . Desta forma, o processo estocástico tem média constante que pode ser definida por:

$$\mu = E[z_t] = \int_{-x}^x zp(z) dz \quad (2.1)$$

A média define o nível sob o qual a série temporal flutua.

Outra característica constante do processo estocástico é a variância, que pode ser definida por:

$$\sigma_z^2 = E[(z_t - \mu)^2] = \int_{-x}^x (z - \mu)^2 p(z) dz \quad (2.2)$$

E mede como a o processo estocástico se propaga nesse nível de tempo.

### 2.2.3 Operadores

Seguindo ainda o trabalho de [Box et al., 1994], serão definidos a seguir alguns operadores.

O operador *Backward Shift* (Operador de atraso)  $B$  de uma série temporal é definido como  $Bz_t = z_{t-1}$ , onde  $z_t$  e  $z_{t-1}$  são duas observações consecutivas da série temporal. Consequentemente,  $B^j$  pode ser definido como  $B^j z_t = z_{t-j}$ .

Para alguns *shocks* aleatórios (sendo eventos em que o valor da série temporal se afasta da média da série temporal), extraídos de uma distribuição fixa com média zero e variância  $\sigma_a^2$ , uma sequência  $a_t, a_{t-1}, a_{t-2}, \dots$ , é chamada de processo de ruído branco.

O processo de ruído branco é transformado no processo  $z_t$  por um filtro linear. Essa operação tem como entrada a soma ponderada de *shocks* aleatórios anteriores  $a_t$ , de forma que:

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t \quad (2.3)$$

Dada a sequência  $\psi_1, \psi_2, \dots$  formada pelos pesos, se  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , o filtro é dado como estável e o processo  $z_t$  é estacionário. O parâmetro  $\mu$  é então a média sobre a qual o processo varia.

O operador *Backward Shift* e o Processo de Ruído Branco descrevem os recursos intrínsecos da série temporal: observações adjacentes são dependentes. [Yang et al., 2012].

O operador inverso ao *Backward Shift* é o operador *Forward Shift*  $F = B^{-1}$ , onde  $F_{z_t} = z_{t+1}$  e, conseqüentemente,  $F_{z_t}^m = z_{t+m}$ .

O último operador apresentado é o operador *Backward Difference*  $\nabla$ , que pode ser escrito em termos de  $B$ , uma vez que  $\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$ .

E [Box et al., 1994] ainda afirma que certas séries temporais não estacionárias podem ser obtidas por modelos de média móvel integrada autorregressiva (modelos ARIMA).

## 2.2.4 Funções de Correlação

### Função de Autocorrelação (ACF)

A Autocorrelação é uma característica bastante relevante para o estudo das séries temporais, uma vez que esta consiste na medida da correlação entre os valores de uma série temporal separados por um valor em unidades de tempo  $t$ .

Na prática, esta função é utilizada como ferramenta para escolha da ordem dos parâmetros MA (*Moving Average*) do modelo a ser construído.

Portanto, a partir de um gráfico de ACF é possível determinar a partir de qual atraso a correlação entre os dados já não é significativa para compor o modelo a partir da série temporal estudada.

O coeficiente de Autocorrelação no atraso  $k$  é definido como:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sqrt{E[(z_t - \mu)^2]E[(z_{t+k} - \mu)^2]}} \quad (2.4)$$

Que pode ser escrito ainda como:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t+k} - \mu)]}{\sigma_z^2} \quad (2.5)$$

Para um modelo MA( $q$ ), a função de ACF  $\rho_k$  será não nula para  $k \leq q$  e zero para  $k > q$ . Ou seja, a função ACF tem uma queda a partir do atraso  $q$ .

E, uma vez que para processos estacionários a variância  $\sigma_z^2 = \gamma_0$  é a mesma nos tempos  $t$  e  $t + 1$ . Portanto, a autocorrelação no atraso  $k$ , ou seja, a correlação entre  $z_t$  e  $z_{t+k}$

é:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.6)$$

A função de Correlação é o plot dos coeficientes de autocorrelação  $\rho_k$  como função do atraso  $k$ . Tal função é adimensional, ou seja, independente da escala de medida da série temporal.

### Função de Autocorrelação Parcial (PACF)

A função de autocorrelação parcial (PACF), por sua vez, na prática auxilia na escolha dos parâmetros Autoregressivos (AR).

Analogamente ao ACF com os termos MA, o número de termos AR pode ser sugerido. Ou seja, a função de PACF é uma ferramenta que explora o fato de que a correlação indireta pode ser decisiva.

Desta forma, a partir da análise do gráfico de PACF, é possível determinar em quais atrasos podem existir termos autoregressivos significantes.

A autocorrelação parcial pode ser descrita em termos de  $p$  funções não nulas de autocorrelação. Seja  $\phi_{kj}$  o  $j$ 'ésimo coeficiente em uma representação autoregressiva de ordem  $k$ , de forma que  $\phi_{kk}$  é o último coeficiente, como pode ser visto na equação abaixo:

$$\rho_j = \phi_{k1}\rho_{j-1} + \dots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad (2.7)$$

Para um modelo AR( $p$ ), a função de PACF  $\phi_{kk}$  será não nula para  $k \leq p$  e zero para  $k > p$ . Ou seja, a função PACF tem uma queda a partir do atraso  $p$ .

## 2.3 Modelos Estocásticos

Nesta seção são apresentados os tipos de modelos utilizados para descrever séries temporais.

São eles:

- Modelos Auto-Regressivos (AR)
- Modelos de Médias Móveis (MA)
- Modelos Auto-regressivos de Médias Móveis (ARMA)
- Modelos Auto-Regressivos Integrados com Médias Móveis (ARIMA)
- Modelos auto-regressivos Integrados com Médias Móveis Sazonal (SARIMA)



### 2.3.1 Modelos Auto-Regressivos (AR)

Neste modelo, o valor corrente do processo é expresso por um agregado finito e linear de valores prévios do processo e um choque  $a_t$ .

Sejam os valores de um processo espaçados igualmente no tempo  $t, t-1, t-2, \dots$  iguais a  $z_t, z_{t-1}, z_{t-2}$ . Sejam também  $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}$  desvios de  $\mu$ , de forma que  $\tilde{z}_t = z_t - \mu$ , tem-se que a equação:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t \quad (2.8)$$

Este processo é chamado de Processo auto-regressivo (AR) de ordem  $p$ . Por exemplo, um modelo AR(p) de ordens  $p = 1$  e  $p = 2$  podem ser definidos como:

AR(1)

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t \quad (2.9)$$

AR(2)

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \quad (2.10)$$

[Box et al., 1994] ainda definem um operador auto-regressivo de ordem  $p$ , como:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \tilde{z}_t = a_t \quad (2.11)$$

Que pode ser reescrito como:

$$\phi(B) \tilde{z}_t = a_t \quad (2.12)$$

Que implica que:

$$\tilde{z}_t = \frac{1}{\phi(B)} a_t = \phi^{-1}(B) a_t \equiv \psi(B) a_t \quad (2.13)$$

Desta forma, o modelo autoregressivo pode ser entendido como a saída  $\tilde{z}_t$  de um filtro linear com função de transferência  $\phi^{-1}(B) = \psi(B)$  quando a entrada é um ruído branco  $a_t$

O modelo possui  $p + 2$  parâmetros desconhecidos,  $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_a^2$ , onde o último é a variância do processo de ruído branco  $a_t$ . Os parâmetros são estimados a partir dos dados coletados.

### 2.3.2 Modelos de Médias Móveis (MA)

Outro tipo de modelo utilizado para representar as séries temporais observadas é o de Médias Móveis (MA). Neste modelo,  $\tilde{z}_t$  depende linearmente de um número  $q$  finito de  $a$ 's anteriores, de forma que:

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.14)$$

é um Modelo de Médias Móveis de ordem  $q$ .

Exemplos de modelos de Médias Móveis de ordem  $q = 1$  e  $q = 2$

MA(1):

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} \quad (2.15)$$

MA(2):

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (2.16)$$

[Box et al., 1994] define ainda o operador de Médias Móveis de ordem  $q$  a partir do operador  $B$ :

$$\tilde{z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t \quad (2.17)$$

Que ainda pode ser escrito como:

$$\bar{z}_t = \theta(B) a_t \quad (2.18)$$

Consequentemente, de forma análoga ao modelo anterior (AR), o modelo MA pode ser entendido como a saída  $\tilde{z}_t$  de um filtro linear com função de transferência  $\theta(B)$  quando a entrada é o ruído branco  $a_t$ .

Este modelo possui  $q + 2$  parâmetros desconhecidos  $\mu, \theta_1, \dots, \theta_q, \sigma_a^2$ . De forma similar ao modelo AR, os parâmetros devem ser estimados a partir dos dados observados.

### 2.3.3 Modelos Auto-regressivos de Médias Móveis (ARMA)

A fim de se obter uma flexibilidade maior em modelar uma série temporais, é, muitas vezes, vantajoso incluir termos auto-regressivos e de média móvel em um modelo, o que leva ao modelo Auto-regressivo de Média Móvel (ARMA), definido por:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.19)$$

Dessa forma, existem  $p + q + 2$  parâmetros desconhecidos,  $\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2$ , que são estimados a partir dos dados observados.

Ou ainda:

$$\phi(B) \tilde{z}_t = \theta(B) a_t \quad (2.20)$$

Um exemplo de modelo ARMA(1,1):

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t - \theta_1 a_{t-1} \quad (2.21)$$

E, a partir da equação 2.20, temos que:

$$\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t = \frac{\theta(B)a_t}{\phi(B)} \quad (2.22)$$

Ou seja, do mesmo modo que os modelos apresentados anteriormente, o modelo ARMA também pode ser imaginado como a saída  $\tilde{z}_t$  de um filtro linear, cuja função de transferência é a razão entre os dois operadores polinomiais  $\theta(B)$  e  $\phi(B)$ , quando a entrada é o ruído branco  $a_t$ .

Uma vez que  $\tilde{z}_t = z_t - \mu$ , onde  $\mu$ , onde  $\mu = E[z_t]$  é a média da série estacionária, o modelo pode ser escrito como:

$$\phi(B)z_t = \theta_0 + \theta(B)a_t \quad (2.23)$$

Onde o termo constante  $\theta_0$  é:

$$\theta_0 = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu \quad (2.24)$$

### 2.3.4 Modelos Auto-Regressivos Integrados com Médias Móveis (ARIMA)

Como definido na seção 2.2.1, muitas séries temporais reais apresentam um comportamento não-estacionário, em particular, não variam em torno de uma média fixa. De toda forma, elas podem exibir comportamento homogêneo de algum tipo.

Embora o nível geral em que as flutuações estão ocorrendo sejam diferentes em diferentes períodos de tempo, o comportamento amplo das séries, quando diferenças de nível são permitidas, podem ser similares [Box et al., 1994].

Desta forma, modelos não-estacionários que apresentam esse comportamento homogêneo podem ser descritos por um modelo que exige um número  $d$  de diferenciações para ser estacionário, que na prática está entre 0, 1 ou no máximo 2.

Esse modelo por ser representado por:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.25)$$

onde  $w_t = \nabla^d z_t$ . A operação inversa a esta é  $z_t = S^d w_t$ , onde  $S = \nabla^{-1} = (1 - B)^{-1}$ . Este operador é denominado operador Somatório  $S$ .

A forma geral do modelo ARIMA é:

$$\varphi(B)z_t = \phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \quad (2.26)$$

Algumas considerações precisam ser feitas:  $\phi(B)$  é chamado de Operador Autoregressivo, e se assume que ele é estacionário, ou seja, as raízes de  $\phi(B) = 0$  estão fora do círculo unitário.  $\varphi(B) = \phi(B)\nabla^d$  é chamado de Operador Autoregressivo generalizado. E  $\theta(B)$  é o Operador de Média Móvel.

Exemplo de um modelo ARIMA(1,1,1):

$$\nabla z_t - \phi_1 \nabla z_{t-1} = a_t - \theta_1 a_{t-1} \quad (2.27)$$

Que pode ser reescrito como:

$$(1 - \phi_1 B)\nabla z_t = (1 - \theta_1 B)a_t \quad (2.28)$$

Portanto, o modelo ARIMA pode ser obtido pela soma (ou integração, daí o nome do termo adicional) do modelo ARMA  $w_t$ ,  $d$  vezes.

### 2.3.5 Modelos Sazonais

Os modelos apresentados até aqui possuem componentes não-sazonais, e são muito úteis para representar séries temporais que não possuem essa característica.

Segundo [Box et al., 1994], uma série exibe um comportamento com periodicidade  $s$ , quando similaridades na série ocorrem após  $s$  intervalos de tempo. Em uma série mensal, por exemplo, a tendência da série temporal em repetir um certo tipo de padrão dentro do período de um ano é de uma periodicidade de  $s = 12$ .

#### Modelos auto-regressivos Integrados com Médias Móveis Sazonal (SARIMA)

Um modelo SARIMA pode ser explicado como ARIMA (p, d, q) (P, D, Q) s, onde (p, d, q) é o componente não sazonal do modelo e (P, D, Q) s é o componente sazonal do modelo no qual está a ordem de regressão automática não sazonal, d é o número de diferenciação regular, q é a ordem de Média Móvel não sazonal, P é a ordem de regressão automática sazonal, D é o número de diferenciação sazonal, Q é a ordem da Média Móvel sazonal e s é a duração da temporada [Mirzavand and Ghazavi, 2014].

Pode ser definido pela equação geral:

$$\phi_p(B)\Phi_P(B^s)\nabla^d \nabla_s^D z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (2.29)$$

Exemplo, SARIMA(1,1,1) (1,1,1)<sub>12</sub>

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12}) \quad (2.30)$$

# Capítulo 3

## Método para Previsão

O método de previsão utilizado nesse trabalho é baseado no método iterativo Box-Jenkins, descrito em [Box et al., 1994], pois se trata de um dos mais estudados e aplicados no campo de análise de previsões, devido às suas propriedades e por suportar fortemente evidências empíricas, aplicado a série temporais. Também foi levado em consideração a possibilidade de utilizar modelos sazonais que consideram a sazonalidade nos dados analisados, que é o caso da série estudada neste trabalho.

Seu objetivo principal é realizar a previsão de uma série tomando como referência seus valores passados através de uma correlação temporal existente, gerando uma função de predição onde o erro médio quadrático de  $z_{t+l} - \hat{z}_t(l)$  seja o menor possível, onde  $\hat{z}_t(l)$  indica a previsão feita a partir da origem  $t$  do dado estimado  $z_{t+l}$  em um tempo futuro  $t + l$ , ou seja, no *lead time*  $l$ .

### 3.1 Método Box-Jenkins

Este ciclo é iterativo e composto por quatro etapas: identificação, estimação, diagnóstico e previsão, como pode ser observado na Figura 3.1.

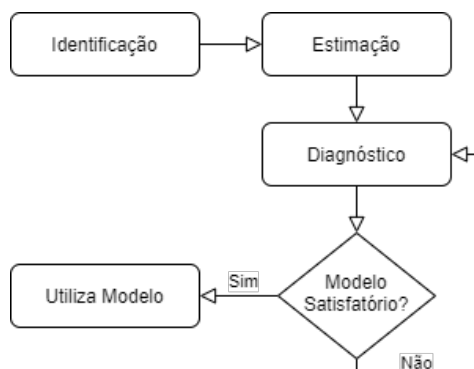


Figura 3.1: Fluxograma simplificado do método aplicado. Fonte: autor

Que podem ser brevemente descritas da seguinte maneira:

- Etapa de identificação: consiste em verificar qual modelo descreve o comportamento da série, valendo-se, por exemplo, de inspeção visual da série temporal e da interpretação dos correlogramas das funções de autocorrelação (ACF) e das funções de autocorrelação parciais (PACF);
- Etapa de estimação: realiza-se a estimativa dos parâmetros do modelo a partir do método da máxima verossimilhança;
- Etapa de diagnóstico: consiste em analisar se o modelo escolhido descreve adequadamente o comportamento da série, por meio da análise dos resíduos;
- Etapa de previsão: etapa que representa o objetivo do trabalho, que consiste em gerar uma função *forecast* a partir da minimização do erro quadrático médio (MSE).

Estas etapas serão mais bem discutidas a seguir.

### 3.1.1 Identificação

“Por identificação, entendemos o uso dos dados e de qualquer informação sobre como a série foi gerada, para sugerir uma subclasse de modelos parcimoniosos dignos de serem cogitados.” [Box et al., 1994]

Ou seja, é o processo aplicado ao conjunto de dados da série temporal para indicar o tipo de modelo que vale a pena ser analisado. Mais especificamente, aponta para indicativos dos valores de  $p$ ,  $d$  e  $q$ , necessários à um modelo ARIMA, por exemplo.

A identificação é inexata, porque a questão de que tipos de modelos ocorrem na prática e em que circunstâncias é uma propriedade do comportamento do mundo físico e, portanto, não pode ser decidida por argumentos puramente matemáticos

Nossa tarefa, então, é identificar uma subclasse apropriada de modelos da família ARIMA, que tem forma:

$$\phi(B)\nabla^d z_t = \theta_0 + \theta(B)a_t \quad (3.1)$$

A abordagem, então, se dá da seguinte maneira:

1. Verificar a estacionariedade do processo  $z_t$  e se necessário aplicar a diferenciação para conseguí-la, de modo que o processo é avaliado em termos de  $w_t$ , a partir de um modelo ARMA:

$$\phi(B)w_t = \theta_0 + \theta(B)a_t \text{ onde,} \quad (3.2)$$

$$w_t = (1 - B)^d z_t = \nabla^d z_t \quad (3.3)$$

Ordem	ACF	PACF
(1,0,d)	Decai exponencialmente	Apenas o primeiro atraso ( $\rho_1$ ) não nulo
(0,d,1)	Apenas o primeiro atraso ( $\phi_{11}$ ) não nulo	Decai exponencialmente
(1,d,1)	Decai exponencialmente a partir do primeiro atraso	Decai exponencialmente a partir do primeiro atraso

Tabela 3.1: Comportamento das Funções de Autocorrelação e Autocorrelação Parcial, adaptado de [Box et al., 1994]

2. Identificar o processo ARIMA resultante, a partir de técnicas de identificação, utilizando principalmente as funções de autocorrelação (ACF) e autocorrelação parcial (PACF).

### Técnicas de Identificação

Como citado acima, deve-se primeiro determinar se a série temporal que se deseja analisar é estacionária. Caso contrário, deve-se transformar a série temporal em uma série de valores estacionários.

E, conforme visto anteriormente na seção 2.2.1, a estacionariedade implica que a média ( $\mu$ ) e a variância ( $\sigma^2$ ) são invariantes na série temporal.

Essas características podem ser analisadas de diversas formas. Uma maneira simples é analisar o gráfico plotado da série temporal, que pode ser vista flutuando através de uma média fixa com uma propagação constante.

Uma vez observado que a série temporal não é estacionária, é comum aplicar uma diferenciação na série a fim de estabilizar sua média. A diferenciação é frequentemente utilizada para transformar uma série temporal não estacionária em uma série temporal estacionária [Zaharim et al., 2009].

Nesse contexto, a amostra ACF (Função de Autocorrelação) e a amostra PACF (Função de Autocorrelação Parcial) são usadas para revelar a estrutura correlativa das séries temporais. A autocorrelação mede o grau de associação entre uma observação no tempo  $t$  e uma observação  $p$  períodos à frente.

A inspeção visual das funções de ACF e PACF fornece uma maneira útil de construir um modelo ARMA ( $p, q$ ), a partir da Tabela 3.1.

### Teste Estatísticos

A estacionariedade de uma série temporal pode ser verificada também através de testes estatísticos, e, como visto anteriormente, essa condição é fundamental para a realização das demais etapas do método utilizado.

Na literatura são utilizados diversos testes estatísticos e neste trabalho foram utilizados os testes de Dickey-Fuller [Dickey and Fuller, 1979], de *Phillips-Perron* [Phillips and Perron, 1988], de KPSS [Kwiatkowski et al., 1992] e o de *Leybourne-McCabe* [Leybourne and HcCabe, 1999] a fim de se confirmar a estacionariedade da série analisada.

Segundo o Teste de Dickey-Fuller e Phillips-Perron, sua hipótese nula é não estacionariedade, logo, se o resultado do teste for  $h = 0$ , indica que falhou em rejeitar a hipótese nula, portanto não há evidências que possam sugerir a estacionariedade da série.

Ou seja:

$$\begin{aligned} h = 1, Z_n \text{ é estacionária} \\ h = 0, Z_n \text{ não é estacionária} \end{aligned} \quad (3.4)$$

Em contrapartida, os testes de KPSS e *Leybourne-McCabe* testam a hipótese nula da estacionariedade contra a hipótese alternativa, da seguinte maneira:

$$\begin{aligned} h = 0, Z_n \text{ é estacionária} \\ h = 1, Z_n \text{ não é estacionária} \end{aligned} \quad (3.5)$$

### 3.1.2 Estimação

“Por estimativa, entendemos o uso eficiente dos dados para fazer inferências sobre parâmetros condicionais à adequação do modelo cogitado” [Box et al., 1994].

Uma vez que processo de identificação nos levou a uma formulação provisória do modelo, agora nós precisamos obter estimativas eficientes dos parâmetros. Após a estimação, o modelo será sujeito a uma verificação de diagnóstico e testes de qualidade de ajuste.

No presente trabalho, seguindo o que foi proposto por [Box et al., 1994], a estimação dos parâmetros foi feita a partir do Método de Máxima Verossimilhança.

#### Estimação por Máxima Verossimilhança

Neste método, o problema de estimação é baseado nos resultados obtidos pela série temporal, determinando-se sua distribuição com a maior probabilidade de ter gerado esta série.

#### A Função de Verossimilhança

Seja  $z$  uma série temporal de  $N$  amostras, que será associada a uma variável randômica com  $N$  dimensões e distribuição de probabilidade conhecida  $p(z|\xi)$ , que depende dos parâmetros desconhecidos  $\xi$ . No caso particular de um modelo ARIMA, o  $\xi$  se refere ao vetor de parâmetros  $p + q + 1$  ( $\phi, \theta, \sigma$ ).



Antes de utilizar os dados,  $p(z|\xi)$  associará uma densidade a cada observação da série temporal, para um  $\xi$  fixo. Após isso, o método nos leva a contemplar os vários valores de  $\xi$  que podem ter dado origem ao conjunto fixo de observações  $z$ . A função utilizada para este propósito é a Função de Verossimilhança  $L(\xi|z)$ , que é da mesma forma que  $p(z|\xi)$ , porém agora o conjunto de observações  $z$  é fixo, enquanto  $\xi$  pode variar livremente.

Segundo [Box et al., 1994], é conveniente trabalhar com o logaritmo da função de verossimilhança  $\ln[L(\xi|z)] = l(\xi|z)$ .

Assumindo que esta série  $z$  pode ser gerada por um modelo ARIMA (p,d,q). A partir dessas observações é possível gerar uma outra série temporal  $w$  de  $n$  observações ( $n = N - d$ )  $w_1, w_2, \dots, w_n$ , onde  $w_t = \nabla^d z_t$ .

Destarte, o problema de ajuste dos parâmetros  $\phi$  e  $\theta$  de um modelo ARIMA é equivalente ao ajuste da série estacionária  $w$  por um modelo ARMA(p,q), que pode ser escrito como:

$$a_t = w_t - \phi_1 w_{t-1} - \phi_2 w_{t-2} - \dots - \phi_p w_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_q a_{t-q} \quad (3.6)$$

Como visto anteriormente na seção (citar), para  $d > 0$  é apropriado assumir que  $\mu = 0$ . E assumindo que  $a$  possui distribuição normal, sua densidade de probabilidade é:

$$p(a_1, a_2, \dots, a_n) = (\sigma_a)^{-n} \exp \left[ - \left( \sum_{t=1}^n \frac{a_t^2}{2\sigma_a^2} \right) \right] \quad (3.7)$$

Logo, dado o conjunto de dados  $w$ , a função logarítmica associada aos parâmetros  $(\phi, \theta, \sigma_a)$  seria:

$$l_*(\phi, \theta, \sigma_a) = -n \ln(\sigma_a) - \frac{S_*(\phi, \theta)}{2\sigma_a^2} \quad (3.8)$$

Onde:

$$S_*(\phi, \theta) = \sum_{t=1}^n a_t^2(\phi, \theta | w_*, a_*, w) \quad (3.9)$$

Na equação acima, o asterisco subscripto é utilizado para enfatizar que são condicionais a escolha dos valores iniciais.

## Análise dos modelos

O critério de informação de Akaike (AIC) [Akaike, 1974] utiliza uma estrutura rigorosa de análise de informações com base no conceito de entropia. A ideia principal por trás da AIC é examinar a complexidade do modelo, juntamente com a qualidade de seu ajuste aos dados da amostra, e produzir uma medida que se equilibre entre os dois. Sua fórmula é [Brockwell and Davis, 2016]

$$AIC : c_n(k) = 2 \times (k/n) - 2 \times \ln(L)/n \quad (3.10)$$

onde  $L$  é a função de probabilidade,  $k$  é o número de parâmetros no modelo e  $n$  é o número de observações.

O *Hannon-Quinn Information Criterion* (HIC) e o *Bayesian Information Criterion* (BIC) são descritos como: [Hannan and Quinn, 1979] e [Schwarz et al., 1978].

$$HIC : c_n(k) = 2k \ln(\ln(n))/n - 2 \ln(L)/n \quad (3.11)$$

$$BIC : c_n(k) = k \ln(n)/n - 2 \ln(L)/n \quad (3.12)$$

Um outro critério que pode ser utilizado é o *Final Prediction Error* (FPE), descrito como: [Brockwell and Davis, 2016]

$$FPE : c_n(k) = [(1 + k/n)/(1 - k/n)] \times L \quad (3.13)$$

### 3.1.3 Diagnóstico

“Por verificação de diagnóstico, queremos dizer a verificação do modelo ajustado em sua relação com os dados, com a intenção de revelar as inadequações do modelo e, assim, obter melhorias no modelo ” [Box et al., 1994].

Uma vez que o modelo foi identificado e seus parâmetros estimados, uma verificação de diagnóstico é então aplicada ao modelo ajustado a fim de se verificar o quão bem o modelo se ajusta à série temporal observada.

Ou seja, agora que o modelo foi identificado e estimado, deve-se decidir se ele é adequado, ou se deve ser modificado e voltar ao início do ciclo apresentado na Figura 3.2.

É importante, portanto, descobrir de que maneira o modelo é inadequado, para que as mudanças corretas possam ser realizadas e para que o modelo reajustado seja mais adequado.

#### ***Overfitting***

Uma das técnicas que pode ser utilizada para checagem de diagnóstico é o *overfitting*. A partir de um modelo identificado como explicitado na seção 3.1.1 e estimado, mostrado na seção 3.1.2 e acreditando que se trata de um modelo correto, ajusta-se um modelo mais elaborado.

Esta técnica pode ser perigosa porque quanto mais elaborado o modelo, mais parâmetros adicionais que podem cobrir direções de discrepância. Porém não havendo ganhos devido à adição dos parâmetros, pode-se perceber que a tentativa falhou.

### Análise de Resíduos

Para utilizar o método do *overfitting*, supomos que conhecemos as discrepâncias que podem estar inerentes ao acréscimo de parâmetros. Um processo menos dependente desse conhecimento prévio é o de Análise de Resíduos.

Supondo um modelo  $\phi(B)\tilde{w}_t = \theta(B)a_t$  que foi identificado e estimado a partir do método da Máxima Verossimilhança aplicado à uma série temporal foram obtidos os parâmetros  $(\hat{\phi}, \hat{\theta})$ , a quantidade:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)\tilde{w}_t \quad (3.14)$$

é referida como os resíduos, que são calculados recursivamente, de forma:

$$\hat{a}_t = \tilde{w}_t - \sum_{j=1}^p \hat{\phi}_j \tilde{w}_{t-j} + \sum_{j=1}^q \hat{\theta}_j \tilde{a}_{t-j} \quad (3.15)$$

Além disso, plotagens de ACF para o resíduo podem ser utilizadas para testar as correlações seriais entre os valores residuais usando limites estatísticos em um nível de confiança predefinido.

Para confirmar essa avaliação, além de uma inspeção visual da função de ACF dos resíduos, o *Ljung-Box Q-Test* [Ljung and Box, 1978] também pode ser utilizado para verificar as autocorrelações dos resíduos estimados.

O teste é definido como:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{\hat{r}_j^2}{(n-j)} \quad (3.16)$$

De forma que se o modelo for apropriado a estatística  $Q(k)$  terá aproximadamente uma distribuição qui-quadrado. A hipótese de ruído branco para os resíduos é rejeitada para valores grandes de  $Q(K)$

Uma vez que o diagnóstico apresentado indica que o modelo não é adequado, é necessário considerar de que maneiras ele pode ser modificado.

### 3.1.4 Previsão

Nesta última etapa, a partir de um modelo ARIMA que foi identificado, estimado e passado pela checagem de diagnóstico, podemos utilizá-lo para realizar previsões.

#### Previsão a partir do Erro Quadrático Médio (MSE)

Seja um modelo ARIMA:

$$\varphi(B)z_t = \theta(B)a_t \quad (3.17)$$

onde  $\varphi(B) = \phi(B)\nabla^d$ .

Será considerada previsão um valor  $z_{t+l}$ ,  $l \geq 1$ , quando parte-se do tempo corrente  $t$ , ou seja, é uma previsão que se inicia na origem  $t$  para um tempo futuro  $l$ .

Logo, uma observação futura  $z_{t+l}$  gerada por um modelo ARIMA, pode ser escrita em termos da equação diferença:

$$z_{t+l} = \varphi_1 z_{t+l-1} + \dots + \varphi_{p+d} z_{t+l-p-d} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q} + a_{t+l} \quad (3.18)$$

Supondo agora que, a partir da origem  $t$ , será feita uma previsão  $\hat{z}_t(l)$  de  $z_{t+l}$  que deve ser uma função linear das observações atual e anteriores  $z_t, z_{t-1}, z_{t-2}, \dots$ . Portanto, será também uma função linear dos valores atual e anteriores dos choques  $a_t, a_{t-1}, a_{t-2}, \dots$

E supondo também que a melhor previsão seria:

$$\hat{z}_{t+l} = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \dots \quad (3.19)$$

Onde os pesos  $\psi_l^*, \psi_{l+1}^*, \dots$  serão determinados.

Como visto anteriormente na seção 2, um modelo ARIMA também pode ser escrito como:

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} = \psi(B)a_t \quad (3.20)$$

Logo, pode-se escrevê-la da forma:

$$z_{t+l} = \sum_{j=0}^{\infty} \psi_j a_{t+l-j} \quad (3.21)$$

Consequentemente, o erro quadrático médio pode ser escrito como:

$$E[z_{t+l} - \hat{z}_t(l)]^2 = (1 + \psi_1^2 + \dots + \psi_{l-1}^2) \sigma_a^2 + \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_a^2 \quad (3.22)$$

Que é minimizado quando  $\psi_{l+j}^* = \psi_{l+j}$

Portanto, tem-se:

$$z_{t+l} = (a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1}) + (\psi_l a_t + \psi_{l+1} a_{t-1} + \dots) = e_t(l) + \hat{z}_t(l) \quad (3.23)$$

Onde  $e_t(l)$  é o erro da previsão  $\hat{z}_t(l)$  em um tempo futuro  $l$ .

Quando  $\hat{z}_t(l)$  é considerado uma função de  $l$  para um  $t$  fixo, será chamado de Função de Previsão.

Uma vez realizadas as previsões a partir dos modelos selecionados, estas serão avaliadas

quanto a sua acuracidade por meio do valor do *Mean Absolute Percentage Error* (MAPE), que pode ser definido como:

$$MAPE = \frac{1}{n} \sum \left| \frac{z_t - \hat{z}_t}{z_t} \right| \times 100 \quad (3.24)$$

O MAPE é o valor do erro percentual médio absoluto. Uma vez que utiliza valores percentuais se torna independente da escala da série temporal. Desta forma, quanto menor o seu valor, melhor a acurácia da previsão.

Outra métrica que é utilizada para verificar a qualidade da previsão é o *Mean Squared Error* (MSE), é definido como:

$$\frac{\sum (z_t - \hat{z}_t)^2}{n} \quad (3.25)$$

O MSE é a média da soma dos quadrados dos erros, ou seja, atribui um maior relevância aos erros de maior magnitude, sendo assim, sensíveis à presença de *outliers*.

A Figura 3.2 ilustra o método iterativo descrito.

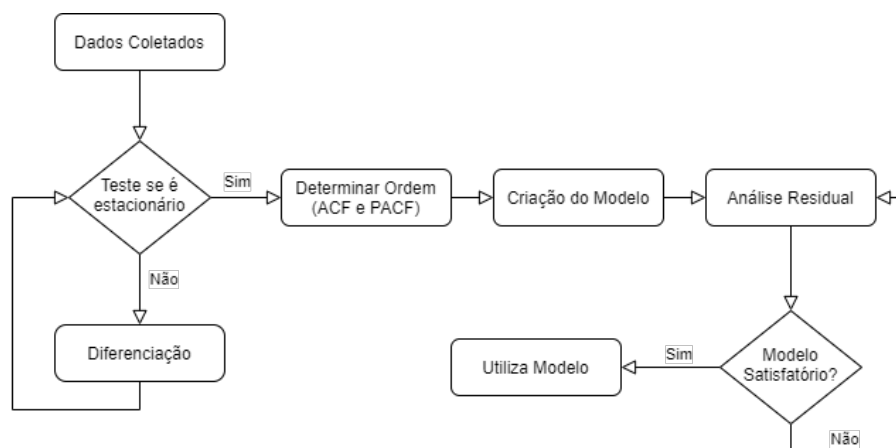


Figura 3.2: Fluxograma expandido do método aplicado. Fonte: autor

# Capítulo 4

## Energia Solar

### 4.1 Energia Solar e Matriz Energética

Após a revolução industrial que ocorreu por volta do século XVIII, percebe-se, para além de suas transformações sociais, econômicas e de relações de trabalho, um aumento no consumo de energia e a dependência das fontes energéticas da época.

Dessas, o carvão tornou-se a principal fonte energética utilizada em todo o mundo. Com o passar dos anos, outras fontes de energia foram sendo incluídas na matriz energética mundial, como pode ser observado na Figura 4.1.

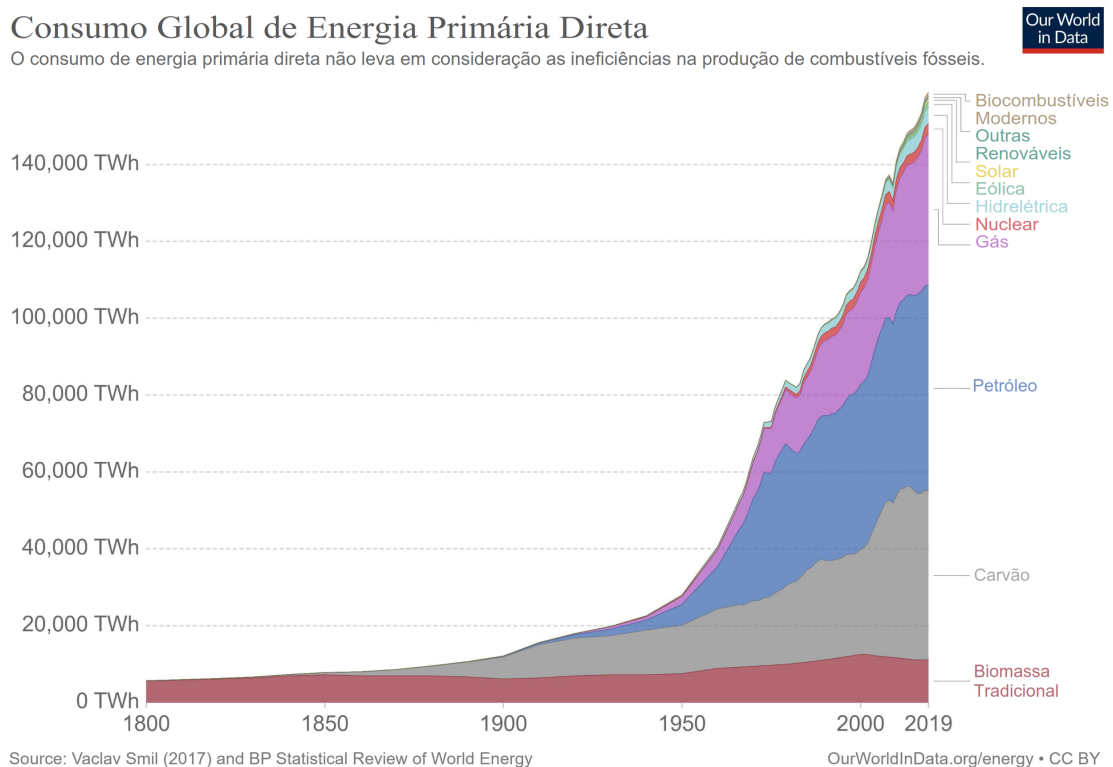


Figura 4.1: Evolução do consumo de energia global. Adaptado de: [Ritchie and Roser, 2020]

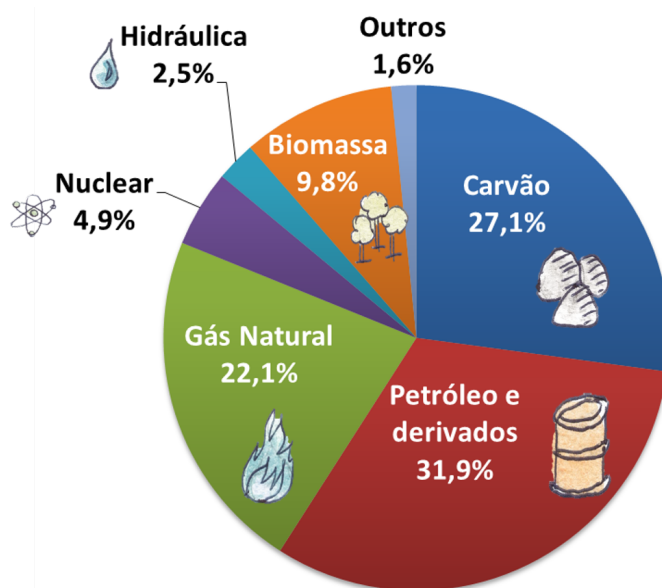


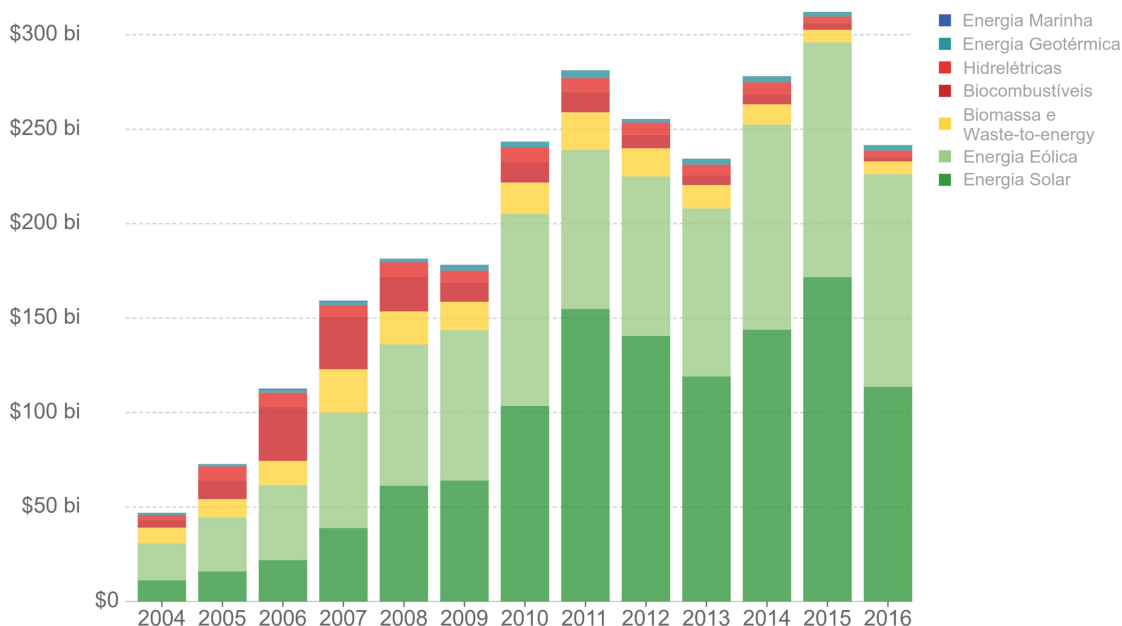
Figura 4.2: Matriz energética mundial. [EPE, 2020]

O mundo inteiro ainda depende de carvão, petróleo e gás (combustíveis fósseis) para mais de 80% das necessidades atuais de energia - situação que mostra poucos sinais de mudança no médio prazo, sem mudanças drásticas na política. A demanda por energia deverá aumentar em 50% nas duas décadas seguintes. [Ashok Kumar et al., 2015]

### Investimento em Energias Renováveis, por tecnologia

investimento global em tecnologias de energia renovável, medido em dólares por ano. Nota: os números do investimento excluem esquemas hidrelétricos de grande escala

Our World  
in Data



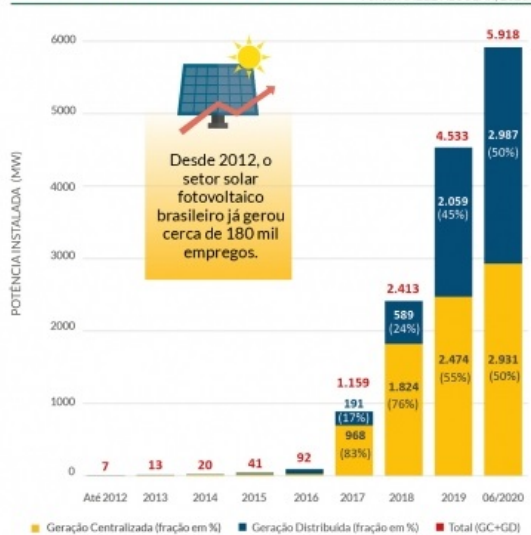
Source: International Renewable Energy Agency (IRENA)

OurWorldInData.org/energy-production-and-changing-energy-sources/ • CC BY

Figura 4.3: Investimento em energia renovável por tecnologia. Adaptado de: [Ritchie and Roser, 2020]

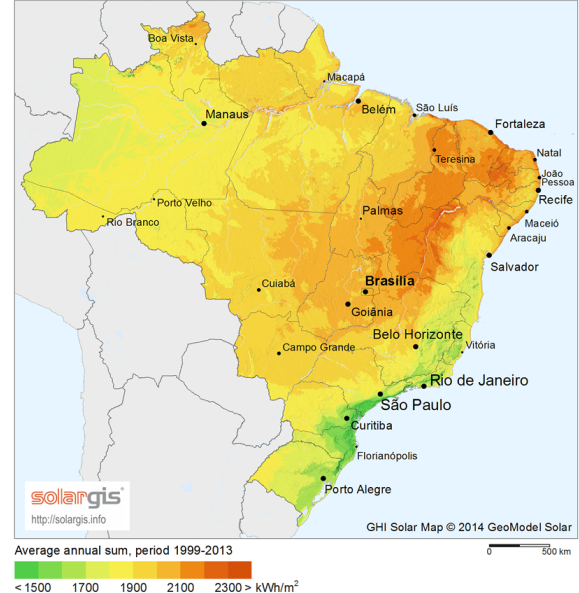
A partir da Figura 4.2, percebe-se uma forte dependência dos combustíveis fósseis no mundo, o que pode ser danoso em dois aspectos: primeiro, trata-se de fontes não-renováveis de energia, ou seja, fontes finitas. E segundo, são fontes produtoras de gases de efeito estufa, que aceleram as mudanças climáticas.

### Evolução da Fonte Solar Fotovoltaica no Brasil



(a) Evolução da Fonte Solar Fotovoltaica no Brasil. [ABSOLAR, 2020]

### Global Horizontal Irradiation (GHI) Brazil



(b) Potencial de radiação no Brasil. [Solargis, 2019]

Figura 4.4: Energia Solar no Brasil

Esses efeitos da dependência dos combustíveis fósseis têm gerado um esforço global pela inserção de fontes energéticas mais limpas, e diversos países como Alemanha e Reino Unido são reconhecidos pelo alto investimento em pesquisas relacionadas à busca de diversificação da matriz energética.

Investir em energias renováveis para diversificar a matriz energética é uma abordagem esclarecida para supervisionar os interesses energéticos e, conseqüentemente, para lidar com as emissões de carbono. Atualmente, na maioria das nações, regulamentos e forças motivadoras orçamentárias são criadas para fortalecer a eficiência energética.

Neste contexto, se destacam como fontes de energias alternativas renováveis as fontes solar e eólica, pois nos últimos anos tem havido um crescimento acentuado no investimento nessas fontes de energia como pode ser observado na Figura 4.3, o que pode indicar um elevado potencial de crescimento para os próximos anos.

A partir dela, pode-se verificar o destaque para investimentos em energia solar e eólica, a tendência, portanto, é que esse investimento continue crescendo e, conseqüentemente, que os custos da tecnologia diminuam.

No Brasil, de acordo com a Associação Brasileira de Energia Solar Fotovoltaica, do total da matriz elétrica, 1,6% da energia produzida é proveniente de sistemas solares



fotovoltaicos, o equivalente a 2,9GW de potência instalada. Esse valor era de 1,19GW, que correspondia a 0,75% da matriz energética naquele ano [ABSOLAR, 2020].

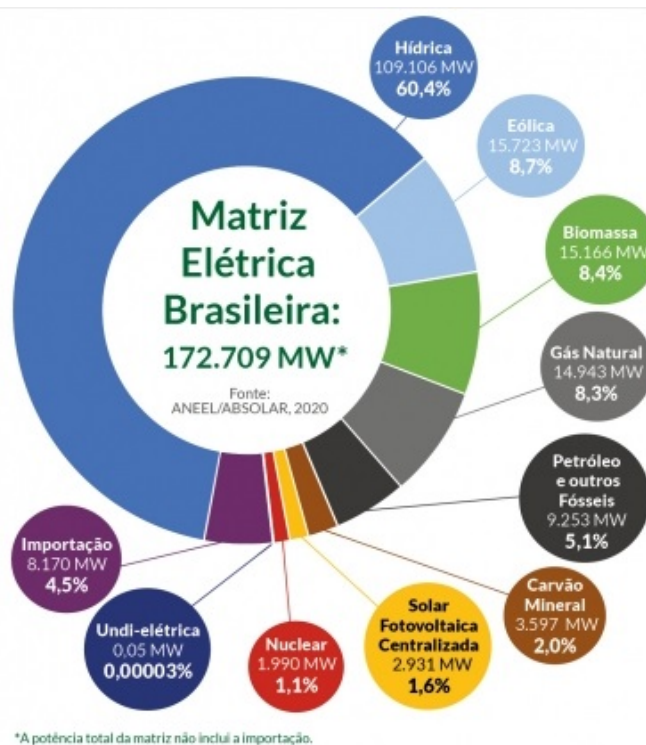


Figura 4.5: Matriz elétrica brasileira. [ABSOLAR, 2020]

Esse crescimento da energia solar no Brasil pode ser visto na Figura 4.4a e pode ser entendido uma vez que o potencial de irradiação solar no país é a maior do mundo, segundo o Brazilian Atlas for Solar Energy Resource: SWERA Results [Martins et al., 2008], contando com uma insolação superior a 3000 horas por ano. Nesse contexto a energia solar surge como alternativa.

Em 2010, a contribuição da energia fotovoltaica correspondia a uma fração muito pequena da matriz energética. Todavia, nos últimos anos foi observado um rápido crescimento na instalação de energia solar fotovoltaica e uma forte tendência de crescimento.

No cenário de novas políticas, a energia gerada a partir de energia solar fotovoltaica em 2035 é mais de 26 vezes a de 2010, expandindo de 32 Twh para 846 Twh. A energia solar fotovoltaica instalada aumentaria de 67 GW em 2011 para 600 GW em 2035, devido à redução de despesas e auxílios governamentais [Ashok Kumar et al., 2015].

#### 4.1.1 Radiação Solar

A radiação solar que atinge o plano horizontal da célula fotovoltaica é composto pela radiação direta, difusa e refletida, como pode ser observado na Figura 4.6. A primeira se refere à incidência direta dos raios solares no plano, e a segunda, aos raios que chegam de maneira indireta devido a desvios (ex: nuvens) e a última se refere à radiação que é

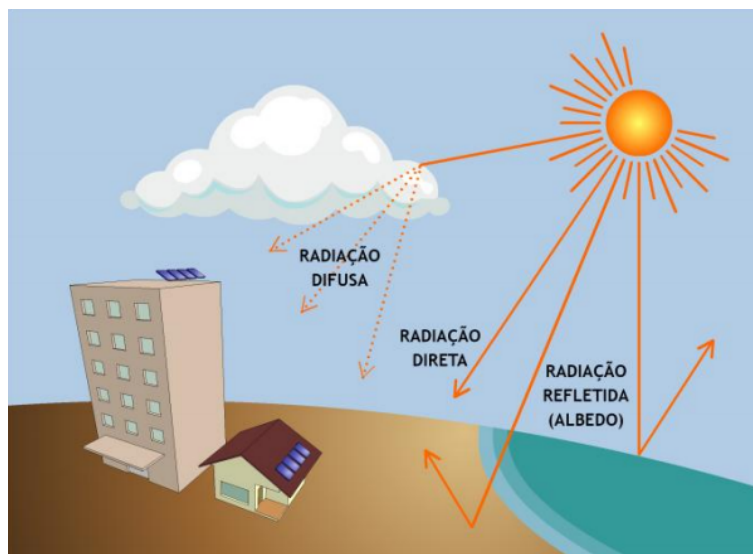


Figura 4.6: Componentes da Radiação Solar: difusa, direta e refletida. Fonte: [Viana et al., 2012]

reenviada para o espaço.

Nos sistemas fotovoltaicos é importante conhecer a quantidade de luz do dia acessível em uma área específica em um determinado momento. A radiação solar, ou irradiância, é uma quantidade instantânea de energia em unidades de  $W/m^2$ . A radiação solar muda durante o dia de  $0 W/m^2$  durante a noite para um valor de aproximadamente  $1000W/m^2$ .

### 4.1.2 Terminologias Importantes

A NBR 10899:2006 [ABNT, 2006] intitulada Terminologia sobre Energia Solar Fotovoltaica, define os principais termos técnicos sobre energia solar. A seguir, apresentamos os que são essenciais a este trabalho:

- Irradiância solar ( $G$ ) - Taxa na qual a radiação solar incide em uma superfície, por unidade de área, normalmente medida em watt por metro quadrado ( $W/m^2$ ).
- Irradiância direta ( $G_{DIR}$ ) - Irradiância solar incidente em uma superfície, sem ter sido espalhada pela atmosfera, podendo ser horizontal ou normal.
- Irradiância difusa ( $G_{DIF}$ ) - Potência radiante do céu, recebida em uma unidade de área em uma superfície horizontal, excluída a irradiância direta
- Irradiância global ( $G_{HOR}$ ) - Potência radiante solar, recebida em uma unidade de área em uma superfície horizontal, que é igual à irradiância direta mais a irradiância difusa.
- Irradiação solar (I ou H) - Irradiância solar integrada durante um intervalo de tempo especificado, normalmente uma hora ou um dia, é medida em watt hora por metro

quadrado ou Joule por metro quadrado ( $\text{Wh/m}^2$  ou  $\text{J/m}^2$ ). É simbolizada por “I”, quando integrada em uma hora e por “H”, quando integrada em um dia.

- Corrente de curto-circuito ( $I_{sc}$ ) - Corrente de saída de um conversor fotovoltaico, na condição de curto-circuito e para valores preestabelecidos de temperatura e irradiância total.
- Tensão de circuito aberto ( $V_{oc}$ ) - Tensão gerada através de um conversor fotovoltaico sem carga (aberto), para valores preestabelecidos de temperatura e irradiância total.
- Condições-padrão de referência para ensaio (STC) - Temperatura de junção da célula em  $25^\circ\text{C}$ , irradiância total de  $1.000 \text{ W/m}^2$  normal à superfície de ensaio com espectro AM 1,5.
- Potência máxima ( $P_{mp}$ ) - Potência em um ponto da curva característica de um conversor fotovoltaico, onde o produto da corrente (I) pela tensão (V) é máximo, no quadrante de geração.
- Potência de pico ( $W_p$ ) - Potência nominal de saída de um conversor fotovoltaico, sob as condições padrão de referência para ensaio

### 4.1.3 Sistemas Fotovoltaicos

A transformação da energia contida na radiação luminosa em energia elétrica é um fenômeno físico conhecido como efeito fotovoltaico [Zilles et al., 2016]. Um dispositivo que tem como característica converter energia solar em eletricidade é denominado como Gerador Fotovoltaico.

Células fotovoltaicas são dispositivos construídos a partir de semicondutores, que possibilita o efeito fotovoltaico para a produção de energia elétrica. A maioria das células fotovoltaicas utilizam o silício como base para sua fabricação.

A Figura 4.7 apresenta os detalhes de uma célula fotovoltaica de silício monocristalino.

As células fotovoltaicas mais utilizadas são capazes de gerar na máxima potência, numa condição de sol de  $1.000\text{W/m}^2$  e temperatura de célula de  $25^\circ\text{C}$ , uma corrente da ordem de  $32\text{mA/cm}^2$ , ou seja, cada uma gera uma corrente entre  $1,5\text{A}$  ( $50 \text{ cm}^2$ ) e  $4,5\text{A}$  ( $150 \text{ cm}^2$ ) numa tensão entre  $0,46\text{V}$  e  $0,48\text{V}$ . Na prática, as células fotovoltaicas são agrupadas em associações série e paralelo para produzir corrente e tensão adequadas às aplicações elétricas a que se destinam [Zilles et al., 2016].

Um Módulo Fotovoltaico é resultado do agrupamento das células fotovoltaicas para a configuração desejada, como está ilustrado na figura 4.9. Estes módulos fotovoltaicos são representados através de características elétricas como sua potência nominal (W), tensão (V) e corrente(I).

A Figura 4.8 apresenta a simbologia de um módulo fotovoltaico.

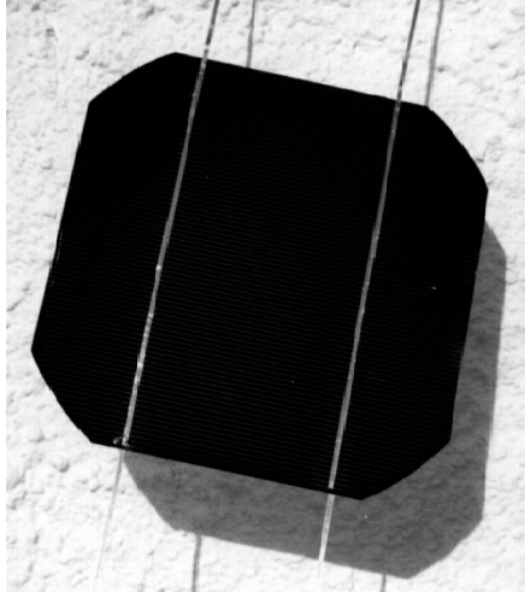


Figura 4.7: Célula fotovoltaica de silício monocristalino. Fonte: [Zilles et al., 2016]

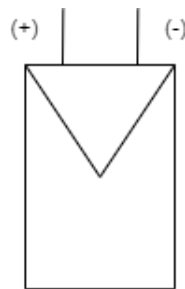


Figura 4.8: Simbologia de um módulo fotovoltaico. Fonte: autor

A máxima potência ( $P_{mp}$ ) fornecida em watt-pico ( $W_p$ ), que é obtida quando se apresentam corrente de máxima potência ( $I_{mp}$ ) e a tensão de máxima potência ( $V_{mp}$ ).

Outros parâmetros também são importantes para a modelagem e representação de um módulo fotovoltaico, são eles: corrente de curto circuito ( $I_{sc}$ ) e tensão de circuito aberto ( $V_{oc}$ ).

Um dos principais parâmetros de uma célula fotovoltaica é a eficiência de conversão de energia ( $\eta$ ), e é definida como a razão entre a máxima potência gerada e a potência incidente, que depende da radiação no plano da célula. De forma que:

$$\eta = \frac{P_{GERADA}}{P_{INCIDENTE}} \quad (4.1)$$

Os sistemas de energia solar fotovoltaica podem ser divididos basicamente em duas categorias:

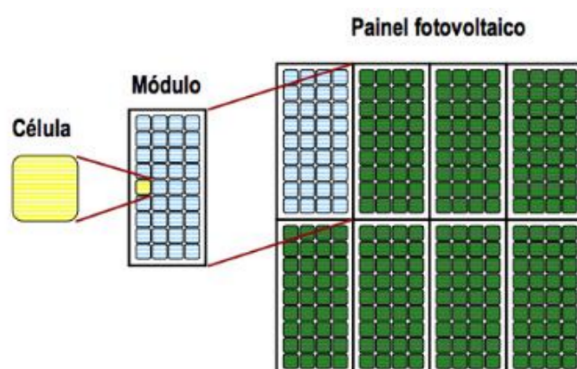


Figura 4.9: Processo de agrupamento de painel. [de Leva et al., 2004]

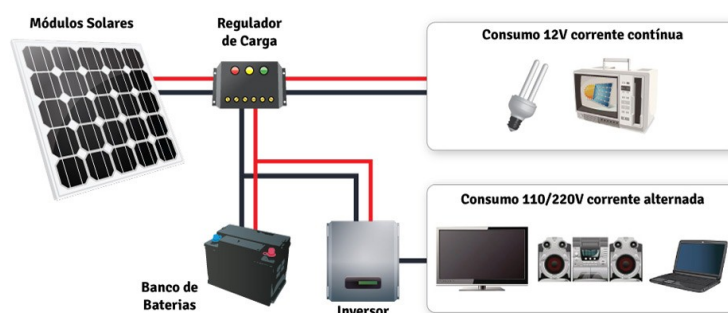


Figura 4.10: Esquemático do SFVI. [RIBSOL, 2020]

## Sistemas Fotovoltaicos Isolados - SFVI

As primeiras aplicações terrestres da tecnologia fotovoltaica ocorreram principalmente com sistemas isolados, capazes de abastecer cargas distantes da rede convencional de distribuição de eletricidade. [Zilles et al., 2016]

Os sistemas isolados não fazem conexão com a rede elétrica pública, tendo seu funcionamento baseado em armazenamento de energia em baterias, que garantem a autonomia do sistema. Desta forma, são comumente utilizados em locais onde não há abastecimento do sistema público de energia. Seu esquemático pode ser observado 4.10

Composição de um SFVI:

**Módulo Fotovoltaico:** como visto anteriormente, tem a função de converter a radiação solar em energia elétrica através do efeito fotovoltaico.

**Controlador de carga:** tem como função proteger o banco de baterias de uma possível sobrecarga ou de uma descarga profunda, monitorando e controlando sua carga e descarga.

**Inversor:** converte a tensão contínua oriunda dos painéis fotovoltaicos ou do banco de baterias em corrente alternada. Necessário caso as cargas funcionem em corrente alternada.

**Bateria:** tem como função armazenar a energia excedente a fim de ser utilizada enquanto não há geração de energia, seja à noite ou em dias chuvosos.

## Sistemas Fotovoltaicos Conectados à rede - SFVCR

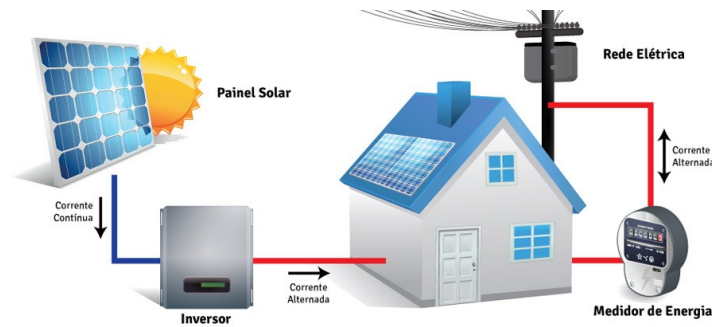


Figura 4.11: Esquemático do SFVCR. [RIBSOL, 2020]

Os sistemas fotovoltaicos conectados à rede, por sua vez, possuem conexão com a rede pública de abastecimento. Sendo, portanto, desnecessária a utilização de dispositivos para armazenamentos, ao contrário dos SFVI.

A energia elétrica gerada em excesso pelo sistema é injetada na rede elétrica, de forma que durante os momentos em que não há geração de energia, o sistema utiliza a rede pública como fonte.

Quanto à sua composição, difere do SFVI por não necessitar do banco de baterias e consequentemente do controlador de cargas e se faz necessária a instalação do medidor bidirecional, que é responsável por monitorar a energia que é injetada na rede. No demais, os dispositivos são os mesmos: painéis fotovoltaicos responsáveis por gerar a energia elétrica e o inversor que transforma a corrente contínua em alternada.

## 4.2 Sistema Fotovoltaico Utilizado

Para este trabalho, os dados foram coletados de uma estação meteorológica situada no Instituto de Computação (IC) da Universidade Federal de Alagoas (Ufal).

O sistema fotovoltaico em questão compõe uma estação solarimétrica e possui 4 painéis solares de 330 Watts de potência.

Nesta estação foi utilizado um piranômetro de baixo custo da empresa LuZdoSOL, possuindo um intervalo de medição entre 0 e  $1479W/m^2$



(a) Estação solarimétrica



(b) Painéis fotovoltaicos

Figura 4.12: Estação solarimétrica e painéis fotovoltaicos [Bastos, 2020]

# Capítulo 5

## Desenvolvimento

Como citado anteriormente, os dados utilizados no presente trabalho foram obtidos a partir do sistema fotovoltaico da Figura 4.12.

Originalmente a série temporal de irradiância fora capturada segundo a segundo, pelo sistema fotovoltaico descrito na seção 4.2.

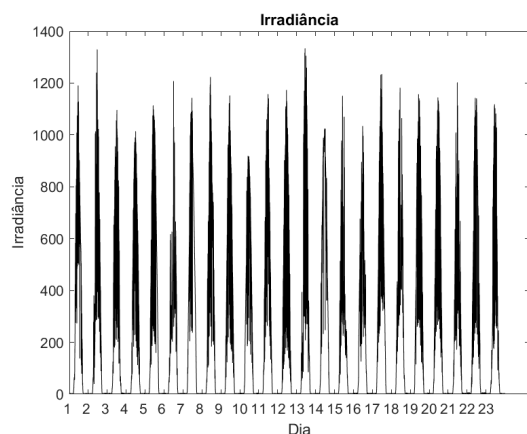
Porém, para efeito de cálculos e dadas as limitações de poder computacional, julgou-se conveniente a utilização de dados com resolução de 1 minuto, sendo este a média aritmética dos 60 segundos passados.

Os cálculos apresentados na seção a seguir foram realizados no MATLAB.

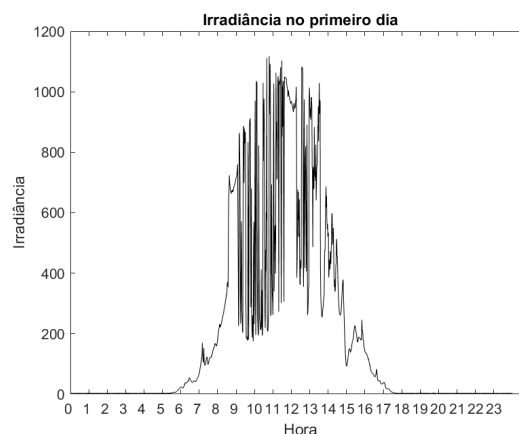
### 5.1 Método Box-Jenkins

#### 5.1.1 Identificação

Visualizamos a seguir, na Figura 5.1a um *plot* de uma série temporal de 23 dias da radiação solar minuto a minuto.



(a) Série Temporal da Irradiância em 23 dias



(b) Série Temporal da Irradiância, dia 1

Figura 5.1: Séries temporais de irradiância



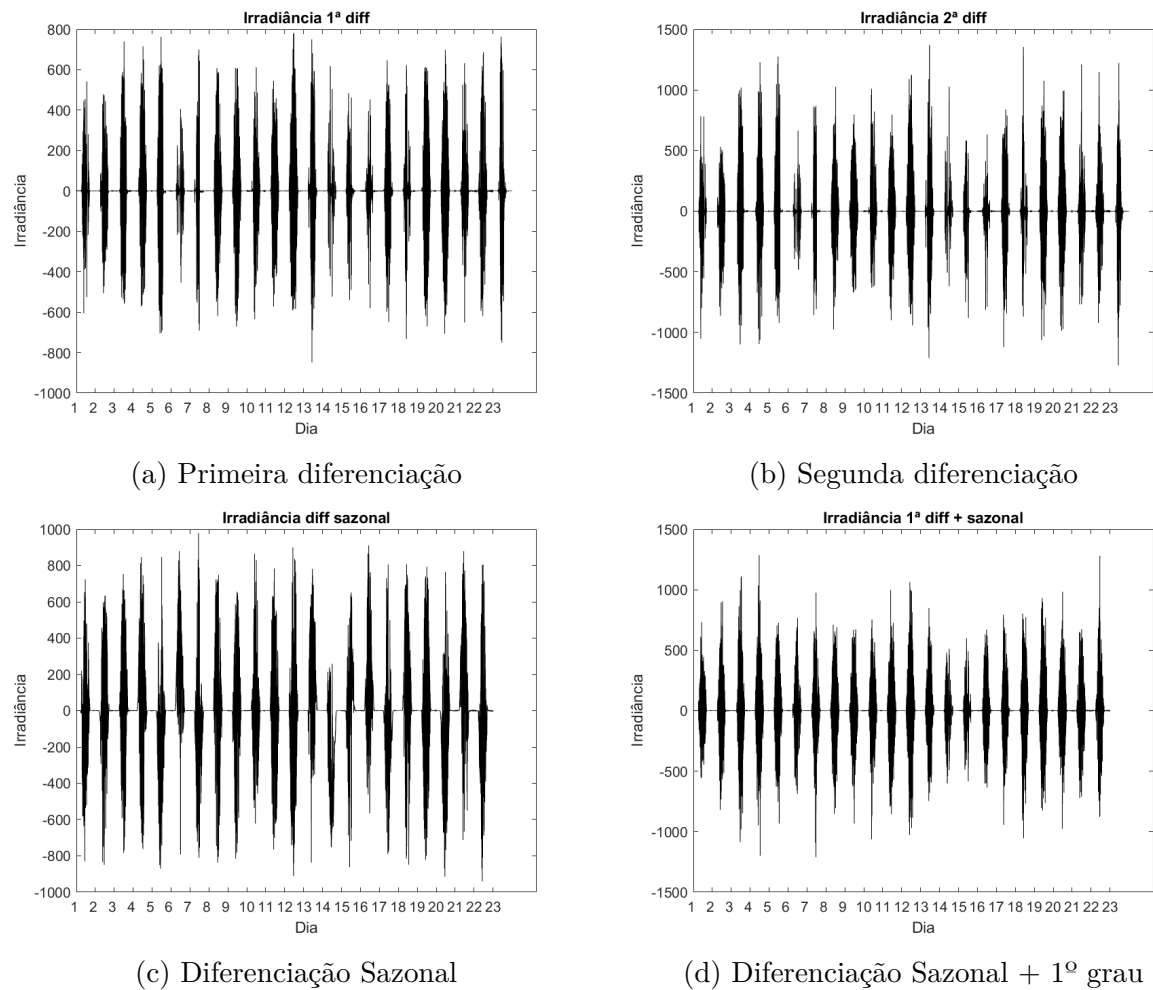


Figura 5.2: Diferenciações a partir da série original

### Inspeção Visual

A partir de uma inspeção visual, é possível perceber que, como esperado, a irradiância solar apresenta um padrão de sazonalidade que ocorre todos os dias.

A irradiância cresce a partir de aproximadamente 5:30h, tem sua máxima por volta das 12h e próximo às 18h se aproxima novamente da irradiância nula.

Esse comportamento pode ser observado com mais clareza a partir da Figura 5.1b

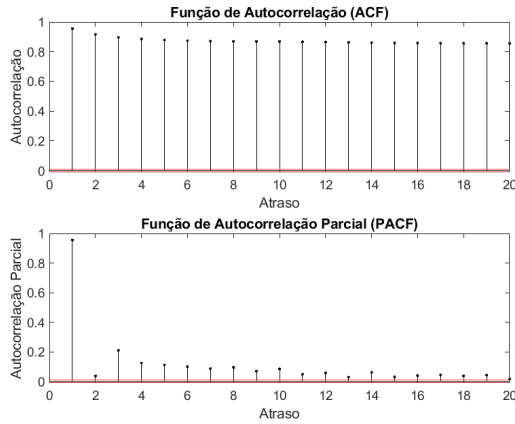
Pode-se notar, outrossim, que retirada a sazonalidade a partir de uma ou mais diferenciações sazonais, a série não apresentaria tendência nem média  $\mu$  oscilante, logo obteríamos após a diferenciação uma série estacionária. Esta condição será verificada nas etapas seguintes.

### Análise de Correlação

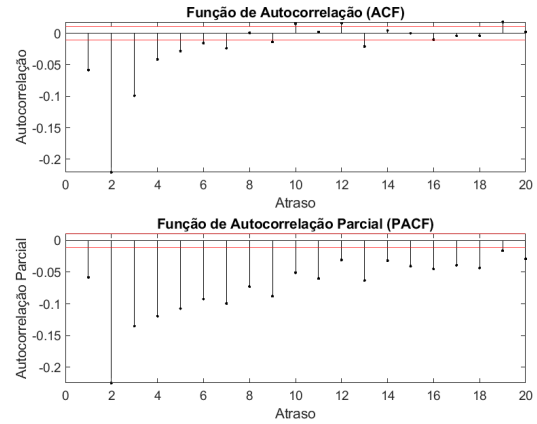
Nesta seção, após a análise visual, recorreremos às funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) a fim de se identificar padrões nos dados. A partir disso é possível deduzir quantos termos utilizar no treino do modelo. A ideia é identificar a

presença de componentes AR e MA.

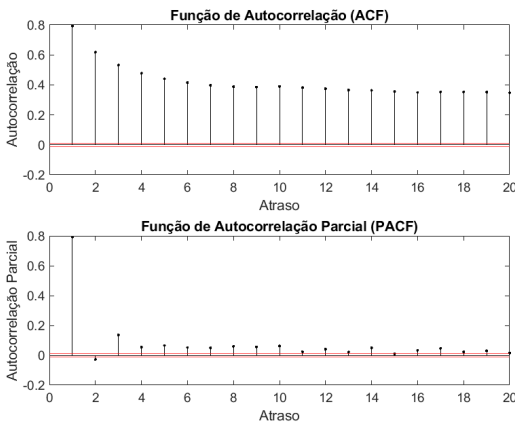
Na Figura 5.3, temos os gráficos de ACF e PACF para as séries original e obtidas através de algumas diferenciações.



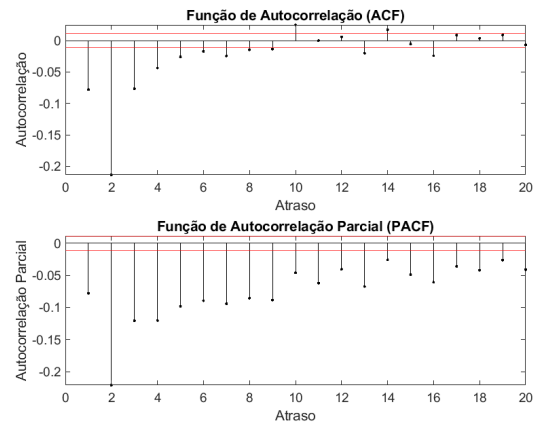
(a) ACF e PACF da serie de irradiância



(b) ACF e PACF da serie de irradiância D1



(c) ACF e PACF da serie de irradiância DS



(d) ACF e PACF da serie de irradiância DS

Figura 5.3: Análise dos gráficos de ACF e PACF

Observando a Figura 5.3a, é possível inferir que esta série não é estacionária, pois o gráfico do ACF apresenta decaimento gradual lento.

Dado isso, é necessário aplicar uma diferenciação nos dados. Após a primeira diferenciação, os gráficos de ACF e PACF se comportam conforme a Figura 5.3b. E a série temporal produto desta operação pode ser visualizada na Figura 5.2a.

Após uma primeira análise, é possível notar que, com o auxílio da Tabela 3.1, esta série temporal poderia ser modelada apresentando termos MA e possivelmente AR, o que será verificado na próxima etapa.

Uma vez que foi identificado comportamento sazonal da série, pode-se aplicar uma diferenciação sazonal a fim de se obter uma série temporal estacionária.

Sabendo-se que a resolução dos dados é de 1 minuto, a diferenciação será aplicada ao atraso correspondente ao atraso 1440, que pode ser escrito, em forma diferencial e com auxílio do operador de atraso como  $(1 - B^{1440})$ .

	Dickey-Fuller	Phillips-Perron	KPPS	Leybourne-McCabe
Série Original	1	1	1	1
1ª Diferenciação	1	1	0	0
2ª Diferenciação	1	1	0	0
Diferenciação Sazonal	1	1	1	1
1ª Diferenciação + Sazonal	1	1	0	0

Tabela 5.1: Testes estatísticos. Fonte: o autor

A série temporal resultante da diferenciação sazonal pode ser vista na Figura 5.2c, e os gráficos das funções de ACF e PACF correspondentes podem ser visualizados na Figura 5.3c.

O resultado de uma diferenciação seguida da diferenciação sazonal é apresentado na Figura 5.2d, e os gráficos de ACF e PACF são vistos na Figura 5.3d.

Analisando os gráficos de ACF e PACF, nota-se uma semelhança com os obtidos após a primeira diferenciação, e até então pode-se inferir apenas a provável presença de termos MA e talvez de AR para a construção do modelo. Estas conjecturas serão analisadas na próxima seção.

Pode-se, ainda, realizar uma segunda diferenciação, como observado na Figura 5.2b, que, por sua vez, visivelmente não traz melhorias.

A partir da inspeção visual nos gráficos, portanto, serão estimados na próxima seção modelos com  $d = 1$  e  $D = 1$ , com o vetor de parâmetros sazonais e não sazonais variando até 5, tirando vantagem do método de *overfitting* explicado anteriormente.

## Testes estatísticos

A fim de confirmar a possível estacionariedade das séries temporais, aplicaram-se testes estatísticos que indicam a presença (ou ausência) de uma raiz unitária.

Na Tabela 5.1 estão compilados os resultados dos testes.

A partir da Tabela 5.1 é possível observar que os testes de Dickey-Fuller e Phillips-Perron não são conclusivos, uma vez que não são capazes de distinguir a diferença entre as séries original e com diferenciação, pois obteve-se  $h = 1$  para todas as séries testadas, conseqüentemente não é possível afirmar que nenhuma das séries é estacionária a partir destes resultados.

Os testes de KPPS e Leybourne-McCabe, por sua vez, indicam que mesmo após uma diferenciação sazonal, a série temporal permanece estacionária, pois  $h = 1$  em ambos os resultados. Porém, aplicando-se uma primeira diferenciação seguida da diferenciação sazonal, obtém-se uma série estacionária, pois os resultados de ambos os testes apresentam  $h = 0$ , ou seja, falhou em rejeitar a hipótese nula de estacionariedade da série, portanto, a série é estacionária.

### 5.1.2 Estimação

A partir da identificação dos modelos que podem ser úteis para a série temporal estudada, os parâmetros foram estimados a partir do Método da Máxima Verossimilhança.

Como supracitado na seção anterior 5.1.1, a partir da inspeção visual e de testes estatísticos, foram construídos modelos cujos atrasos variam de 1 a 5, e cujos parâmetros sazonais também são acrescentados até o grau 3.

Utilizando o método iterativo de Box-Jenkins, pôde-se eliminar os modelos de menor qualidade a cada grau de sazonalidade, a partir dos parâmetros de avaliação já descritos.

Para os grau de sazonalidade 1, foram estimados os modelos apresentados na Tabela 5.2.

A partir daí, notou-se a pouca relevância do modelo cujos parâmetros não-sazonais são iguais a 1. Desta forma, para a próxima iteração essa configuração foi descartada.

Os modelos construídos para sazonalidade 2 pode ser observado na Tabela 5.3 e os modelos com sazonalidade 3 na Tabela 5.4.

### 5.1.3 Diagnóstico

Em posse dos modelos, pode-se realizar a checagem de diagnóstico, a fim de se identificar a qualidade do modelo, ou melhorar seu ajuste.

No Apêndice A, são apresentados todos os gráficos utilizados para análise de diagnóstico: resíduos normalizados, o ACF dos resíduos, o gráfico QQ e ainda os valores de  $p$  resultantes do teste de Ljung-Box para 20 atrasos.

Para a análise dos resíduos, o que se espera é que eles possuam um comportamento aleatório, semelhante à um ruído branco. O teste de Ljung-Box, como a análise do ACF dos resíduos, são utilizado para testar a independência dos resíduos estudados.

Para o modelo SARIMA(1,1,1)(1,1,1), a partir da Figura A.1, observando o gráfico de ACF vemos que nos atrasos 1 e 6 já existem informações da série temporal que não são cobertas, uma vez que a correlação é significativa nestes atrasos. O que pode ser confirmado no gráfico que apresenta os valores de  $p$  do teste Ljung-box. O que já é um mal indicativo para a qualidade do modelo em análise.

No modelo seguinte, SARIMA(2,1,2)(1,1,1), notamos que o gráfico de PACF da Figura A.2 apresenta correlação significativa no sexto e também no oitavo atraso, visto que o valor de ACF está acima dos limites de confiabilidade nestes atraso. Os modelos SARIMA(3,1,3)(1,1,1) e SARIMA(4,1,4)(1,1,1) apresentam um comportamento semelhante.

E, finalmente, no modelo SARIMA(5,1,5)(1,1,1), observamos, a partir da Figura A.5, o modelo perde qualidade adicionando estes termos AR e MA em relação aos modelos anteriores, uma vez que nos atrasos 1, 2 e 4 existem correlações significativas.

Visto que ao adicionar termos AR e MA não-sazonais ao modelo SARIMA(4,1,4)(1,1,1) não foi eficiente, analisar-se-ão os modelos com mais parâmetros AR e MA sazonais.

Sazonalidade: 1				
Modelo	Parâmetros	p-value	AIC	BIC
<i>SARIMA</i> (1, 1, 1)(1, 1, 1) <sub>1440</sub>	$\phi_1 = 0.679$	0	392697	392747
	$\theta_1 = -0.973$	0		
	$\Phi_{1440} = -0.101$	5.9594e-258		
	$\Theta_{1440} = -0.837$	0		
<i>SARIMA</i> (2, 1, 2)(1, 1, 1) <sub>1440</sub>	$\phi_1 = 0.411$	7.984e-133	392138	392205
	$\phi_2 = 0.65$	1.3282e-15		
	$\theta_1 = -0.620$	4.1657e-312		
	$\theta_2 = -0.328$	5.5046e-93		
	$\Phi_{1440} = -0.107$	5.8653e-290		
	$\Theta_{1440} = -0.836$	0		
<i>SARIMA</i> (3, 1, 3)(1, 1, 1) <sub>1440</sub>	$\phi_1 = 1.409$	0	392037	392121
	$\phi_2 = -0.390$	1.7561e-45		
	$\phi_3 = -0.036$	0.0022573		
	$\theta_1 = -1.622$	0		
	$\theta_2 = 0.366$	2.7094e-31		
	$\theta_3 = 0.256$	4.0362e-58		
	$\Phi_{1440} = -0.107$	7.5067e-287		
	$\Theta_{1440} = -0.836$	0		
<i>SARIMA</i> (4, 1, 4)(1, 1, 1) <sub>1440</sub>	$\phi_1 = 0.349$	1.3169e-99	392015	392116
	$\phi_2 = 1.016$	0		
	$\phi_3 = -0.263$	8.5289e-53		
	$\phi_4 = -0.130$	5.2235e-31		
	$\theta_1 = -0.565$	9.915e-268		
	$\theta_2 = -1.267$	0		
	$\theta_3 = 0.453$	1.1248e-146		
	$\theta_4 = 0.379$	1.4016e-149		
	$\Phi_{1440} = -0.106$	1.51e-281		
	$\Theta_{1440} = -0.837$	0		
<i>SARIMA</i> (5, 1, 5)(1, 1, 1) <sub>1440</sub>	$\phi_1 = 0.468$	7.0568e-19	392873	392990
	$\phi_2 = 1.109$	0		
	$\phi_3 = -0.412$	1.1285e-12		
	$\phi_4 = -0.415$	1.727e-118		
	$\phi_5 = 0.226$	9.3928e-61		
	$\theta_1 = -0.701$	3.8616e-40		
	$\theta_2 = -1.338$	0		
	$\theta_3 = 0.694$	9.6959e-21		
	$\theta_4 = 0.647$	2.2109e-162		
	$\theta_5 = -0.302$	2.8572e-26		
	$\Phi_{2880} = -0.166$	0		
	$\Theta_{1440} = -0.828$	0		

Tabela 5.2: Modelos estimados com sazonalidade 1. Fonte: o autor

Sazonalidade: 2				
Modelo	Parâmetros	p-value	AIC	BIC
<i>SARIMA</i> (2, 1, 2)(2, 1, 2) <sub>1440</sub>	$\phi_1 = 0.386$	3.4181e-136	390674	390758
	$\phi_2 = 0.112$	3.4181e-136		
	$\theta_1 = -0.595$	0		
	$\theta_2 = -0.356$	1.0381e-126		
	$\Phi_{1440} = -0.861$	0		
	$\Phi_{2880} = 0.05$	1.0998e-51		
	$\Theta_{1440} = -0.836$	4.6217e-31		
	$\Theta_{2880} = -0.05$	0		
<i>SARIMA</i> (3, 1, 3)(2, 1, 2) <sub>1440</sub>	$\phi_1 = -0.612$	5.0538e-262	390654	390755
	$\phi_2 = 0.459$	0		
	$\phi_3 = 0.168$	9.1521e-56		
	$\theta_1 = 0.403$	3.3312e-119		
	$\theta_2 = -0.881$	0		
	$\theta_3 = -0.425$	2.4578e-209		
	$\Phi_{1440} = -0.861$	0		
	$\Phi_{2880} = -0.05$	7.2216e-53		
<i>SARIMA</i> (4, 1, 4)(2, 1, 2) <sub>1440</sub>	$\phi_1 = 0.326$	1.1603e-36	390605	390722
	$\phi_2 = 0.889$	3.964e-258		
	$\phi_3 = -0.286$	2.0003e-51		
	$\phi_4 = -0.056$	1.6043e-05		
	$\theta_1 = -0.547$	8.5476e-97		
	$\theta_2 = -1.153$	0		
	$\theta_3 = 0.453$	1.2303e-76		
	$\theta_4 = 0.269$	3.7979e-48		
	$\Phi_{1440} = -0.861$	0		
	$\Phi_{2880} = -0.05$	7.2261e-51		
<i>SARIMA</i> (5, 1, 5)(2, 1, 2) <sub>1440</sub>	$\phi_1 = 0.812$	0.00072864	390622	390757
	$\phi_2 = 0.702$	5.5863e-09		
	$\phi_3 = -0.759$	0.00070049		
	$\phi_4 = 0.274$	0.029397		
	$\phi_5 = -0.06$	6.0683e-05		
	$\theta_1 = -1.033$	1.7166e-05		
	$\theta_2 = -839$	1.2033e-06		
	$\theta_3 = 1.000$	0.00026973		
	$\theta_4 = -0.154$	0.37225		
	$\theta_5 = 0.028$	0.47594		
	$\Phi_{1440} = -0.860$	0		
	$\Phi_{2880} = -0.048$	2.5235e-47		
	$\Theta_{1440} = -0.060$	1.5173e-29		
$\Theta_{2880} = -0.711$	0			

Tabela 5.3: Modelos estimados com sazonalidade 2. Fonte: o autor

Sazonalidade: 3				
Modelo	Parâmetros	p-value	AIC	BIC
<i>SARIMA</i> (3, 1, 3)(3, 1, 3) <sub>1440</sub>	$\phi_1 = 1.3815$	1.1515e-65	389381	389499
	$\phi_2 = -0.341$	7.6244e-160		
	$\phi_3 = -0.057$	3.0453e-21		
	$\theta_1 = -1.585$	3.6453e-14		
	$\theta_2 = 0.290$	0		
	$\theta_3 = 0.296$	3.5856e-127		
	$\Phi_{1440} = -1$	3.2453e-45		
	$\Phi_{2880} = -0.800$	2.4587e-120		
	$\Phi_{4320} = -0.045$	0		
	$\Theta_{1440} = 0.092$	4.1452e-23		
	$\Theta_{2880} = -0.086$	3.4312e-76		
	$\Theta_{4320} = -0.662$	0		
<i>SARIMA</i> (4, 1, 4)(3, 1, 3) <sub>1440</sub>	$\phi_1 = 0.434$	0	390175	390309
	$\phi_2 = 0.872$	7.1248e-23		
	$\phi_3 = -0.314$	0.00035949		
	$\phi_4 = -0.059$	0.064826		
	$\theta_1 = -0.639$	2.0448e-05		
	$\theta_2 = -1.112$	5.4516e-12		
	$\theta_3 = 0.477$	1.5143e-07		
	$\theta_4 = 0.278$	0.0002753		
	$\Phi_{1440} = -0.924$	0.85425		
	$\Phi_{2880} = -207$	0.68594		
	$\Phi_{4320} = -0.124$	0		
	$\Theta_{1440} = -0.003$	5.4515e-89		
$\Theta_{2880} = -0.697$	1.4853e-56			
$\Theta_{4320} = -0.043$	0			

Tabela 5.4: Modelos estimados com sazonalidade 3. Fonte: o autor

Iniciando pelo SARIMA(2,1,2)(2,1,2), observando a Figura A.6, o gráfico do ACF dos resíduos mostra uma correlação significativa nos atrasos 2, 4 e 6. Este comportamento também pode ser observado no modelo SARIMA(3,1,3)(2,1,2), e de forma semelhante também nos modelos SARIMA(4,1,4)(2,1,2) e SARIMA(5,1,5)(2,1,2).

A partir desses resultados, manteve-se a constante de parâmetros não sazonais e aumentou-se apenas os sazonais. Os resultados podem ser observados nas Figuras do apêndice A..

Nos modelos SARIMA(3,1,3)(3,1,3), e SARIMA(4,1,4)(3,1,3) os valores de ACF estão dentro dos limites de confiabilidade. Desta forma, estes modelos, possivelmente, apresentarão melhores resultados.

Após a previsão, que será apresentada na seção seguinte, analisar-se-á se estas melhorias em diagnostico de resíduos refletem diretamente na qualidade da previsão.



# Capítulo 6

## Resultados e Discussões

Nesta seção serão analisados os resultados das previsões a partir dos modelos estimados e modificados nas seções anteriores.

Foram utilizadas como métricas para avaliação das previsões apresentadas o MSE e o MAPE.

### 6.1 Previsão

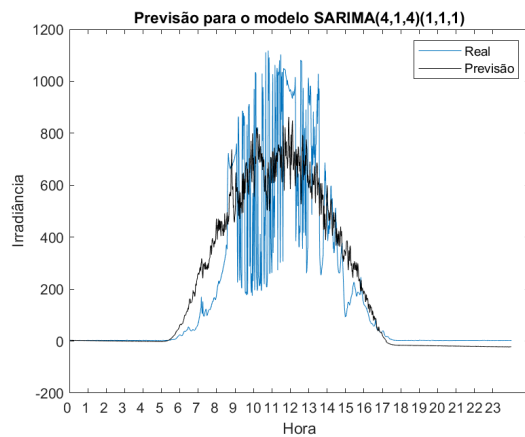
Tendo em vista que a análise dos resíduos realizada na etapa anterior demonstrou que os modelos selecionados são adequados para representar a série temporal estudada, nesta etapa é realizada a previsão para o período de um dia seguinte à série em questão.

Valores de MAPE e MSE		
Modelo	MAPE	MSE
<i>SARIMA</i> (1, 1, 1)(1, 1, 1) <sub>1440</sub>	0.270388811845513	8.364535312564251e+07
<i>SARIMA</i> (2, 1, 2)(1, 1, 1) <sub>1440</sub>	0.311088660382182	1.137727859086927e+08
<i>SARIMA</i> (3, 1, 3)(1, 1, 1) <sub>1440</sub>	1.406681891823677	4.692420151651514e+07
<i>SARIMA</i> (4, 1, 4)(1, 1, 1) <sub>1440</sub>	2.529048716993082	4.270970300203889e+07
<i>SARIMA</i> (5, 1, 5)(1, 1, 1) <sub>1440</sub>	5.522707966104619	4.492382070542155e+07
<i>SARIMA</i> (2, 1, 2)(2, 1, 2) <sub>1440</sub>	0.298566625403429	9.811575638213997e+07
<i>SARIMA</i> (3, 1, 3)(2, 1, 2) <sub>1440</sub>	0.312116962107178	9.583388334946437e+07
<i>SARIMA</i> (4, 1, 4)(2, 1, 2) <sub>1440</sub>	8.101238029873288	7.131312027740571e+07
<i>SARIMA</i> (5, 1, 5)(2, 1, 2) <sub>1440</sub>	0.913096720908498	4.959064551993600e+07
<i>SARIMA</i> (3, 1, 3)(3, 1, 3) <sub>1440</sub>	2.798600238482275	4.619016557519928e+07
<i>SARIMA</i> (4, 1, 4)(3, 1, 3) <sub>1440</sub>	2.771101232923249	5.558621387396077e+07

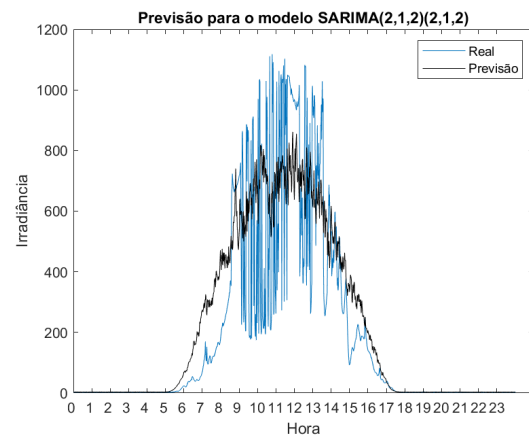
Tabela 6.1: Valores de MAPE e MSE calculados a partir dos modelos estimados

Na Figura 6.1 podem ser observadas as previsões a partir dos modelos selecionados quando comparadas com os dados reais do período.

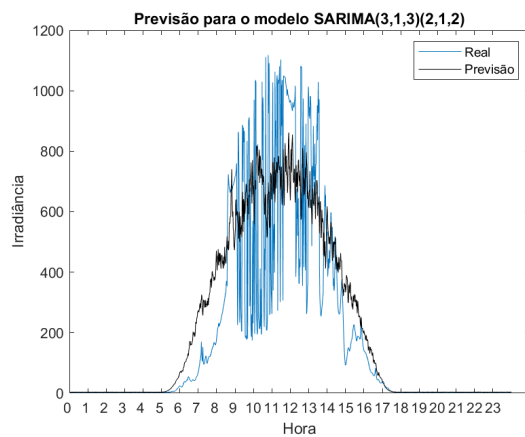
Como se pode notar, os modelos com maior grau de sazonalidade possuem melhoras significativas até  $p = q = 2$ , a partir daí não são observados avanços consideráveis para o



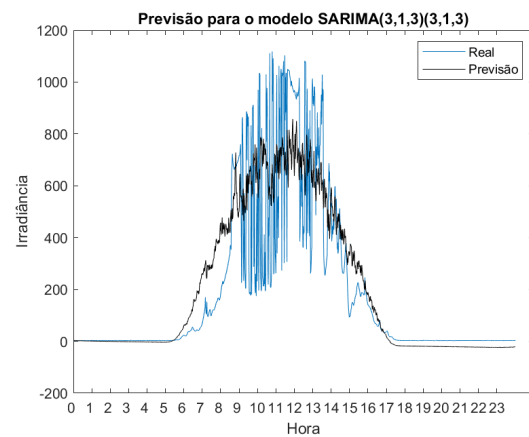
(a) Previsão do dia seguinte a partir do modelo SARIMA(4,1,4)(1,1,1)



(b) Previsão do dia seguinte a partir do modelo SARIMA(2,1,2)(2,1,2)



(c) Previsão do dia seguinte a partir do modelo SARIMA(3,1,3)(2,1,2)



(d) Previsão do dia seguinte a partir do modelo SARIMA(3,1,3)(3,1,3)

Figura 6.1: Previsões realizadas a partir dos modelos SARIMA que apresentaram melhores índices de qualidade

acréscimo de complexidade no modelo.

Os modelos com termos sazonais de grau 1 apresentam uma certa tendência, que pode ser observada, por exemplo, na Figura 6.1a, onde a partir de aproximadamente 17h a previsão apresenta valores acima de zero, o que não é um comportamento esperado.

Este problema é superado ao serem adicionados os parâmetros sazonais com grau 2.

Ao adicionar mais um grau de sazonalidade, a previsão volta a apresentar uma tendência, como pode ser visto na Figura 6.1d.

Adicionalmente, a partir da Tabela 6.1, os menores valores de MAPE são dos modelos  $SARIMA(1, 1, 1)(1, 1, 1)_{1440}$ ,  $SARIMA(2, 1, 2)(2, 1, 2)_{1440}$  e  $SARIMA(3, 1, 3)(2, 1, 2)_{1440}$ . E os modelos com menores valores de MSE são:  $SARIMA(4, 1, 4)(1, 1, 1)_{1440}$ ,  $SARIMA(3, 1, 3)(3, 1, 3)_{1440}$  e  $SARIMA(5, 1, 5)(1, 1, 1)_{1440}$ .

Porém, uma vez que apenas os modelos com grau de sazonalidade 2 não apresentaram

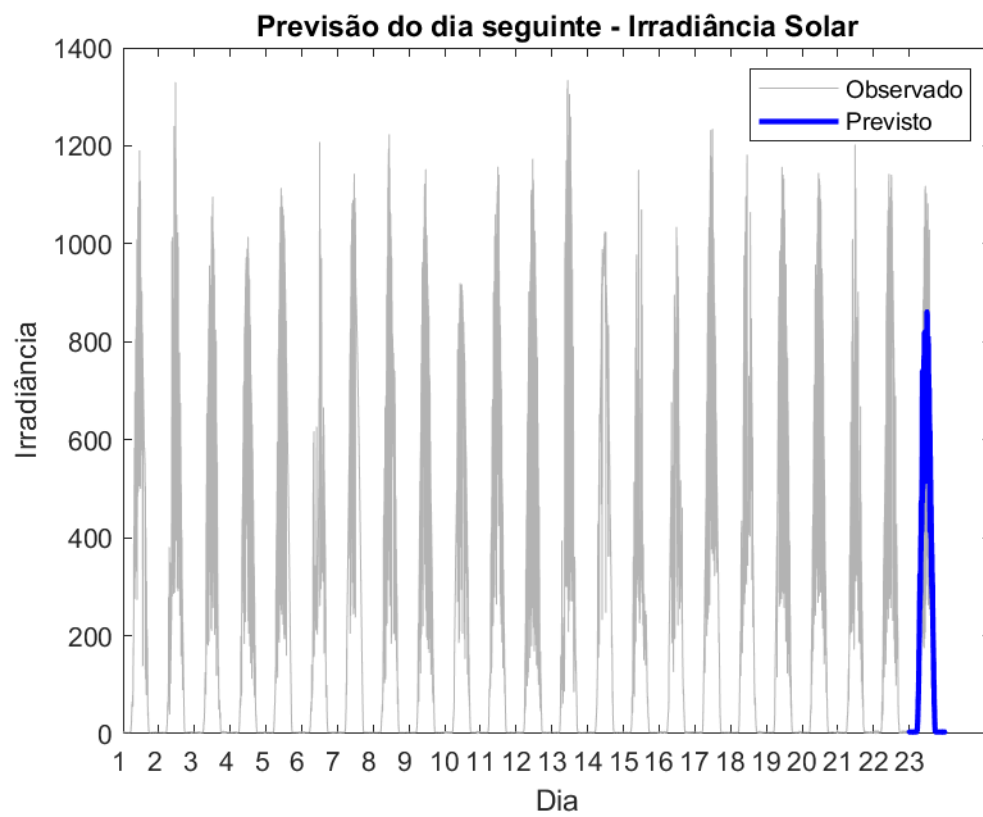


Figura 6.2: Previsão do dia seguinte à série temporal utilizada

tendência em sua previsão, estes serão considerados mais adequados.

Na Figura 6.2 podemos observar a série original de irradiância solar seguida da previsão a partir do modelo  $SARIMA(3, 1, 3)(2, 1, 2)_{1440}$ .

# Capítulo 7

## Conclusão

O presente trabalho teve como objetivo desenvolver e comparar modelos adequados para previsão de irradiância solar para o dia seguinte. Para esta finalidade, foram utilizados dados da estação meteorológica instalada no Instituto de Computação (IC) na Universidade Federal de Alagoas (Ufal).

A fim de se otimizar o processo de modelagem, os dados que foram coletados a cada segundo, neste trabalho foi utilizada sua média por minuto. Utilizando os dados de irradiância, aplicou-se o método de identificação Box-Jenkins para a análise da série temporal baseada em suas observações passadas.

A partir da análise dos dados percebeu-se que a série temporal utilizada possui um padrão que se repete diariamente, como esperado. Foram considerados apenas modelos ARIMA sazonais (SARIMA).

Após a identificação do modelo, foi possível definir classes do modelo ARIMA que poderiam se adequar à série temporal de irradiância. Estes modelos, por sua vez, foram estimados a partir do método da Máxima Verossimilhança, onde uma função de verossimilhança, baseada na função de densidade de probabilidade do modelo, é maximizada. Todos os cálculos foram realizados no MATLAB.

Em seguida, os modelos passaram por uma checagem de diagnóstico, a partir do estudo dos resíduos, que possibilitou realizar mudanças e construir novos modelos is adequados.

Tendo, portanto, posse dos modelos que teoricamente seriam os mais adequados para a série temporal estudada, foi possível realizar a previsão do dia seguinte da radiação solar através do método de MSE aplicado diretamente à equação diferença do modelo.

Esta previsão foi então comparada com os dados reais obtidos. Foi possível observar que a previsão ficou próxima dos dados reais, que pôde ser confirmado através de métricas como MSE e MAPE.

Pode-se concluir, portanto, que os objetivos do trabalho foram atingidos, uma vez que os resultados obtidos foram aceitáveis dentro da métrica e dos limites propostos, e visto ainda que a previsão de irradiância é de muita importância para a produção de energia solar.

Neste trabalho foi possível descrever e implementar técnicas de predição nas séries temporais com dados coletados, possibilitando um estudo de caso real. Como sugestão de continuação do estudo, a utilização da irradiância como entrada exógena em um modelo ARIMAX para previsão da potência do sistema poderia apresentar resultados significantes.

# Bibliografia

- [ABNT, 2006] ABNT (2006). *NBR 10899: Energia Solar Fotovoltaica - Terminologia*. Associação Brasileira de Normas Técnicas.
- [ABSOLAR, 2020] ABSOLAR (2020). Infográfico absolar. Acessado em 15 de julho de 2020.
- [Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- [Ashok Kumar et al., 2015] Ashok Kumar, L., Sumathi, S., and Surekha, P. (2015). Solar pv and wind energy conversion systems: An introduction to theory, modeling with matlab/simulink, and the role of soft computing techniques.
- [Bastos, 2020] Bastos, I. G. P. (2020). Previsão de geração fotovoltaica a partir de dados meteorológicos utilizando rede lstm.
- [Bayer and Souza, 2010] Bayer, F. M. and Souza, A. M. (2010). Wavelets e modelos tradicionais de previsão: um estudo comparativo. *Revista Brasileira de Biometria*, 28:40–61.
- [Box et al., 1994] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time series analysis: forecasting and control*. Prentice Hall.
- [Brockwell and Davis, 2016] Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.
- [Cargnelutti Filho et al., 2010] Cargnelutti Filho, A., Matzenauer, R., Maluf, J. R. T., and Radin, B. (2010). Modelo harmônico para a estimativa da temperatura do ar média mensal em diferentes locais do estado do rio grande do sul. *Ciência Rural*, 40:1888 – 1896.
- [Chu et al., 2015] Chu, Y., Pedro, H. T., Li, M., and Coimbra, C. F. (2015). Real-time forecasting of solar irradiance ramps with smart image processing. *Solar Energy*, 114:91 – 104.

- [de Leva et al., 2004] de Leva, F. F., Salerno, C. H., Camacho, J. R., and Guimarães, S. C. (2004). Modelo de um projeto de um sistema fotovoltaico. *ENCONTRO ENERGIA MEIO RURAL*, 5.
- [Dickey and Fuller, 1979] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.
- [EPE, 2020] EPE, E. d. P. E. (2020). Matriz energética e elétrica. Acessado em 16 de julho de 2020.
- [Hammer et al., 1999] Hammer, A., Heinemann, D., Lorenz, E., and Lückehe, B. (1999). Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, 67(1):139 – 150.
- [Hannan and Quinn, 1979] Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195.
- [Ji and Chee, 2011] Ji, W. and Chee, K. C. (2011). Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn. *Solar Energy*, 85(5):808 – 817.
- [Kwiatkowski et al., 1992] Kwiatkowski, D., Phillips, P. C., Schmidt, P., Shin, Y., et al. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, 54(1-3):159–178.
- [Leybourne and HcCabe, 1999] Leybourne, S. J. and HcCabe, B. (1999). Modified stationarity tests with data-dependent model-selection rules. *Journal of Business & Economic Statistics*, 17(2):264–270.
- [Ljung and Box, 1978] Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.
- [Martins et al., 2008] Martins, F. R., Pereira, E. B., De Abreu, S. L., and Colle, S. (2008). Brazilian atlas for solar energy resource: Swera results. In *Proceedings of ISES World Congress 2007 (Vol. I–Vol. V)*, pages 2651–2655. Springer.
- [Mellit and Pavan, 2010] Mellit, A. and Pavan, A. M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy. *Solar Energy*, 84(5):807 – 821.
- [Mirzavand and Ghazavi, 2014] Mirzavand, M. and Ghazavi, R. (2014). A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. *Water Resources Management*, 29.

- [Murat et al., 2018] Murat, M., Malinowska, I., Gos, M., and Krzyszczak, J. (2018). Forecasting daily meteorological time series using arima and regression models. *International Agrophysics*, 32(2):253–264.
- [Pazikadin et al., 2020] Pazikadin, A. R., Rifai, D., Ali, K., Malik, M. Z., Abdalla, A. N., and Faraj, M. A. (2020). Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ann): A review of five years research trend. *Science of The Total Environment*, 715:136848.
- [Phillips and Perron, 1988] Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.
- [Reikard, 2009] Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342 – 349.
- [Reikard et al., 2017] Reikard, G., Haupt, S. E., and Jensen, T. (2017). Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models. *Renewable Energy*, 112:474 – 485.
- [RIBSOL, 2020] RIBSOL (2020). Sistemas fotovoltaicos. Acessado em 16 de julho de 2020.
- [Ritchie and Roser, 2020] Ritchie, H. and Roser, M. (2020). Energy. Acessado em 15 de julho de 2020.
- [Schwarz et al., 1978] Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- [Shams et al., 2016] Shams, M. B., Haji, S., Salman, A., Abdali, H., and Alsaffar, A. (2016). Time series analysis of bahrain’s first hybrid renewable energy system. *Energy*, 103:1 – 15.
- [Silva et al., 2008] Silva, M. I., Guimarães, E. C., and Tavares, M. (2008). Previsão da temperatura média mensal de uberlândia, mg, com modelos de séries temporais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 12(5):480–485.
- [Solargis, 2019] Solargis (2019). Global solar atlas 2.0, solar resource data. Acessado em 15 de julho de 2020.
- [Valipour, 2015] Valipour, M. (2015). Long-term runoff study using sarima and arima models in the united states. *Meteorological Applications*, 22(3):592–598.
- [Viana et al., 2012] Viana, T. d. S. et al. (2012). Potencial de geração de energia elétrica com sistemas fotovoltaicos com concentrador no brasil.



- [Voyant et al., 2017] Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569 – 582.
- [Wang et al., 2018] Wang, Y., Wang, C., Shi, C., and Xiao, B. (2018). Short-term cloud coverage prediction using the arima time series model. *Remote Sensing Letters*, 9(3):274–283.
- [Yang et al., 2012] Yang, D., Jirutitijaroen, P., and Walsh, W. M. (2012). Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy*, 86(12):3531 – 3543. Solar Resources.
- [Zaharim et al., 2009] Zaharim, A., Razali, A., Gim, T., and Sopian, K. (2009). Time series analysis of solar radiation data in the tropics. *European Journal of Scientific Research*, 25:672–678.
- [Zhang et al., 1998] Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, 14(1):35–62.
- [Zheng and Zhong, 2011] Zheng, F. and Zhong, S. (2011). Time series forecasting using a hybrid rbf neural network and ar model based on binomial smoothing. *World Academy of Science, Engineering and Technology*, 51:1464–1468.
- [Zilles et al., 2016] Zilles, R., Macêdo, W. N., Galhardo, M. A. B., and de Oliveira, S. H. F. (2016). *Sistemas fotovoltaicos conectados à rede elétrica*. Oficina de Textos.

# Apêndice A

## Resíduos

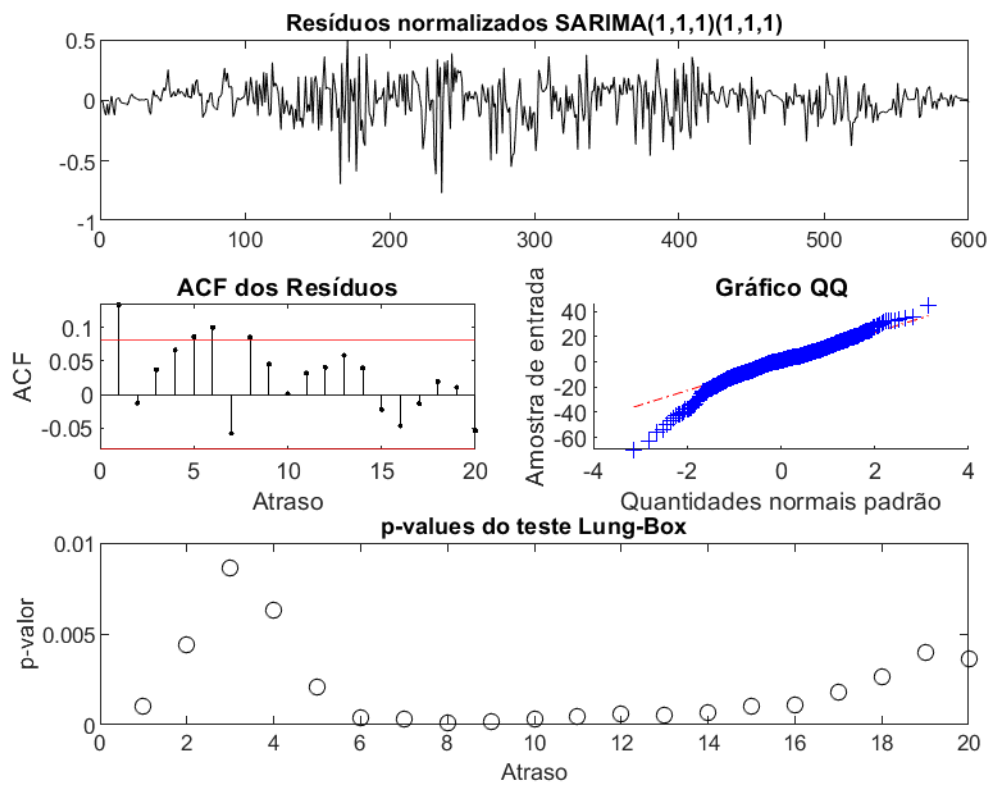


Figura A.1: Análise de Resíduos do modelo SARIMA(1,1,1)(1,1,1)

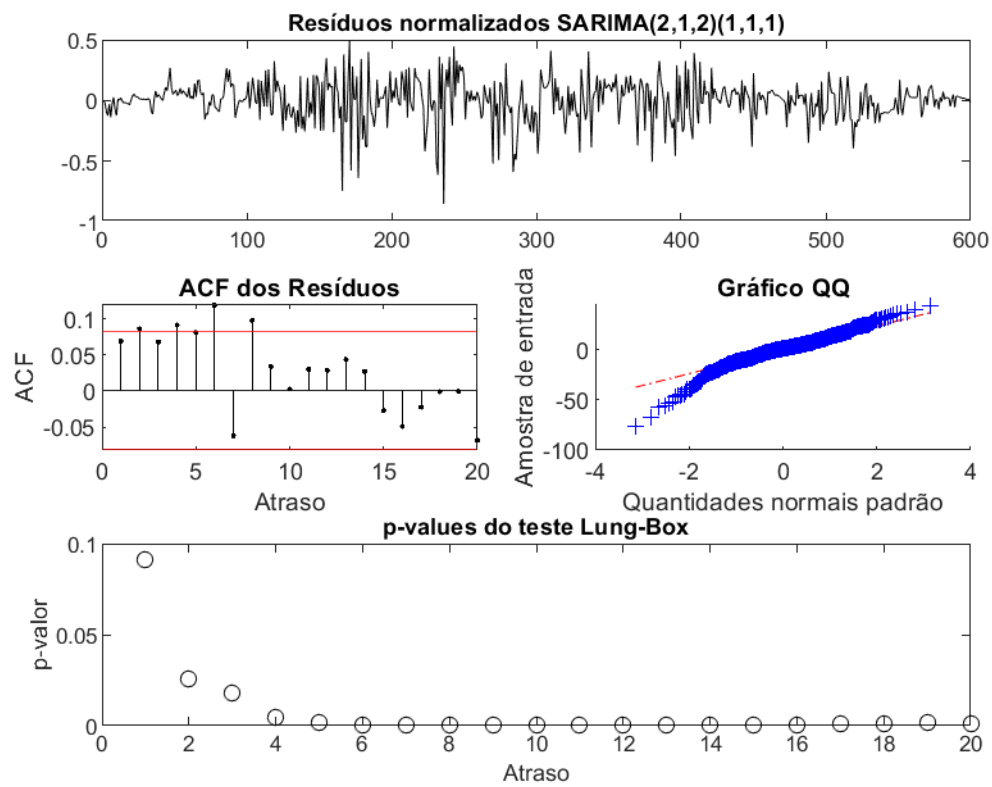
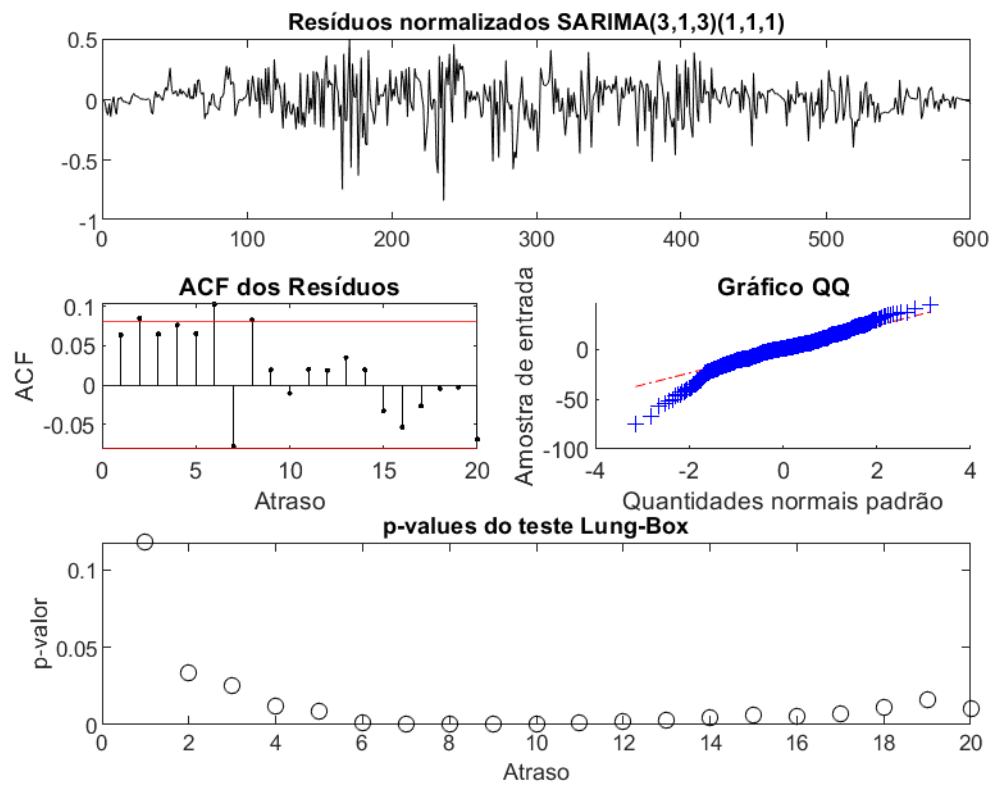
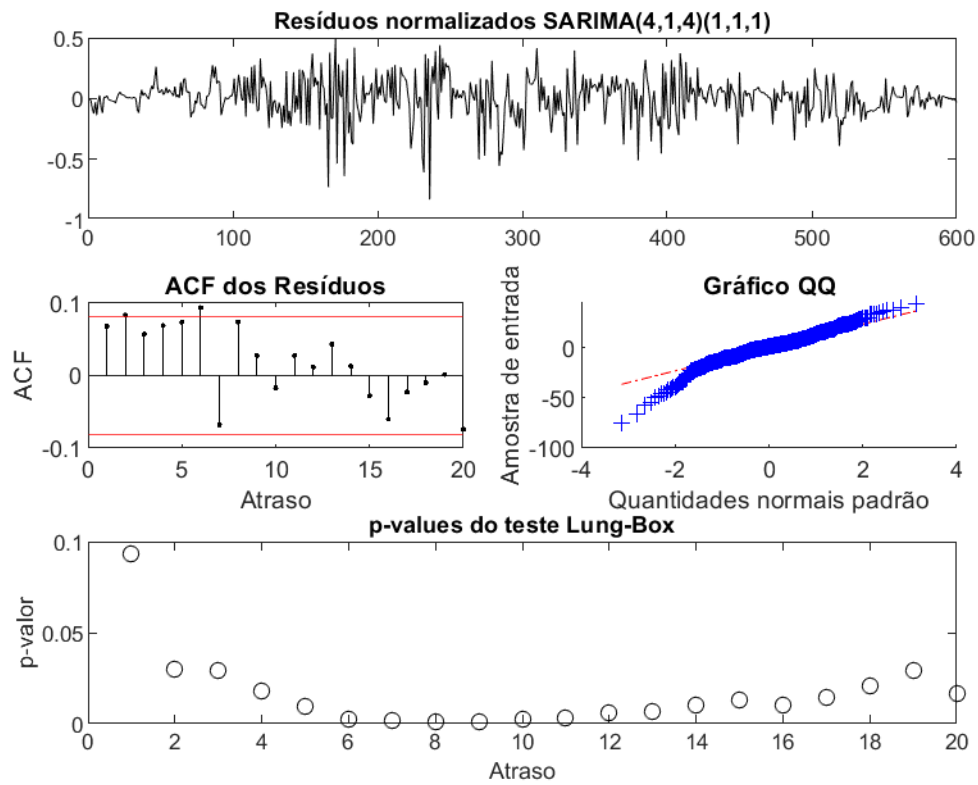


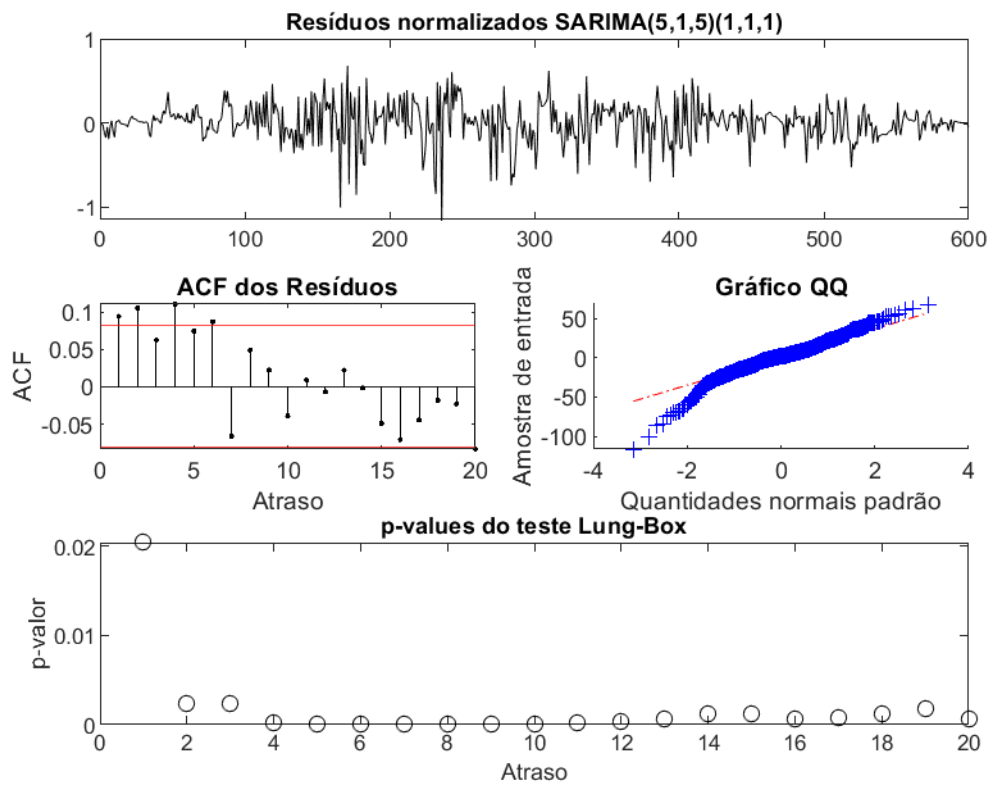
Figura A.2: Análise de Resíduos do modelo SARIMA(2,1,2)(1,1,1)



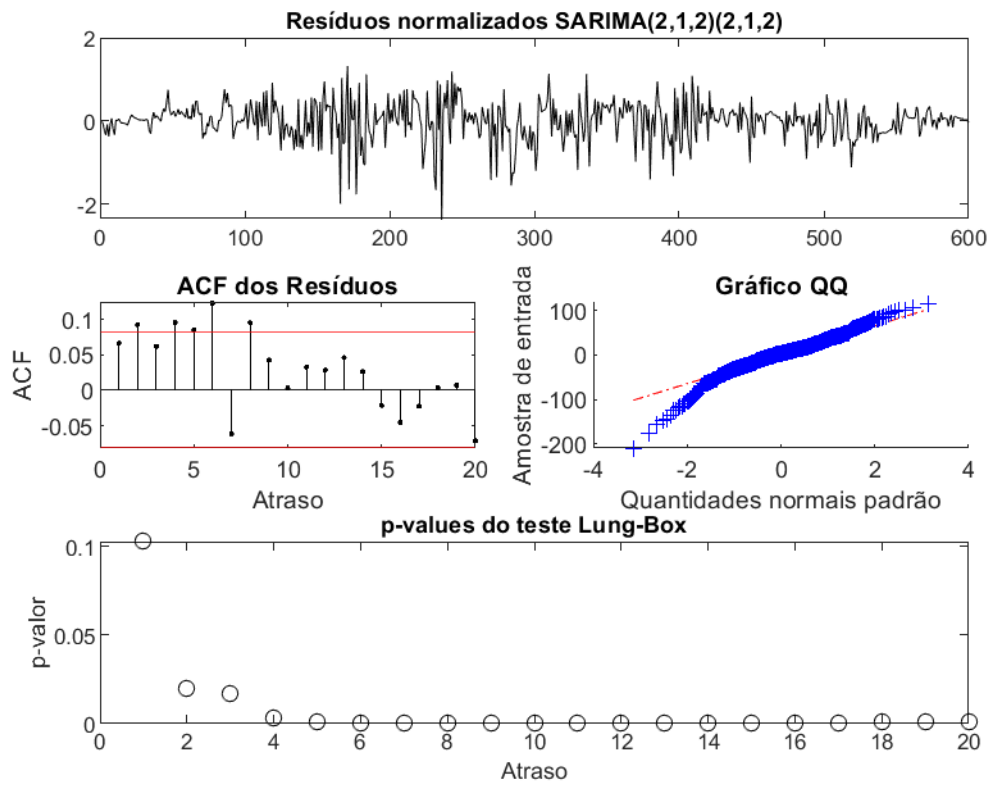
Análise de Resíduos do modelo SARIMA(3,1,3)(1,1,1)



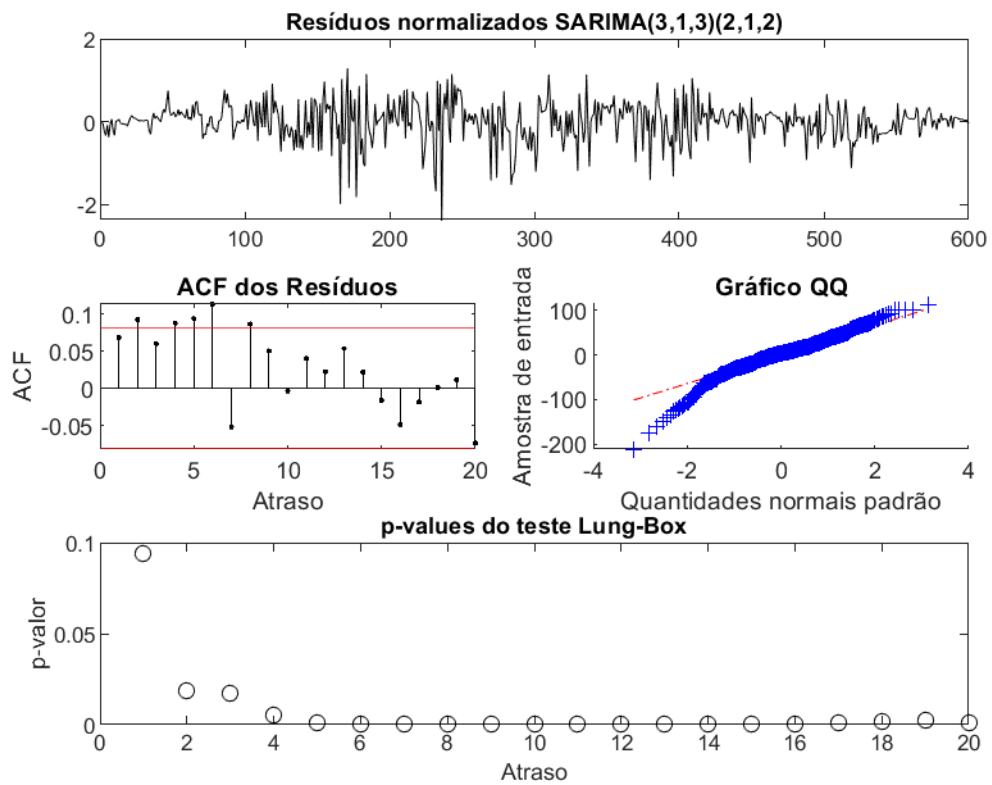
Análise de Resíduos do modelo SARIMA(4,1,4)(1,1,1)



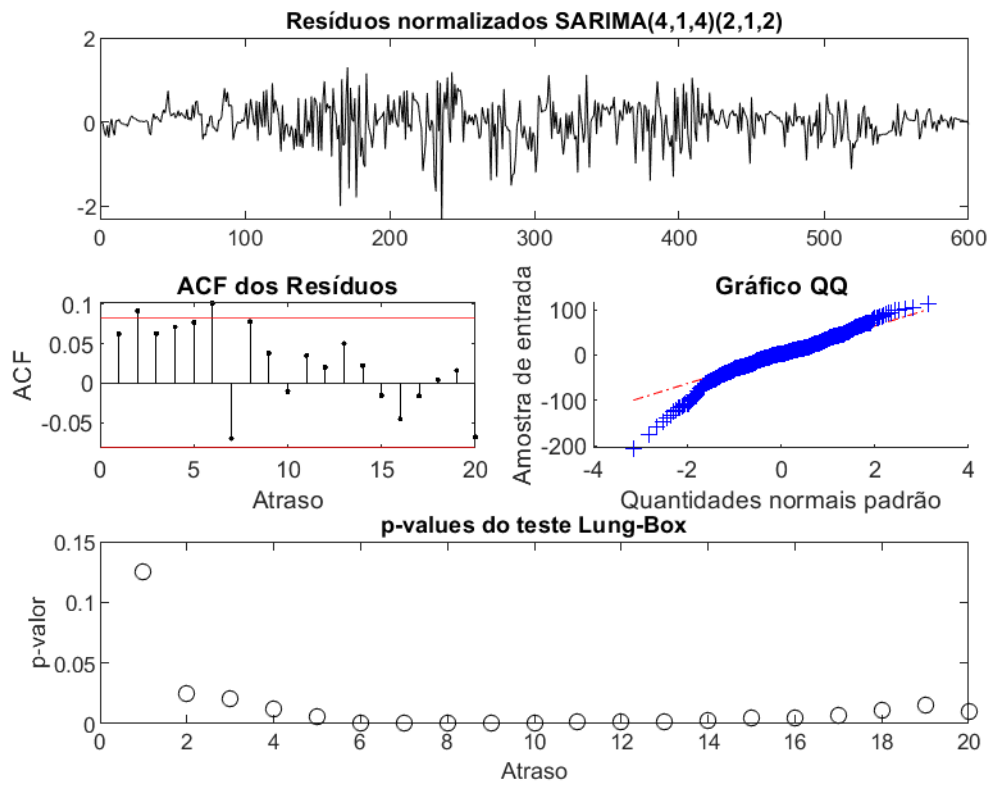
Análise de Resíduos do modelo SARIMA(5,1,5)(1,1,1)



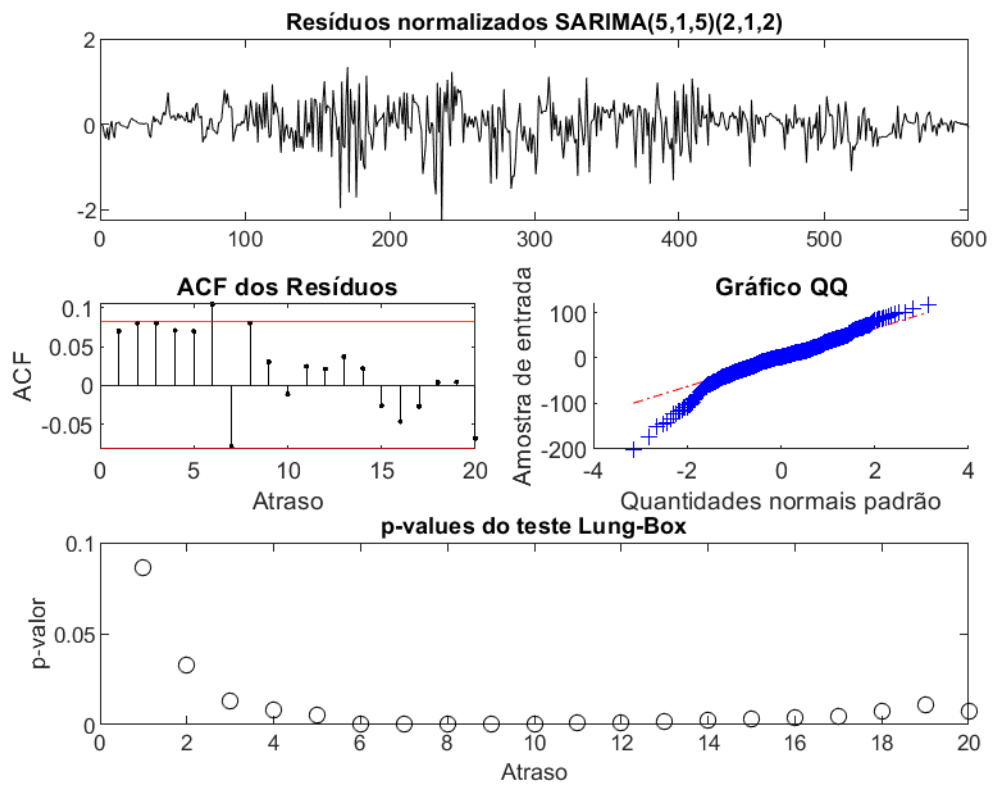
Análise de Resíduos do modelo SARIMA(2,1,2)(2,1,2)



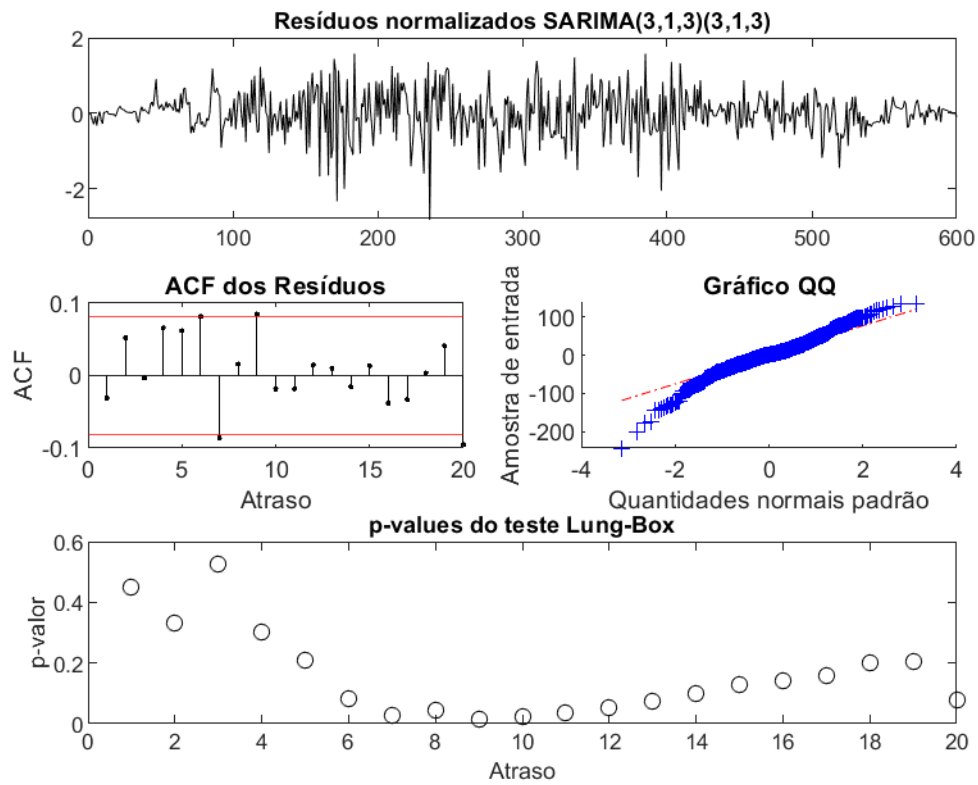
Análise de Resíduos do modelo SARIMA(3,1,3)(2,1,2)



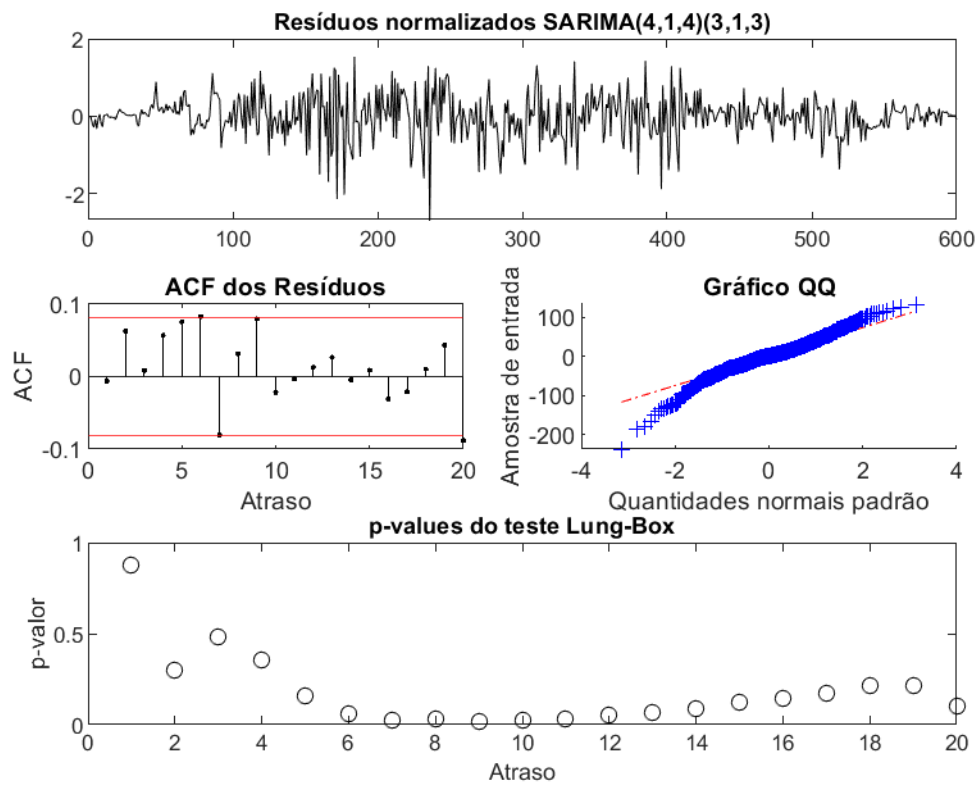
Análise de Resíduos do modelo SARIMA(4,1,4)(2,1,2)



Análise de Resíduos do modelo SARIMA(5,1,5)(2,1,2)



Análise de Resíduos do modelo SARIMA(3,1,3)(3,1,3)



Análise de Resíduos do modelo SARIMA(4,1,4)(3,1,3)