

INSTITUTO DE ENGENHARIA NUCLEAR

JOSÉ LUIZ RODRIGUES NEVES CUNHA

**UTILIZAÇÃO DA DECOMPOSIÇÃO STL JUNTAMENTE COM O MÉTODO DE
SUAVIZAÇÃO EXPONENCIAL SIMPLES E MODELO ARMA PARA PREVISÃO
DAS MEDIDAS DE VELOCIDADE DO VENTO NAS TORRES METEOROLÓGICAS
DA CNAAA**

Rio de Janeiro

2022

JOSÉ LUIZ RODRIGUES NEVES CUNHA

UTILIZAÇÃO DA DECOMPOSIÇÃO STL JUNTAMENTE COM O MÉTODO DE SUAVIZAÇÃO EXPONENCIAL SIMPLES E MODELO ARMA PARA PREVISÃO DAS MEDIDAS DE VELOCIDADE DO VENTO NAS TORRES METEOROLÓGICAS DA CNAAA

Dissertação apresentada ao Programa de Pós-graduação em Ciência e Tecnologia Nucleares do Instituto de Engenharia Nuclear da Comissão Nacional de Energia Nuclear como parte dos requisitos necessários para a obtenção do Grau de Mestre em Ciência em Engenharia Nuclear – Profissional em Métodos Computacionais Aplicados.

ORIENTADOR: PROF. DR. CLÁUDIO MÁRCIO NASCIMENTO ABREU PEREIRA

Rio de Janeiro

2022

CUNH CUNHA Rodrigues Neves, José Luiz

Utilização da Decomposição STL Juntamente com o Método de Suavização Exponencial Simples e Modelo ARMA para previsão das medidas de velocidade do vento nas torres meteorológicas da CNAAA / José Luiz Rodrigues Neves Cunha – Rio de Janeiro: CNEN/IEN, 2022.

xxi, 156f. : il.; 31 cm.

Orientador: Cláudio Márcio Nascimento Abreu Pereira

Dissertação (Mestrado em Ciência e Tecnologia Nucleares) – Instituto de Engenharia Nuclear, PPGIEN, 2022.

1. Previsão em Séries Temporais. 2. Decomposição STL. 3. Métodos de Suavização Exponencial Simples, Holt e Holt-Winters. 4. Modelos ARMA e SARIMA. 5. Decomposição Clássica. 6. Velocidade do Vento. 7. Sistema de Controle Ambiental.

UTILIZAÇÃO DA DECOMPOSIÇÃO STL JUNTAMENTE COM O MÉTODO DE SUAVIZAÇÃO EXPONENCIAL SIMPLES E MODELO ARMA PARA PREVISÃO DAS MEDIDAS DE VELOCIDADE DO VENTO NAS TORRES METEOROLÓGICAS DA CNAAA.

José Luiz Rodrigues Neves Cunha

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA NUCLEARES DO INSTITUTO DE ENGENHARIA NUCLEAR COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA E TECNOLOGIA NUCLEARES.

Aprovada por:

Prof. Cláudio Márcio do Nascimento Abreu Pereira, D.Sc.

Prof. Cláudio Henrique dos Santos Grecco, D.Sc.

Prof. Roberto Schirru, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2022

AGRADECIMENTOS

Ao professor e amigo Cláudio Márcio Nascimento Abreu Pereira, pela orientação e confiança. Referência profissional e humana, autor intelectual deste trabalho, que através da sua sábia e hábil orientação o tornou possível.

Neste momento, cabe enaltecer o empenho de todo corpo docente do IEN, que diante das restrições pandêmicas se reinventaram e realizaram a proeza de oferecer um curso de alta qualidade.

Ao professor e amigo Roberto Schirru, com quem trabalhei por mais de trinta anos, por todo aprendizado adquirido durante essa jornada, contribuindo de maneira inestimável para a minha evolução profissional e acadêmica.

Ao professor emérito Luiz Pereira Calôba, do qual tive o privilégio de ser aluno e responsável pelos meus primeiros passos em análise de séries temporais e redes neurais.

Ao professor Alexandre Cunha Costa, pelo esmero do seu curso de análise de séries temporais e pela atenção e compromisso com seus alunos.

Ao engenheiro e amigo Paulo Fernando Bittencourt Sampaio da usina nuclear Angra 1 que, em confiança, disponibilizou os dados meteorológicos aqui utilizados.

Aos amigos Dra. Célia Maria Oliveira Carvalho, Dra. Maria Emília Carvalho e Rav Shmuel Lemle pelo importante apoio em momentos cruciais.

Aos amigos Vinícius Drumond Gonzaga - pelo incansável incentivo e ajuda - e professor Alan Miranda Monteiro de Lima pela disponibilidade e auxílio.

À secretaria do IEN, L. Carla Brum, e a todo o corpo técnico desse instituto pelo apoio prestado no decorrer do curso.

Aos colegas de turma, que fizeram dessa um exemplo de união e solidariedade.

Ao Criador, por tudo.

RESUMO

No caso de um acidente nuclear com liberação de efluentes radioativos para o meio ambiente, o Sistema de Controle Ambiental (SCA) existente no complexo nuclear CNAAA, monitora o deslocamento da pluma radioativa produzida e faz uma previsão para os períodos de uma e duas horas à frente. Tal monitoração está calcada sobre o tipo de acidente ocorrido e os dados atmosféricos coletados de torres meteorológicas instaladas nas proximidades do complexo. Desta forma, o SCA monitora a distribuição espacial de doses, taxas de dose e a progressão da pluma radioativa gerada, esboçando-a sobre um mapa da região de interesse. A previsão de uma e duas horas realizada pelo SCA utiliza o método conhecido por *naïve method* ou método ingênuo de previsão, que repete o último valor conhecido. Seus resultados se mostram satisfatórios à medida que o próximo valor a ser coletado não difira substancialmente do último conhecido. Entretanto, observando-se dados reais, adquiridos *in loco*, percebe-se que esse fato não ocorre na prática, o que pode tornar a previsão consideravelmente errônea.

A proposta deste trabalho consiste em agregar um modelo híbrido ao método de previsão do SCA de modo a torná-lo mais preciso diante de situações especiais. Para tal foi utilizada a decomposição tendência-sazonal usando Loess (STL, do inglês, *Seasonal and Trend decomposition using Loess*) juntamente com o método de suavização exponencial simples (SES, do inglês, *Simple Exponential Smoothing*) e o modelo autorregressivo de médias móveis (ARMA, do inglês, *AutoRegressive-Moving-Average*) para previsão de parâmetros meteorológicos no sítio de Angra, pois o STL conta com um método de decomposição matemática mais robusto, capaz de lidar satisfatoriamente com séries temporais que possuam alta frequência de aquisição, as chamadas séries intradiárias. Esse modelo utiliza o método Loess, que é uma técnica de suavização fundamentada em regressões ajustadas localmente. No escopo deste trabalho, o método é aplicado na previsão de medidas de velocidade do vento feitas em uma das torres meteorológicas utilizadas pelo SCA. Com a utilização do modelo híbrido desenvolvido foi alcançado um acréscimo de acurácia em relação à previsão do SCA de 101.39% e com a desejada estabilidade estatística.

Palavras-Chave: Previsão em séries temporais, Decomposição STL, Métodos de Suavização Exponencial Simples, Método de Holt, Método Hot-Winters, Modelos ARMA e SARIMA, Decomposição Clássica, Velocidade do Vento, Sistema de Controle Ambiental.

ABSTRACT

In the case of a nuclear accident with release of radioactive effluents into the environment, the Environmental Control System (ECS) existing in the CNAAA nuclear complex, monitors the displacement of the radioactive plume produced and makes a forecast for periods of one and two hours at front. Such monitoring is based on the type of accident that occurred and the atmospheric data collected from meteorological towers installed in the vicinity of the complex. In this way, the ECS monitors the spatial distribution of doses, dose rates and the progression of the generated radioactive plume, outlining it on a map of the region of interest. The one- and two-hour prediction performed by the ECS uses a method known as the naïve method or naive prediction method, which repeats the last known value. Its results are satisfactory as long as the next value to be collected does not differ substantially from the last known one. However, observing real data, acquired in loco, it is clear that this fact does not occur in practice, which can make the forecast considerably erroneous.

The purpose of this work is to add a hybrid model to the ECS forecasting method in order to make it more accurate in special situations. To this end, the Seasonal and Trend decomposition using Loess (STL) was used together with the Simple Exponential Smoothing (SES) method and the AutoRegressive Moving Average (ARMA) model to forecast weather parameters at the Angra site, because STL has a more robust mathematical decomposition method, capable of satisfactorily dealing with time series that have a high frequency of acquisition, the so-called intraday series. This model uses the Loess method, which is a smoothing technique based on locally adjusted regressions. In the scope of this work, the method is applied to forecast wind speed measurements made at one of the meteorological towers used by the ECS. Using the hybrid model developed, an increase in accuracy compared to the ECS forecast of 101.39% was achieved, with the desired statistical stability.

Keywords: Forecast in time series, STL Decomposition, Simple Exponential Smoothing Method, Holt Method, Holt-Winters Method, ARMA and SARIMA Models, Classical Decomposition, Wind Speed, Environmental Control System.

“A teoria produz um bom resultado, mas dificilmente nos aproxima do segredo do Criador. Estou, em todos os casos, convencido de que Ele não joga dados.”

Albert Einstein, dezembro de 1926.

Aos meus pais e irmã, dedico.

LISTA DE FIGURAS

Figura 1 - Seção vertical da atmosfera na área costeira, indicando a brisa marítima e a brisa terrestre	2
Figura 2 - Representação da brisa de vale e montanha.....	3
Figura 3 - Fatores que influenciam o perfil vertical da velocidade do vento	4
Figura 4 - Previsão calculada com intervalo de confiança de 50%	7
Figura 5 - Pluma radioativa gerada durante um Grande Loca	14
Figura 6 - Previsão do deslocamento da pluma radioativa após uma hora	15
Figura 7 - Previsão do deslocamento da pluma radioativa após duas horas	15
Figura 8 - Totais mensais de passageiros aéreos da PAN AM entre os anos 1949 a 1960	17
Figura 9 - Ilustração da estrutura básica de um boxplot vertical	18
Figura 10 - Distribuição das velocidades do vento com e sem o tratamento de <i>outliers</i>	19
Figura 11 - Dispersão das velocidades do vento com e sem o tratamento de outliers	19
Figura 12 - Função de autocorrelação das velocidades do vento com e sem <i>outliers</i>	20
Figura 13 - Covariância positiva indicando tendência crescente	28
Figura 14 - Covariância negativa indicando tendência decrescente	28
Figura 15 - Covariância nula indicando ausência de tendência	29
Figura 16 - Totais mensais de passageiros aéreos da PAN AM entre os anos 1949 a 1960 ...	33
Figura 17 - Série AirPassengers após uma diferenciação de $h = 12$	34
Figura 18 - Série AirPassengers após uma diferenciação de segunda ordem (12 e 1)	35
Figura 19 - Dispersão de AirPassengers em função de $h = 1$	36
Figura 20 - FAC para cada uma das 21 defasagens da série AirPassengers estacionária	38
Figura 21 - Correlogramas para série AirPassengers estacionária. Fonte: Wessa (2017)	38
Figura 22 - Correlogramas do ruído branco gaussiano gerado.....	40
Figura 23 - Conjunto de 100 números randômicos gerados com distribuição normal padrão .	43
Figura 24 - Histograma e curva densidade de probabilidade do conjunto normal padrão gerado	44
Figura 25 - Gráfico Q-Q plot para o conjunto normal padrão gerado	45
Figura 26 - Série Temporal hipotética com suas médias móveis de 3, 5 e 7 períodos	52
Figura 27 - Série temporal cobrindo 15 anos, com periodicidade mensal	54
Figura 28 - Série temporal de vendas com 5 modelos de tendência	55
Figura 29 - Série temporal e média móvel centrada, ambas com periodicidade igual a 12 ...	56
Figura 30 - Determinação dos Índices Sazonais Aditivos	57

Figura 31 - Determinação dos Índices Sazonais Multiplicativos	60
Figura 32 - Plotagem dos índices de CI calculados no modelo aditivo	61
Figura 33 - Plotagem dos índices de CI calculados no modelo multiplicativo	62
Figura 34 - Modelo AR(1) gerado.....	72
Figura 35 - Correlogramas da FAC e FACP do modelo AR(1) gerado	73
Figura 36 - Modelo AR(2) gerado	74
Figura 37 - Correlogramas da FAC e FACP do modelo AR(2) gerado	74
Figura 38 - Modelo MA(1) gerado	76
Figura 39 - Correlogramas da FAC e FACP do modelo MA(1) gerado	76
Figura 40 - Modelo MA(2) gerado	77
Figura 41 - Correlogramas da FAC e FACP do modelo MA(2) gerado	77
Figura 42 - Roteiro para escolha de um modelo.....	81
Figura 43 - Decomposição STL sobre a série calls mostrando sua múltipla sazonalidade	85
Figura 44 - Disposição das torres meteorológicas no sítio das usinas nucleares de Angra	90
Figura 45 - Velocidade do vento tomada a cada hora entre os anos 2013 e 2020	103
Figura 46 - Índices CI para o modelo aditivo.....	109
Figura 47 - Índices CI para o modelo multiplicativo	110
Figura 48 - Velocidade do vento registrada pela torre C em maio de 2016	111
Figura 49 - Velocidade do vento dos cinco primeiros dias de maio.....	112
Figura 50 - Série WS_m transformada através da aplicação do LN	113
Figura 51 - Conjunto de Treinamento (vermelho) e Teste (verde) de WS_m após transformação LN.....	114
Figura 52 - 5 últimas horas do conjunto de treinamento e 5 primeiras horas do teste	115
Figura 53 - Ajuste Holt-Winters (vermelho) sobre o conjunto de treinamento (preto).....	115
Figura 54 - Gráfico dos resíduos gerados pelo método Holt-Winters.....	116
Figura 55 - Gráfico QQ-plot dos resíduos gerados pelo modelo Holt-Winters.....	117
Figura 56 - Histograma e curva densidade de probabilidade dos resíduos gerados pelo modelo Holt-Winters	118
Figura 57 - Correlograma FAC dos resíduos gerados pelo modelo Holt-Winters	119
Figura 58 - Parte do conjunto de treinamento (vermelho), todo o conjunto de teste (verde) e a previsão Holt-Winters (azul) sobre o conjunto de teste	120
Figura 59 - Análise dos resíduos após aplicação de ADF com valores default	122
Figura 60 - Análise dos resíduos após aplicação do ADF com lag = 38	123

Figura 61 - Existência de 2 raízes (vermelho) fora do círculo unitário	125
Figura 62 - Série WS_m após uma transformação logarítmica	126
Figura 63 - WS_m após as transformações de Log e tsclen	127
Figura 64 - WS_m após as transformações de Log e tsclen e diferenciação de segunda ordem	128
Figura 65 - Gráfico QQ-plot de WS_m após as transformações de Log, tsclen e diferenciações de segunda ordem	128
Figura 66 - Histograma e curva densidade de probabilidade de WS_m após as transformações e diferenciação de segunda ordem.....	129
Figura 67 - FAC e FACP de WS_m após diferenciação de segunda ordem	130
Figura 68 - Todas as raízes unitárias estão no interior do círculo	132
Figura 69 - Gráfico dos resíduos gerados pelo modelo SARIMA	132
Figura 70 - Gráfico QQ-plot dos resíduos gerados pelo modelo SARIMA	133
Figura 71 - Histograma e curva densidade de probabilidade dos resíduos gerados pelo modelo SARIMA.....	133
Figura 72 - Correlograma FAC dos resíduos gerados pelo modelo SARIMA	134
Figura 73 - Ajuste SARIMA (vermelho) sobre o conjunto de treinamento (preto)	135
Figura 74 - Ajuste SARIMA (azul) sobre o conjunto de teste (preto)	136
Figura 75 - Previsão SARIMA para todo o conjunto de teste.....	137
Figura 76 - Decomposição STL sobre o conjunto de treinamento	139
Figura 77 - Componente trend (preto) e valores replicados para Tt (tendência) (vermelho) 140	
Figura 78 - Dez últimos dias de seasonal (preto) e valores replicados para St (efeito sazonal) (vermelho)	141
Figura 79 - Componente restante (remainder) obtida pela decomposição STL	141
Figura 80 - Correlogramas da componente de restante (remainder) obtida pela decomposição STL.....	142
Figura 81 - Raízes unitárias do modelo ARMA(6,11)	143
Figura 82 - FAC dos resíduos obtidos pelo ajuste do modelo ARMA(6,11)	143
Figura 83 - Parte do conjunto de treinamento (vermelho), todo o conjunto de teste (verde) e a previsão STL/SES/ARMA sobre o conjunto de teste (azul)	144
Figura 84 - Treinamento (vermelho), oito primeiros valores de teste (verde), previsão de oito valores do SCA (preto) e previsão de oito valores do STL/SES/ARMA (azul).....	145

LISTA DE TABELAS

Tabela 1 - Teste de raiz unitária ADF para a série AirPassengers original.....	33
Tabela 2 - Teste de raiz unitária ADF para a série AirPassengers após a diferenciação de 12....	
.....	34
Tabela 3 - Teste ADF para a série AirPassengers após a diferenciação de segunda ordem (12 e 1)	35
Tabela 4 - FAC calculado para defasagens de até 21 meses à frente sobre AirPassengers	37
Tabela 5 - Valor do teste de hipóteses de normalidade Shapiro-Wilker	45
Tabela 6 - Posicionamento das médias móveis de 3, 5 e 7 períodos	52
Tabela 7 - Posicionamento das médias móveis de 4 períodos	53
Tabela 8 - Representante de índices Sazonais Aditivos para cada um dos 12 períodos específicos	58
Tabela 9 - Representante de índices Sazonais Multiplicativos para cada um dos 12 períodos específicos	60
Tabela 10 - Os 12 representantes do índice sazonal aditivo (coluna BR) repetidos para cada ano	64
Tabela 11 - Componentes T, representantes S e CI calculadas para a recomposição no modelo aditivo	65
Tabela 12 - Associação dos modelos AR e MA às FAC e FACP	78
Tabela 13 - Casos particulares do modelo ARIMA	83
Tabela 14 - Dados meteorológicos coletados pela Torre A	91
Tabela 15 - Dados meteorológicos coletados pelas Torres B, C e D	92
Tabela 16 - Parte superior dos resultados da análise realizada sobre a planilha QCD01C30 .	94
Tabela 17 - Parte inferior dos resultados da análise realizada sobre a planilha QCD01C30 .	94
Tabela 18 - Número de registros faltantes e inválidos para cada torre no ano de 2013	95
Tabela 19 - Número de registros faltantes e inválidos para cada torre no ano de 2014	96
Tabela 20 - Número de registros faltantes e inválidos para cada torre no ano de 2015	96
Tabela 21 - Número de registros faltantes e inválidos para cada torre no ano de 2016	96
Tabela 22 - Número de registros faltantes e inválidos para cada torre no ano de 2017	96
Tabela 23 - Número de registros faltantes e inválidos para cada torre no ano de 2018	97
Tabela 24 - Número de registros faltantes e inválidos para cada torre no ano de 2019	97
Tabela 25 - Número de registros faltantes e inválidos para cada torre no ano de 2020	97

Tabela 26 - Total geral de registros faltantes e inválidos para cada torre de 2013 a 2020.....	97
Tabela 27 - Valores mensais inválidos de WS e WD, por quarto, para os anos 2013 e 2014 .	98
Tabela 28 - Valores mensais inválidos de WS e WD, por quarto, para os anos 2015 e 2016 ..	99
Tabela 29 - Valores mensais inválidos de WS e WD, por quarto, para os anos 2017 e 2018 ..	99
Tabela 30 - Valores mensais inválidos de WS e WD, por quarto, para os anos 2019 e 2020 .	100
Tabela 31 - Total anual e geral dos valores inválidos de WS e WD, por quarto, para a torre D	100
Tabela 32 - Valores das métricas para cada modelo de regressão	104
Tabela 33 - Representantes dos índices (coluna BB) e seus valores corrigidos (coluna BC) ..	105
Tabela 34 - Somatório da coluna BB diferente de zero e somatório da BC (corrigido) igual a zero	106
Tabela 35 - Representantes dos índices (Coluna BH) e a quarta correção (coluna BL)	107
Tabela 36 - Somatório de BH diferente de 8760 e somatório de BL (corrigido) igual a 8760 ..	107
Tabela 37 - Valores calculados dos índices de CI para o modelo aditivo (coluna BR)	108
Tabela 38 - Valores calculados dos índices de CI para o modelo multiplicativo (coluna BY)....	109
Tabela 39 - Teste de normalidade Shapiro-Wilker para os resíduos gerados pelo modelo Holt-Winters	118
Tabela 40 - Medidas de acurácia da previsão Holt-Winters para os 8 primeiros valores de teste	120
Tabela 41 - Medidas de acurácia para previsão Holt-Winters sobre todo o conjunto de teste	121
Tabela 42 - Valores do teste de raiz unitária ADF para a série WS_m não diferenciada	123
Tabela 43 - Valores do teste de raiz unitária ADF para a série WS_m após uma diferenciação I(1)	124
Tabela 44 - Valores do teste de raiz unitária KPSS para a série WS_m não diferenciada ...	124
Tabela 45 - Valores do teste de raiz unitária KPSS para a série WS_m após uma diferenciação I(1)	124
Tabela 46 - Valor do teste de hipóteses de normalidade Shapiro-Wilker para WS_m após as transformações e diferenciação de segunda ordem	129
Tabela 47 - Teste de normalidade Shapiro-Wilker para os resíduos gerados pelo modelo SARIMA.....	134

Tabela 48 - Medidas de acurácia da previsão SARIMA para os 8 primeiros valores de teste	137
.....
Tabela 49 - Medidas de acurácia da previsão SARIMA para todo o conjunto de teste	138
Tabela 50 - Medidas de acurácia da previsão STL/SES/ARMA sobre os 8 primeiros valores de teste	145
Tabela 51 - Medidas de acurácia da previsão STL/SES/ARMA sobre todo o conjunto de teste.	
.....	146
Tabela 52 - Medidas de acurácia da previsão realizada pelo SCA para os 8 primeiros valores de teste.....	148
Tabela 53 - Medidas de acurácia da previsão Holt-Winters para os 8 primeiros valores de teste	
.....	148
Tabela 54 - Medidas de acurácia da previsão SARIMA para os 8 primeiros valores de teste.....	149
.....
Tabela 55 - Medidas de acurácia da previsão STL/SES/ARMA para os 8 primeiros valores de teste.....	149

LISTA DE ABREVIATURAS E SIGLAS

AIC	- Critério de Informação Akaike
AICc	- Critério de Informação Akaike corrigido
AR	- Modelo Autorregressivo
ARMA	- Modelo Autorregressivo de Média Móvel
ARIMA	- Modelo Autorregressivo Integrado de Média Móvel
BIC	- Critério de Informação Bayesiano
CNAAA	- Central Nuclear Almirante Álvaro Alberto
CNEN	- Comissão Nacional de Energia Nuclear
DF	- Teste de estacionariedade de Dickey-Fuller
DFA	- Teste de estacionariedade de Dickey-Fuller Aumentado
FAC	- Função de AutoCorrelação
FACP	- Função de AutoCorrelação Parcial
H_0	- Hipótese Nula
H_a	- Hipótese Alternativa
IEN	- Instituto de Engenharia Nuclear
IID	- Independente e Identicamente Distribuído
KPSS	- Teste de estacionariedade de Kwiatkowski-Phillips-Schmidt-Shin
LMP	- Laboratório de Monitoração de Processos - COPPE/UFRJ
MA	- Modelo de Média Móvel
MAE	- Erro Absoluto Médio
MAPE	- Erro Absoluto Percentual Médio
MPE	- Erro Percentual Médio
SARIMA	- Modelo Autorregressivo Integrado de Média Móvel com Sazonalidade
SCA	- Sistema de Controle Ambiental
SES	- Suavização Exponencial Simples
STL	- Decomposição Sazonal e Tendência usando Loess
TBATS	- Sazonalidade Trigonométrica, Transformação Box-Cox, ARMA, Componentes de Tendência e Sazonal

ÍNDICE

1 – INTRODUÇÃO	1
1.1 - CONTEXTUALIZAÇÃO E DEFINIÇÃO DO PROBLEMA	1
1.1.1 - FATORES QUE INFLUENCIAM A VELOCIDADE E DIREÇÃO DO VENTO	1
1.1.2 - METODOLOGIA DE PREVISÃO DO SCA	4
1.1.3 - A PREVISÃO TRATADA PELA ANÁLISE DE SÉRIES TEMPORAIS	4
1.1.4 - RELEVÂNCIA DA PREVISÃO	5
1.1.5 - TRABALHOS RELACIONADOS	7
1.2 - JUSTIFICATIVA	13
1.3 - OBJETIVO	13
2 - FUNDAMENTAÇÃO TEÓRICA	15
2.1. ESTRUTURA E COMPONENTES DE UMA SÉRIE TEMPORAL.....	16
2.2. REPRESENTAÇÃO GRÁFICA DE UMA SÉRIE TEMPORAL.....	16
2.3. ANÁLISE DOS <i>OUTLIERS</i>	17
2.3.1. <i>BOXPLOT</i>	18
2.4 – DIFERENCIADA	20
2.5 - TRANSFORMAÇÃO LOGARÍTMICA	21
2.6 - TRANSFORMAÇÃO BOX-COX	22
2.7 - TESTES DE RAIZ UNITÁRIA	22
2.8 - ESTACIONARIEDADE	22
2.8.1 - TESTE DE ESTACIONARIEDADE DE RAÍZ UNITÁRIA	23
2.9 - ANÁLISE DESCRIPTIVA DE SÉRIES TEMPORAIS	25
2.10 - DISPERSÃO, VARIÂNCIA, DESVIO PADRÃO, COVARIÂNCIA E AUTOCORRELAÇÃO	26
2.10.1 - DISPERSÃO	26
2.10.2 - VARIÂNCIA	26
2.10.3 - DESVIO PADRÃO	26
2.10.4 - COVARIÂNCIA	27
2.10.5 - CORRELAÇÃO	29
2.10.6 - AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL	30
2.11 - CORRELOGRAMA	38

2.12 - RUÍDO BRANCO	40
2.12.1 - RUÍDO BRANCO GAUSSIANO	41
2.13 - MODELAGEM	46
2.14 - RESÍDUOS	46
2.15 - CRITÉRIOS DE INFORMAÇÃO	47
2.16 - DECOMPOSIÇÃO DE SÉRIES TEMPORAIS - MODELO CLÁSSICO	47
2.16.1 - OBTENÇÃO DA COMPONENTE TENDÊNCIA NO MODELO CLÁSSICO	49
2.16.1.1 - OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE MODELOS DE REGRESSÃO ..	49
2.16.1.2 - OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE MÉDIAS MÓVEIS CENTRADAS	50
2.16.1.3 - OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE AJUSTE EXPONENCIAL	53
2.16.2 - OBTENÇÃO DA COMPONENTE SAZONAL NO MODELO CLÁSSICO	55
2.16.2.1 - MODELO ADITIVO	55
2.16.2.2 - MODELO MULTIPLICATIVO	58
2.16.3 - OBTENÇÃO DAS COMPONENTES CICLO E IRREGULAR (CI)	60
2.16.4 - PROCESSO DE RECOMPOSIÇÃO	63
2.16.4.1 - RECOMPOSIÇÃO PELO MODELO ADITIVO	63
2.16.4.2 - RECOMPOSIÇÃO PELO MODELO MULTIPLICATIVO	65
2.16.5 - PROCESSO DE PREVISÃO	65
2.17 - MÉTODOS DE SUAVIZAÇÃO EXPONENCIAL	66
2.17.1 - MÉTODO DE SUAVIZAÇÃO EXPONENCIAL SISMPLES (SES)	67
2.17.2 - MÉTODO DE HOLT	68
2.17.3 - MÉTODO DE HOLT-WINTERS	69
2.18 - MODELOS AR, MA, ARMA, ARIMA E SARIMA	70
2.18.1 - MODELO AUTORREGRESSIVO - AR(p)	71
2.18.2 - MODELOS DE MÉDIA MÓVEL - MA(q)	74
2.18.3 - MODELOS ARMA	79
2.18.3.1 - DETERMINAÇÃO DO MODELO ARMA	80
2.18.4 - MODELOS ARIMA	82
2.18.5 - MODELOS SARIMA	83
2.19 - MÉTODOS DE PREVISÃO PARA SÉRIES TEMPORAIS DE ALTA FREQUÊNCIA	84
2.19.1 - DECOMPOSIÇÃO DE TENDÊNCIA-SAZONAL USANDO LOESS – STL	86

2.19.2 - DECOMPOSIÇÃO STL NO SOFTWARE ESTATÍSTICO R	88
2.20 - TESTE Q DE LJUNG-BOX	89
3 - ANÁLISE DOS DADOS COLETADOS PELAS TORRES METEOROLÓGICAS	90
3.1 - AQUISIÇÃO DOS DADOS METEOROLÓGICOS	90
3.2 - ESTRUTURA DOS ARQUIVOS TRANSMITIDOS PELAS TORRES	91
3.3 - ADEQUAÇÃO DOS ARQUIVOS À ESTRUTURA DE UMA SÉRIE TEMPORAL .	92
3.4 - ANÁLISE DE CONSISTÊNCIA DOS REGISTROS ORIGINAIS	93
3.5 - ESCOLHA DA TORRE E QUARTO DE HORA MAIS ADEQUADOS	95
3.5.1 - ESCOLHA DA TORRE PARA REPRESENTAÇÃO DA SÉRIE TEMPORAL WS_h	95
3.5.2 - QUARTO DE HORA PARA REPRESENTAÇÃO SÉRIE TEMPORAL	98
3.5.3 - PREENCHIMENTO DOS DADOS FALTANTES E INVÁLIDOS	100
3.5.4 - ANÁLISE DE OUTLIERS	102
4 - APLICAÇÃO DOS MÉTODOS E MODELOS DE ANÁLISE DE SÉRIES TEMPORAIS NA PREVISÃO DE VELOCIDADE DO VENTO	103
4.1 - MODELO CLÁSSICO DE ANÁLISE DE SÉRIES TEMPORAIS	103
4.1.1 - DEFINIÇÃO DA SÉRIE TEMPORAL	103
4.1.2 - OBTENÇÃO DA COMPONENTE DE TENDÊNCIA	103
4.1.3 - OBTENÇÃO DA COMPONENTE SAZONAL	104
4.1.3.1 - MÉDIAS MÓVEIS CENTRADAS	104
4.1.3.2 - ÍNDICES SAZONais ADITIVOS E MULTIPLICATIVOS	104
4.1.4 - OBTENÇÃO DAS COMPONENTES CI	107
4.2 - DEFININDO A SÉRIE TEMPORAL PARA OS PRÓXIMOS MÉTODOS E MODELOS	110
4.2.1 - APLICANDO A TRASNFORMAÇÃO LOGARÍTMICA SOBRE A SÉRIE WS_m	112
4.2.2 - DEFINIÇÃO DO CONJUNTO DE TREINAMENTO E TESTE SOBRE A SÉRIE WS_m APÓS SUA TRANSFORMAÇÃO LOGARÍTMICA	113
4.3 - APLICAÇÃO DO MÉTODO DE HOLT-WINTERS	115
4.4 - APLICAÇÃO DO MODELO AUTOREGRESSIVO SARIMA	121
4.4.1 - AVALIAÇÃO DA ESTACIONARIEDADE DA SÉRIE	121
4.4.2 - DETERMINAÇÃO DAS ORDENS DO MODELO SARIMA	125

4.5 - APLICAÇÃO DA DECOMPOSIÇÃO STL/SES/ARMA SOBRE A SÉRIE WS_m	138
5 - ANÁLISE DOS RESULTADOS.....	147
5.1 - ACURÁCIA DA PREVISÃO DO MODELO UTILIZADO PELO SCA	147
5.2 - ACURÁCIA DA PREVISÃO DO MODELO HOLT-WINTERS	148
5.3 - ACURÁCIA DA PREVISÃO DO MODELO SARIMA	148
5.4 - ACURÁCIA DA PREVISÃO DO MODELO STL/SES/ARMA	149
6 - CONCLUSÕES	150
7 - TRABALHOS FUTUROS	152
REFERÊNCIAS	153

1 – INTRODUÇÃO

1.1 – CONTEXTUALIZAÇÃO E DEFINIÇÃO DO PROBLEMA

A segurança é o princípio preponderante em qualquer atividade envolvendo energia nuclear, sobrepondo-se ao próprio objetivo da atividade ou quaisquer outros interesses. Assim, as usinas nucleares contam com sistemas de segurança capazes de entrarem automaticamente em ação para a prevenção de acidentes e com procedimentos bem definidos no caso de alguma emergência. Esses procedimentos consistem em ferramentas computacionais e ações preventivas a serem realizadas diante de um acidente ou situação anormal.

O Sistema de Controle Ambiental (SCA) é um exemplo dessas ferramentas computacionais. Desenvolvido para dar suporte aos gerentes das equipes de proteção radiológica, ele fornece as informações necessárias à tomada de decisões no caso de uma eventual liberação de efluentes radioativos para o meio ambiente. Sendo assim, o SCA é responsável pela previsão da distribuição espacial de doses e taxas de dose nas redondezas da CNAAA e pela estimativa da progressão da pluma radioativa formada a partir da liberação acidental. A previsão do deslocamento da pluma realizada pelo SCA compreende intervalos de uma e duas horas à frente e está calcada sobre o tipo de acidente ocorrido e os dados atmosféricos coletados das torres meteorológicas instaladas nas proximidades da CNAAA.

O transporte dos efluentes radioativos liberados na atmosfera é realizado pela ação do vento e sua distribuição espacial é determinada a partir de sua velocidade e direção, assim como a estabilidade atmosférica e o relevo da região.

1.1.1 – FATORES QUE INFLUENCIAM A VELOCIDADE E DIREÇÃO DO VENTO

Segundo Cavalcanti *et al.* (2009), a proximidade do oceano é um primeiro fator regulador das condições climáticas numa mesma latitude. A grande capacidade térmica do oceano reduz as variações diurnas de temperatura, tanto pela proximidade como pela grande quantidade de vapor d'água proveniente do oceano que se distribui nas proximidades. O alto conteúdo de umidade do ar absorve a radiação térmica emitida pela superfície da terra na faixa do infravermelho, radiação esta reemitida para baixo como parte do chamado efeito estufa. Em locais distantes do oceano, o ar mais seco não retém o calor emitido pela superfície, que é então

perdido para o espaço. Nesses casos, o esfriamento noturno é mais pronunciado, podendo ocasionar temperaturas baixas à noite. Outro fator regulador é a altitude do local. Os fatores mencionados – a proximidade do oceano e a altitude do local – são dois controles muito importantes no clima local e estão associados a regimes de vento local denominados brisa marítima e terrestre e brisa de vale e de montanha, respectivamente. A brisa terrestre ocorre durante à noite, quando a terra se resfria mais rapidamente que o oceano, revertendo o gradiente de temperatura que estava presente durante o dia. Assim, o oceano, por ter capacidade calorífica maior que da terra e por permitir misturas verticais, demora mais a se aquecer e esfriar que a superfície terrestre. Então, durante à noite, o ar sobre a terra fica mais frio que o ar sobre o mar. Isso gera uma diferença de pressão e o ar sopra da terra para o mar. Durante o dia, por volta das 10 horas, ocorre o inverso e a brisa marítima sopra do mar para a terra, Figura 1. Em regiões altas, durante à noite, o ar se resfria mais do que nos vales; o ar frio é mais pesado e flui para os vales formando a brisa de montanha. Durante o dia, o sol aquece as encostas com face para o leste antes de aquecer o vale, gerando a brisa de vales para a montanha, Figura 2.

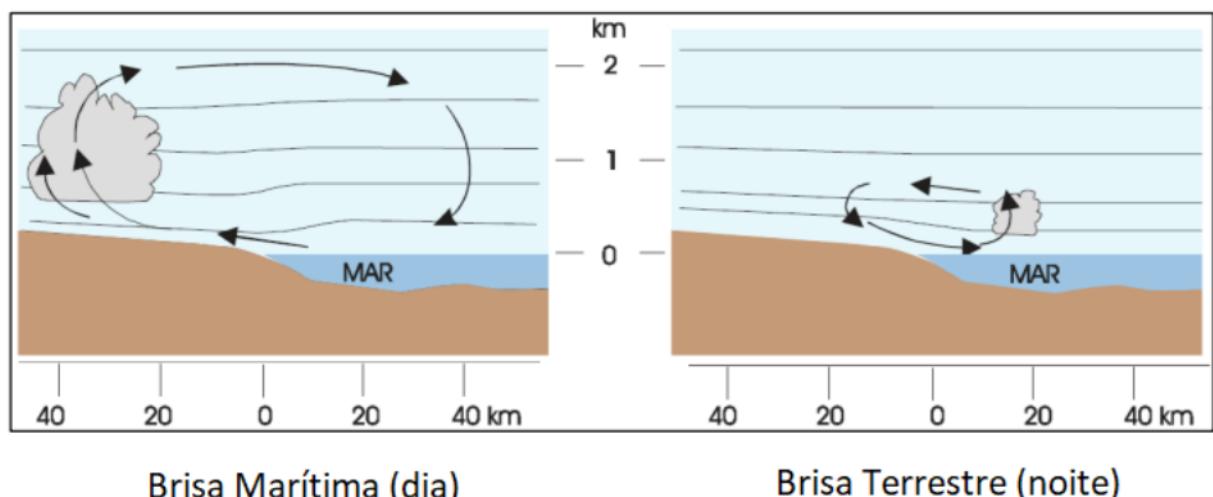


Figura 1 – Seção vertical da atmosfera na área costeira indicando a brisa marítima e a brisa terrestre. As linhas horizontais são isóbaras esquemáticas e as setas indicam a direção do vento e/ou do movimento vertical.

Fonte: Varejão Silva (2006).

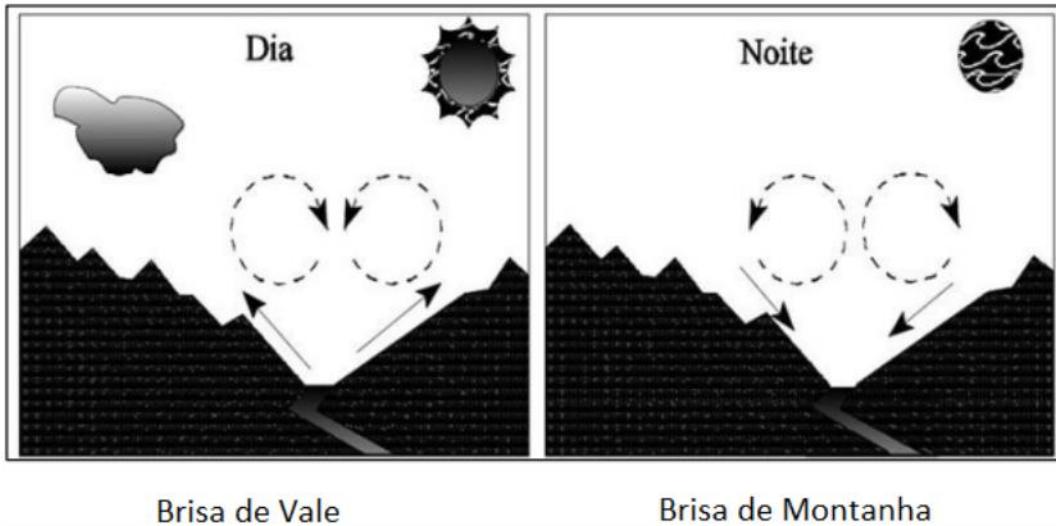


Figura 2 – Representação da brisa de vale e montanha.

Fonte: Silva (2003).

No estudo das circulações atmosféricas, costuma-se categorizar os diversos sistemas de circulação de acordo com sua escala espacial e temporal. Na escala local ou microescala, os processos atmosféricos são da ordem de 1 metro a 1 km, e duração da ordem de 1 segundo até 1 hora. Nessa escala estão presentes os processos de formação das gotas de chuva, tornados, convecção profunda, dispersão de poluentes e interações no sistema solo-planta-atmosfera, redemoinhos de poeira, efeitos aerodinâmicos relacionados à topografia local e a obstáculos naturais ou artificiais, efeitos associados à rugosidade e às variações de fluxo de calor e umidade. Esses dois últimos efeitos produzem fortes variações de alta frequência na velocidade do vento, conhecidas como turbulências atmosféricas.

A Figura 3 apresenta alguns fatores locais que influenciam a direção e a intensidade do vento, a saber, tipo de terreno (plano ou acidentado), rugosidade (seu comprimento se modifica de acordo com as mudanças naturais na paisagem), obstáculos naturais e resultantes de ação humana (ocasionam o efeito de sombreamento que está ligado às dimensões horizontais e à altura dos obstáculos) e orografia (variações na altura do terreno: colinas, vales, depressões etc.).

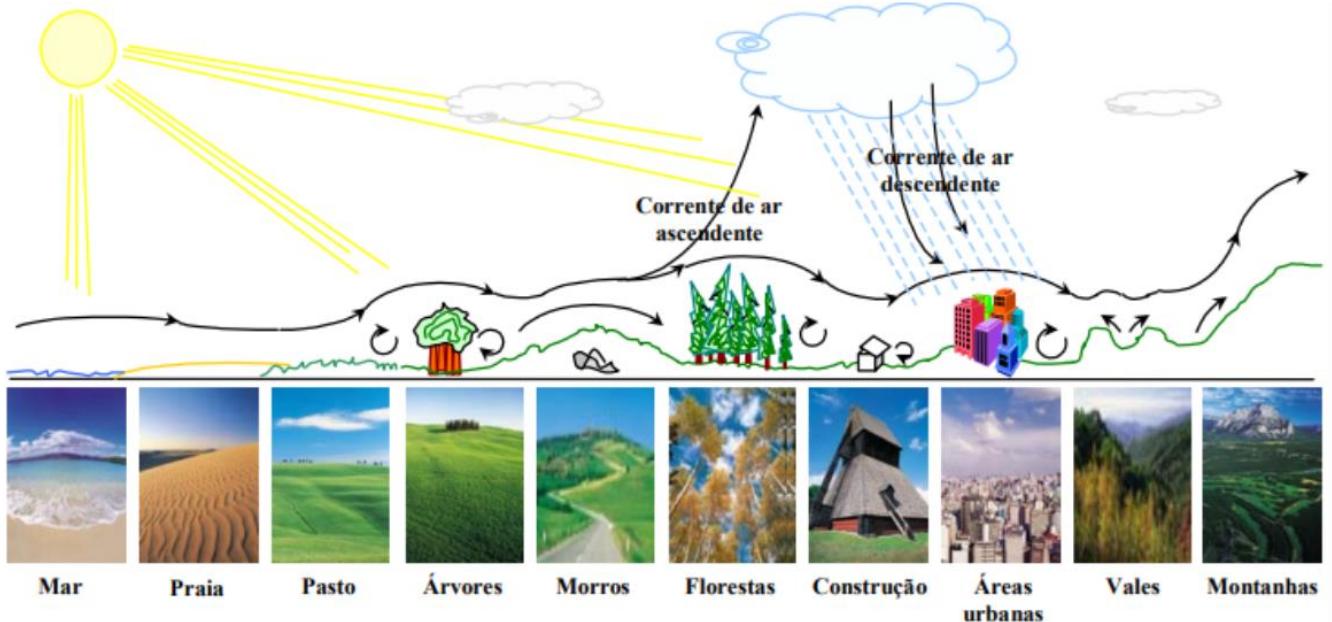


Figura 3 – Fatores que influenciam o perfil vertical da velocidade do vento.

Fonte: Silva (2003).

Excetuando-se o pasto da figura acima, a localização da CNAAA abrange os demais fatores que influenciam o perfil vertical do vento, tornando a tarefa de previsão da velocidade do vento bastante complexa.

1.1.2 – METODOLOGIA DE PREVISÃO DO SCA

A previsão da velocidade do vento realizada pelo SCA para cada quinze minutos, no horizonte de uma e duas horas à frente, é realizada através do método conhecido por *naïve method* ou método ingênuo de previsão, que repete o último valor conhecido. Esse método apresenta resultados eficazes à medida que os dados previstos não se afastem consideravelmente do último valor tomado como referência.

Contudo, como observado *in loco*, devido às altas variações de velocidade do vento na região, situações nas quais o próximo valor de velocidade do vento difere notavelmente do último registrado ocorrem com certa recorrência, podendo tornar este tipo de previsão significantemente errônea.

1.1.3 – A PREVISÃO TRATADA PELA ANÁLISE DE SÉRIES TEMPORAIS

A análise de séries temporais é uma técnica em estatística que lida com dados de séries temporais e análise de tendências. A teoria convencional não se mostra satisfatória o suficiente para lidar com séries temporais provenientes de um cenário geográfico extremamente complexo, incluindo terreno multifacetado, efeitos costeiros e antrópicos. Além disso, a metodologia padrão de análise de séries temporais não é capaz de tratar eficientemente séries temporais com sazonalidades complexas e apresenta limitações estatísticas.

Assim, faz-se necessária a adoção de mecanismos matemáticos mais robustos que possam absorver tais características. A decomposição de séries temporais utilizando o método Loess, que emprega uma estimativa de relacionamentos não lineares ajustadas localmente, se mostra satisfatória no trato de tais séries temporais.

1.1.4 – RELEVÂNCIA DA PREVISÃO

Conforme observado por Hyndman *et al.* (2018), a capacidade de prever o futuro é algo que sempre fascinou a humanidade, acompanhando-a desde os tempos mais remotos até os dias atuais. Por vezes, sendo considerada como uma atividade criminosa e noutras como inspiração divina, cabendo citar o que o profeta Isaías escreveu por volta de 700 AC.:

“Anunciai-nos as coisas que ainda hão de vir, para que saibamos que sois deuses; ou fazei bem, ou fazei mal, para que nos assombremos, e juntamente o vejamos”. (Isaías 41:23)

Nesse sentido, Nielsen (2021) também nos fala sobre as investigações aristotélicas no tocante às previsões meteorológicas, que culminaram com um tratado sobre o tema (Meteorologia) a vigorar até o Renascimento.

Uma boa previsão pode gerar fortunas ao passo que o contrário pode levar seus previsores à bancarrota. Hyndman *et al.* (2018) apresentam algumas previsões famosas sobre computação:

- “Acho que existe um mercado mundial para talvez cinco computadores”. (Presidente da IBM, 1943).
- No futuro, os computadores não podem pesar mais do que 1,5 tonelada. (Mecânica Popular, 1949).

- “Não há razão para que alguém queira um computador em sua casa”. (Presidente da DEC, 1977).

A última frase foi proferida apenas três anos antes da IBM produzir o primeiro computador pessoal, por isso não é surpresa a DEC não mais se encontrar no mercado. A previsão é obviamente uma atividade difícil e as empresas que a fazem bem têm uma grande vantagem sobre aquelas cujas previsões falham.

Numa abordagem macro, Ngai Hang (2010) distingue as previsões em dois tipos, as previsões *ex-post* e *ex-ante*. A primeira trata de previsões quando as observações futuras são conhecidas com algum grau de certeza. Sendo assim, as previsões *ex-post* são usadas como um meio de verificação para avaliação do modelo de previsão e nas previsões do tipo *ex-ante* as observações futuras não estão disponíveis para verificação.

Conforme descrito por George E. P. Box (2016) sobre previsão de curto prazo, estas geralmente são necessárias em um período conhecido como *lead time* ou horizonte, que varia de acordo com cada problema. Por exemplo, uma das séries temporais utilizadas nesse trabalho, que por ora será chamada de z_t , armazena os valores de velocidade do vento registrados em intervalos de tempo discretos e equidistantes, mais precisamente com periodicidade horária. Então a velocidade do vento na hora atual t e as velocidades z_{t-1} , z_{t-2} , z_{t-3} , ... nas horas anteriores podem ser usadas para prever velocidades para períodos de $l = 1, 2, 3, \dots$ horas à frente. Chamando de $\hat{z}_t(l)$ a previsão feita a partir da origem t para velocidades z_{t+l} em algum tempo futuro $t + l$, ou seja, no *lead time* l . A função $\hat{z}_t(l)$, que fornece as previsões na origem t para os *lead times* futuros, com base nas informações disponíveis dos valores atuais (z_t) e anteriores ($z_{t-1}, z_{t-2}, z_{t-3}, \dots$) ao longo do tempo t , será chamada de função de previsão na origem t . O objetivo que devemos empreender é obter uma função de previsão tal que os desvios médios quadráticos entre os valores reais e previstos $z_{t+l} - \hat{z}_t(l)$ seja o menor possível para cada *lead time* l .

Tão importante quanto o cálculo da previsão é a determinação do quão boa essa se apresenta, isto é, qual o grau de acurácia associado à mesma e isto irá variar dependendo da aplicação. A precisão da previsão pode ser expressa pelo cálculo dos limites de probabilidade e seu percentual pode ser estabelecido em função do problema analisado, por exemplo, existem

situações nas quais um erro de aproximadamente 10% é perfeitamente aceitável e, noutras, tal erro seria potencialmente desastroso.

A Figura 4 ilustra os conceitos acima, mostrando os vinte últimos valores observados de uma série temporal hipotética com periodicidade mensal, que culmina no tempo t e apresenta também as treze previsões realizadas a partir da origem t , para os meses $l = 1, 2, \dots, 13$, juntamente com os limites de probabilidade de 50%.

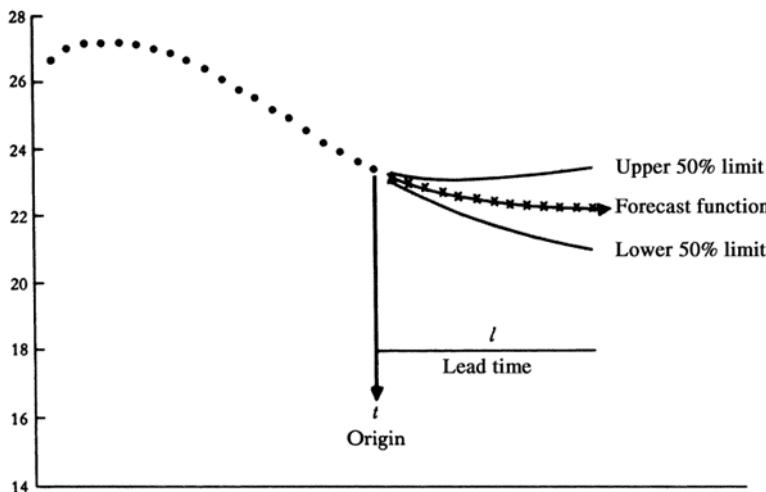


Figura 4 – Previsão calculada com intervalo de confiança de 50%.

Fonte: George E. P. Box (2016).

Da figura acima, observa-se que a distância entre os limites inferior e superior cresce à medida que o horizonte de previsão aumenta, em outras palavras, quanto maior o horizonte de previsão, menor será a confiança no valor previsto. Esta é uma característica recorrente nos métodos e modelos de previsão.

1.1.5 – TRABALHOS RELACIONADOS

Nesta seção, são apresentados alguns trabalhos relacionados à previsão de séries temporais de velocidade do vento envolvendo diversas abordagens e estratégias. Alavancada pelo advento de geração elétrica a partir de fontes eólicas, a literatura apresenta uma vultosa quantidade de estudos tratando o tema. No entanto, um minguado número de publicações lida com séries temporais intradiárias e nenhum trabalho foi encontrado envolvendo esse tipo de

série temporal oriunda de um cenário geográfico tão complexo como o considerado neste trabalho.

Chuanjin Yu *et al.* (2017) argumentam que para aumentar a precisão da previsão da velocidade do vento, a Transformada Wavelet (WT) é amplamente empregada para desagregar uma série original de velocidade do vento em várias subséries antes da previsão. No entanto, a subsérie de maior frequência costuma ter uma grande perturbação na previsão final. No estudo, para aumentar a precisão da previsão, o Singular Spectrum Analysis (SSA) é aplicada para fazer processamento adicional na subsérie de frequência mais alta, em vez de não fazer nenhuma modificação ou eliminá-la. Assim, uma tecnologia de decomposição híbrida chamada Improved WT (IWT) é proposta. Enquanto isso, um novo modelo híbrido IWT-ENN combinado com IWT e Elman Neural Network (ENN) também é projetado. O procedimento do IWT é sistematicamente investigado. Os resultados experimentais mostram que: (1) o desempenho do modelo híbrido IWT-ENN tem uma grande melhoria em comparação com outros, incluindo o método de persistência, ENN, modelo Auto-Regressivo (AR), Back Propagation Rede Neural (BPNN) e Decomposição de Modo Empírico (EMD)-ENN; (2) em comparação com as duas estratégias gerais onde a subsérie de frequência mais alta é sem retratamento ou eliminada, o novo modelo híbrido proposto IWT-ENN tem o melhor desempenho de previsão.

G. W. Chang *et al.* (2017) apresentam um modelo baseado em rede neural de função de base radial aprimorado com um esquema de feedback de erro (IRBFNN-EF) para prever a velocidade do vento e a potência de curto prazo de um parque eólico, onde um fator de forma adicional é incluído na função de base Gaussiana clássica associado a cada neurônio na camada oculta e um método simples de inicialização de parâmetros é proposto para encontrar efetivamente os valores iniciais de dois parâmetros-chave da função de base ao realizar o treinamento da rede neural. Um parque eólico perto da área central de Taiwan conectado ao sistema Taipower é servido como alvo de medição. Fornecido com 24 h de dados de entrada em resolução de 10 min (ou seja, 6×24 etapas de tempo de entrada) para treinar a rede neural proposta, um tempo de antecipação de até 72 h (ou seja, 6×72 etapas de tempo de saída previstas) foi realizado. Casos de teste para diferentes meses ao longo de 2014 são relatados. Os resultados obtidos pelo modelo proposto são comparados com os obtidos por outros quatro métodos de previsão baseados em redes neurais artificiais. Isso mostra que o modelo proposto

leva a uma melhor precisão para prever a velocidade do vento e a potência do vento enquanto a eficiência computacional é mantida.

Henrique do Nascimento Camelo *et al.* (2016) mostram um estudo de previsão da média mensal de velocidade do vento para a cidade de Caucaia, litoral do Estado do Ceará, no nordeste do Brasil. Para isso, foi utilizado o modelo aditivo de Holt-Winters a partir dos dados de velocidade do vento ao nível de 10 m de altura no período de janeiro de 2004 até dezembro de 2005, coletados em torre meteorológica. A previsão mostrou-se eficiente, baseada no fato de que a análise de erro realizada forneceu valores (em geral) reduzidos como, por exemplo, existe um erro percentual médio da previsão comparado ao observado de aproximadamente 3,5%. Dessa forma, foi possível identificar que os dados previstos conseguem acompanhar o perfil dos dados observados, indicando, assim, viabilidade do modelo para a região de estudo. Esse trabalho poderá ser utilizado como uma ferramenta de previsão de velocidade do vento, visando o estudo e avanço da geração eólio-elétrica em diversas regiões do país, proporcionando aos tomadores de decisão a instalação de novos parques eólicos, visto que haverá possibilidade de conhecer o regime local de intensidade de vento futuro.

Hui Liu *et al.* (2012) propõem neste artigo um modelo híbrido denominado EMD-ANN para previsão da velocidade do vento baseado na Decomposição de Modo Empírico (EMD) e nas Redes Neurais Artificiais (ANNs). Para escolher o melhor algoritmo de treinamento para o modelo de ANN, várias simulações experimentais com diferentes algoritmos de treinamento são feitas. Para estimar o desempenho do modelo EMD-ANN, dois casos de previsão são concluídos e os resultados são comparados com o modelo ANN e o modelo ARIMA, respectivamente. Para evitar a aleatoriedade causada pelo modelo ANN ou a parte ANN do modelo híbrido EMD-ANN, todas as simulações neste estudo são repetidas pelo menos 30 vezes para obter a média. Os resultados mostram que: (1) o desempenho do modelo proposto é altamente satisfatório; e (2) o método híbrido EMD-ANN proposto é robusto para lidar com amostragens de saltos em séries de vento não estacionárias.

Hui Liu *et al.* (2013) informam que com base nas teorias de Wavelet, análise clássica de séries temporais, algoritmo genético, otimização de enxame de partículas e redes neurais artificiais, duas estruturas de previsão híbridas [o Wavelet-Genetic Algorithm (GA)-Multilayer Perceptron (MLP) e o Wavelet-Particle Swarm Optimization (PSO)-Multilayer Perceptron (MLP)] são

propostas para prever velocidades de vento não estacionárias. Comparações de desempenho de previsão usando diferentes combinações de algoritmos são fornecidas para investigar a contribuição de diferentes componentes nessas duas estruturas híbridas. Os resultados baseados em três casos experimentais mostram que: (1) ambas as estruturas de previsão híbridas propostas são adequadas para os diversos requisitos de precisão nas previsões de velocidade do vento, que podem ser aplicadas a sistemas de energia eólica; e (2) em ambas as estruturas híbridas, a contribuição do AG e componentes PSO na melhoria do MLP não são estatisticamente significativos, enquanto o do componente Wavelet é estatisticamente significativo.

Ina Khandel Wal *et al.* (2015) apresentam a vantagem do Discrete Wavelet Transform (DWT) para melhorar a precisão da previsão de séries temporais. Este artigo sugere uma nova técnica de previsão, segregando um conjunto de dados de séries temporais em componentes lineares e não lineares através do DWT. No início, o DWT é usado para decompor o conjunto de dados de treinamento na amostra das séries temporais em peças lineares (detalhadas) e não lineares (aproximadas). Em seguida, os modelos de média móvel integrada autorregressiva (ARIMA) e rede neural artificial (RNA) são usados para reconhecer e prever separadamente os componentes detalhados e aproximados reconstruídos, respectivamente. Dessa maneira, a abordagem proposta utiliza taticamente os pontos fortes exclusivos de DWT, ARIMA e ANN para melhorar a precisão da previsão. Este método híbrido é testado em quatro séries temporais do mundo real e seus resultados de previsão são comparados com os dos modelos híbridos ARIMA, ANN e ZHANG. Os resultados mostram claramente que o método proposto atinge a melhor precisão de previsão para cada série.

Jianming Hu *et al.* (2015) sugerem uma abordagem de previsão híbrida que consiste em EWT (Empirical Wavelet Transform), CSA (Coupled Simulated Annealing) e LSSVM (Least Square Support Vector Machine) para melhorar a precisão da previsão da velocidade do vento a curto prazo. O modelo híbrido proposto é demonstrado para prever uma série média de velocidade do vento de meia hora obtida de um parque eólico localizado no noroeste da China.

Ping Jiang *et al.* (2014) mostram um modelo híbrido inteligente para previsão de velocidade do vento a curto prazo. O modelo é baseado na análise de correlação cruzada (CC) e um modelo de regressão de vetor de suporte (SVR) que é acoplado com algoritmos de otimização de

brainstorming (BSO) e pesquisa de cuco (CS), que são utilizados com sucesso para determinação de parâmetros. Os modelos híbridos propostos foram usados para prever velocidades de vento de curto prazo coletadas de quatro turbinas eólicas localizadas em um parque eólico na China. Os resultados de previsão demonstram que os modelos híbridos inteligentes superam modelos individuais para previsão de velocidade de vento de curto prazo, o que resulta principalmente da superioridade de BSO e CS para a otimização de parâmetros.

Reginaldo Nunes da Silva *et al.* (2022) informam sobre a implementação de um modelo de Rede Neural Recorrente multivariada do tipo memória de curto longo prazo (MV-LSTM) para dados de velocidade média, máxima e mínima horária. Uma rede LSTM com única variável também é implementada para comparar os resultados. Os dados utilizados foram coletados por um anemômetro sônico no período de 01 de janeiro a 31 de dezembro de 2015 no cerrado nativo brasileiro na Fazenda Água Limpa – FAL localizada no Distrito Federal. Os resultados mostram que os vetores de velocidade máxima e mínima melhoram a previsão de velocidade do vento e que o modelo MV-LSTM reduz o atraso da previsão quando há uma mudança brusca da velocidade do vento.

Tiago Mendes Dantas *et al.* (2014) apresentam uma abordagem de previsão de Séries Temporais que combina técnicas de Bagging, decomposição STL e o método clássico de Amortecimento Exponencial de Holt-Winters. Os resultados indicam que o método proposto foi capaz de prever com significativa habilidade, superando o desempenho da abordagem de Holt-Winters tradicional e o método Neuro-Fuzzy.

Tong Niu *et al.* (2018) observam que com a chegada do “big data”, a análise de mineração de dados e a previsão de alto desempenho da velocidade do vento estão atraindo cada vez mais atenção. Apesar do fato de que investigações massivas sobre previsão de velocidade do vento na teoria e na prática têm sido conduzidas por vários pesquisadores, estudos sobre previsão de vários passos à frente ainda estão faltando, impedindo o desenvolvimento no campo. Neste estudo, uma nova abordagem híbrida é proposta para a previsão da velocidade do vento com vários passos à frente, utilizando a seleção de recursos ideais e uma rede neural artificial otimizada por um algoritmo de morcego modificado com estratégia de cognição. O modelo híbrido proposto pode remediar em grande parte as deficiências das redes neurais para a previsão multi-etapa à frente, que é validado para diferentes horizontes de previsão e funciona

de forma eficaz. Finalmente, experimentos baseados em três unidades de verificação da cidade de Penglai na China são conduzidos de forma eficaz, ilustrando que o modelo proposto não só tem vantagens quando comparado com modelos de referência, mas também tem grande potencial para aplicação ao sistema de energia eólica.

Yu Jiang et al. (2013) propõem um novo método de séries temporais para a previsão da velocidade do vento em curtíssimo prazo. O modelo de previsão de séries temporais é baseado na teoria bayesiana e na modelagem de quebra estrutural (que indica uma mudança inesperada ao longo do tempo nos parâmetros dos modelos de regressão, o que pode levar a grandes erros de previsão e falta de confiabilidade do modelo em geral), que pode incorporar o conhecimento de domínio sobre a velocidade do vento a priori. Além disso, o modelo de quebra estrutural bayesiana prevê a velocidade do vento como um conjunto de valores possíveis, que é diferente da previsão do modelo clássico de séries temporais univariadas. Este conjunto de valores previstos pode ser usado para várias aplicações, como controle preditivo de turbinas eólicas e programação de energia eólica. O modelo proposto é testado com dados reais de velocidade do vento coletados de turbinas eólicas em escala de utilidade.

Zhenkun Liu et al. (2020) argumentam que muitos métodos de previsão têm sido amplamente empregados para prever a velocidade do vento de curto prazo para ventos irregulares, não lineares e não estacionários. No entanto, eles negligenciam a eficácia do pré-processamento de dados e da otimização dos parâmetros do modelo, representando um enorme desafio para a previsão precisa e estável da velocidade do vento e a operação segura da indústria de energia eólica. Para superar esses desafios e melhorar ainda mais o desempenho e a estabilidade da previsão da velocidade do vento, um sistema de previsão é desenvolvido com base em uma estratégia de pré-tratamento de dados, um algoritmo de otimização multiobjetivo modificado e vários modelos de previsão. Mais especificamente, uma estratégia de pré-tratamento de dados é executada para determinar a tendência dominante de uma série de velocidade do vento e controlar a interferência do ruído. O algoritmo de otimização multiobjetivo pode ajudar a adquirir precisão e estabilidade de previsão mais satisfatórias. Os múltiplos modelos de previsão são integrados para construir um modelo combinado para previsão da velocidade do vento. Para verificar as propriedades do sistema de previsão desenvolvido, dados de velocidade do vento de 10 minutos de 4 parques eólicos adjacentes na Península de Shandong, China são adotados como estudos de caso. Os resultados da previsão pontual e da previsão por intervalo

revelam que nosso sistema de previsão supera positivamente todos os modelos contrastivos em relação à precisão e estabilidade da previsão. Assim, nosso sistema desenvolvido é extremamente útil para aumentar a precisão da previsão e é uma ferramenta razoável e válida para uma programação inteligente da rede elétrica.

1.2 – JUSTIFICATIVA

No cenário de um acidente nuclear com liberação de efluentes radioativos para o meio ambiente, conhecer com antecedência as condições meteorológicas pode ser um grande diferencial entre sucesso e fracasso do planejamento e execução da resposta à emergência nuclear, especialmente em áreas populosas. Nesses termos, há que buscar a melhor relação entre antecedência e acurácia da previsão.

Influenciada por diversos fatores ambientais, a velocidade do vento apresenta características de altas flutuações, não linearidade, autocorrelação e volatilidade estocástica inerente. Então, devido à complexidade caótica e intrínseca dos parâmetros climáticos, a utilização de modelos híbridos para a previsão de velocidade do vento é sempre mais adequada que os modelos únicos.

1.3 – OBJETIVO

Agregar maior precisão ao método de previsão utilizado pelo SCA para períodos de uma e duas horas à frente, de modo a mitigar a vulnerabilidade desse método diante de um valor futuro consideravelmente destoante em relação ao último valor conhecido.

Tal objetivo foi alcançado a partir de um estudo comparativo das principais técnicas de previsão disponíveis na teoria de análise de séries temporais, culminando com o desenvolvimento de um modelo híbrido, que utiliza um método matemático robusto o suficiente para lidar adequadamente com as características particulares das séries temporais intradiárias de velocidade do vento obtidas de uma torre meteorológica localizada num complexo cenário geográfico, como aquele onde se encontra a CNAAA. Uma vez estabelecido o modelo para

uma torre meteorológica, esse pode ser expandido para os outros dados coletados por essa torre, assim como para as demais torres pertencentes à CNAAA.

A Figura 5 mostra a pluma radioativa liberada para o meio ambiente através das simulações de um acidente do tipo Grande Loca e de determinadas condições meteorológicas. As Figuras 6 e 7 representam a evolução prevista dessa pluma após o período de uma e duas horas, respectivamente.

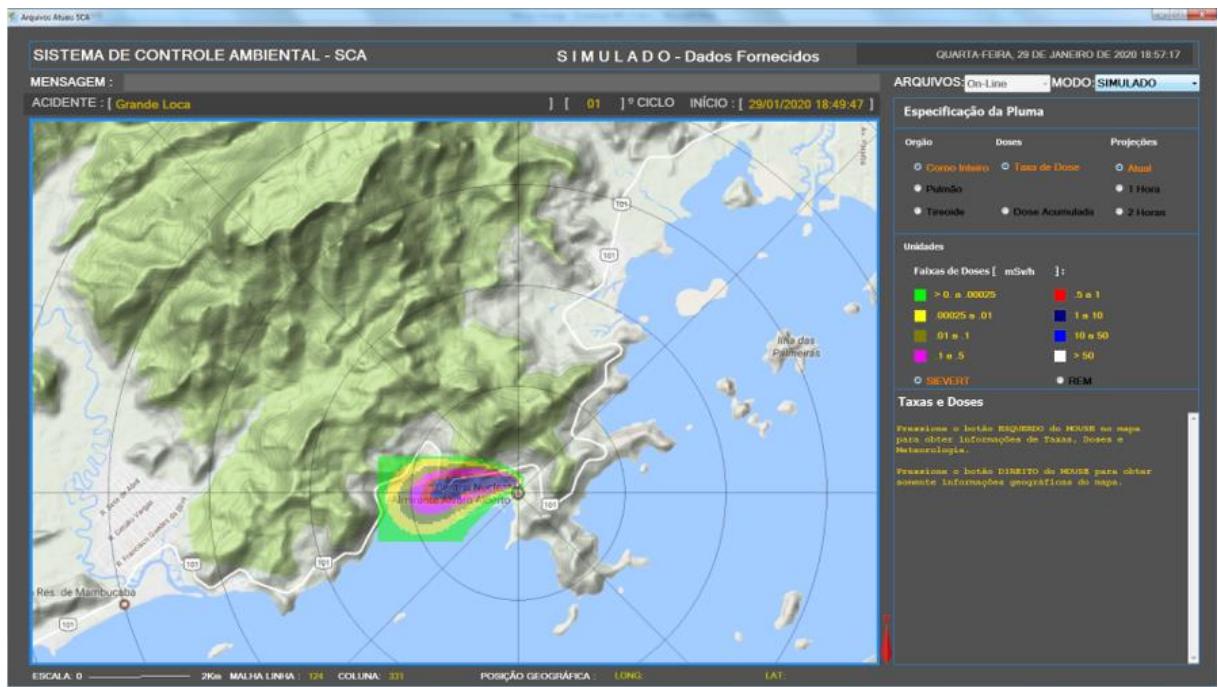


Figura 5 – Pluma radioativa gerada durante um Grande Loca.

Fonte: Laboratório de Monitoração de Processos - LMP (2022).

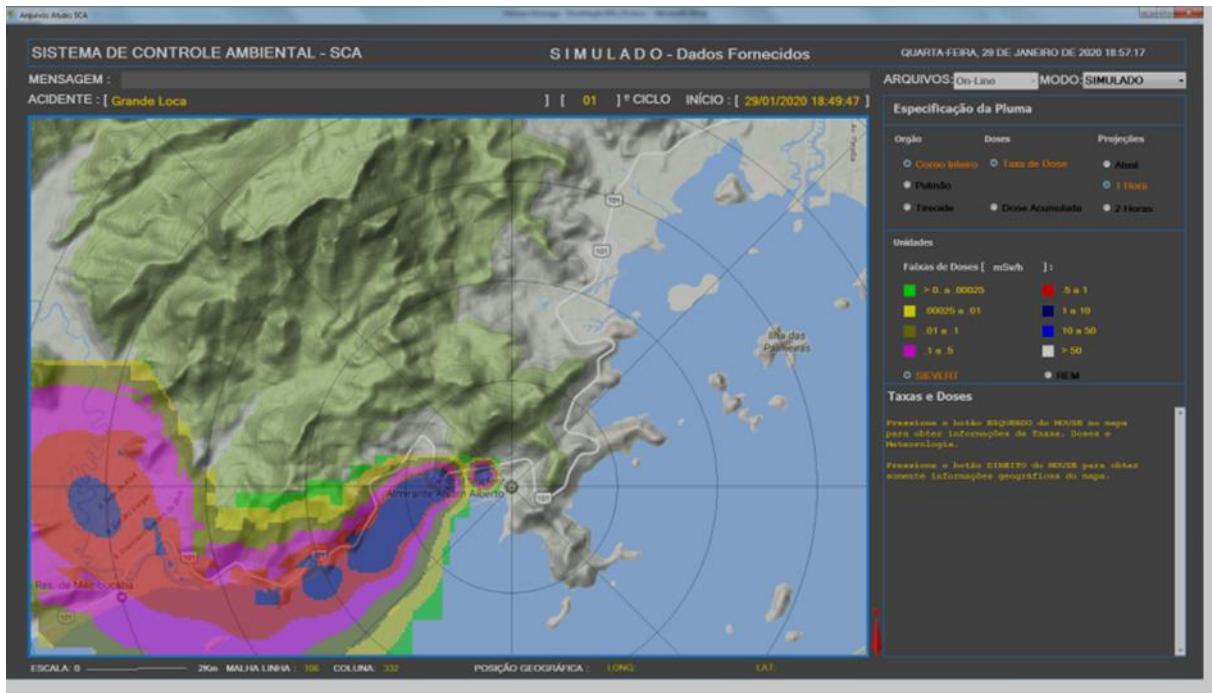


Figura 6 – Previsão do deslocamento da pluma radioativa após uma hora.

Fonte: Laboratório de Monitoração de Processos - LMP (2022).

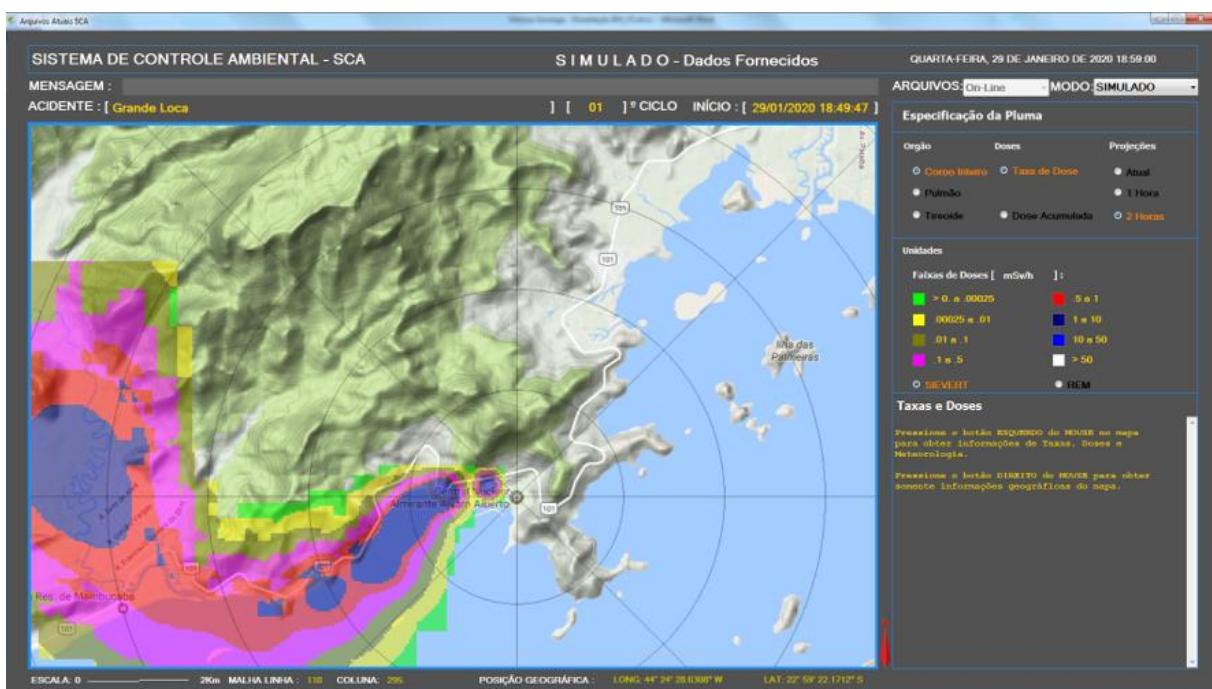


Figura 7 – Previsão do deslocamento da pluma radioativa após duas horas.

Fonte: Laboratório de Monitoração de Processos - LMP (2022).

2 – FUNDAMENTAÇÃO TEÓRICA

2.1 – ESTRUTURA E COMPONENTES DE UMA SÉRIE TEMPORAL

Uma série temporal caracteriza-se pelo registro de um conjunto de valores relacionados a um ou mais eventos e, em grande parte, registrados em intervalos de tempo constante. O intervalo de tempo entre duas medidas consecutivas registradas na série temporal determina a sua periodicidade. Uma característica intrínseca de uma série temporal é que normalmente as observações adjacentes são dependentes. A análise de séries temporais é uma técnica em estatística que lida com dados de séries temporais e análise de tendências, visando ao entendimento da dependência de seus elementos em relação ao tempo com o objetivo de realizar uma previsão de valores futuros. Assim, após o estabelecimento de uma série temporal $\{Y_1 \dots Y_t\}$ normalmente estamos interessados em prever um valor futuro Y_{t+l} , pois tal previsão pode fornecer uma base para diversos tipos de planejamentos.

Uma série temporal composta por valores relacionados a uma única variável (um único evento) é chamada de série temporal univariada e aquela composta por um conjunto de valores relacionados a mais de uma variável (mais de um evento) é denominada de série temporal multivariada. Alguns problemas de interesse podem apresentar séries temporais compostas por mais de uma variável com determinado grau de relação entre si. Frequentemente é possível realizar uma análise mais efetiva, considerando séries univariadas individuais como componentes de uma série temporal multivariada e analisar as séries conjuntamente. O objetivo dessa estratégia consiste em obter uma compreensão das relações dinâmicas ao longo do tempo entre as séries e melhorar a precisão das previsões para séries individuais.

2.2 – REPRESENTAÇÃO GRÁFICA DE UMA SÉRIE TEMPORAL

A representação gráfica de uma série temporal costuma ser a primeira abordagem na análise de séries temporais, pois normalmente oferece um panorama geral das características mais relevantes, tais como tendência, sazonalidade, variância etc. A figura abaixo mostra a representação gráfica da série temporal *AirPassengers* (disponibilizada pelo *software* estatístico R), que registra os totais mensais de passageiros entre os anos 1949 e 1960 da extinta *Pan American World Airways* (Pan Am).

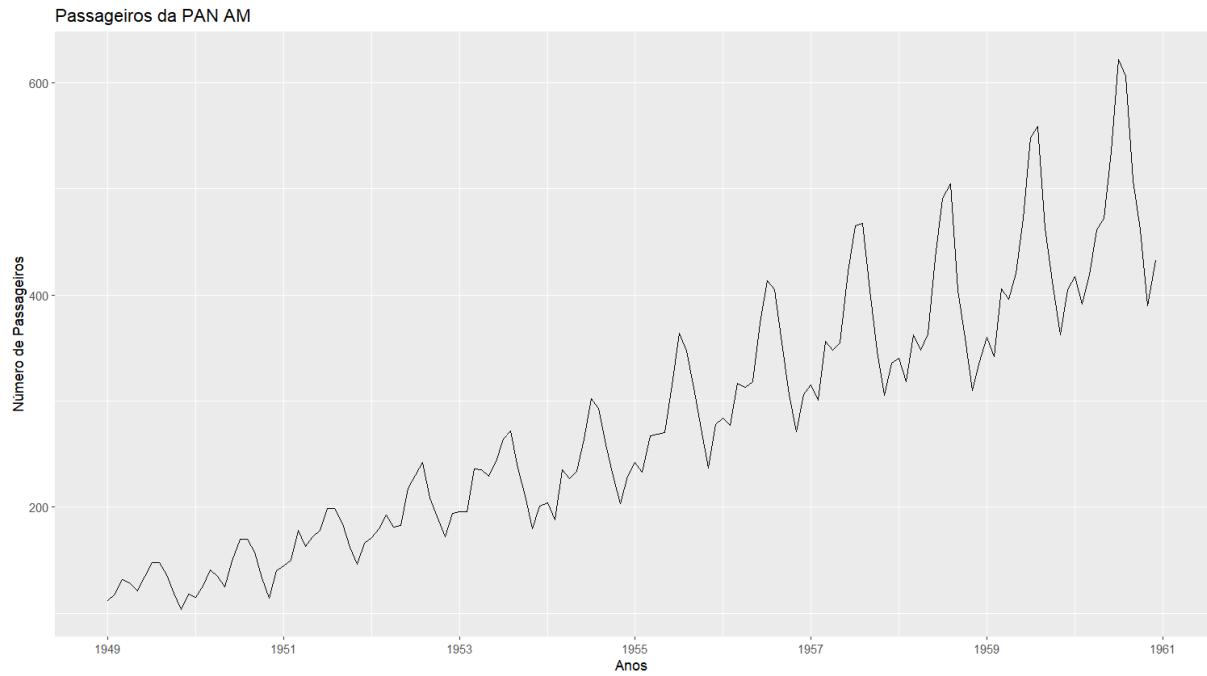


Figura 8 – Totais mensais de passageiros aéreos da PAN AM entre os anos 1949 e 1960.

Observando-se o gráfico acima, fica evidente a presença de uma tendência de crescimento do número de passageiros, assim como os períodos sazonais de alta, que provavelmente se referem a períodos de férias.

Considera-se que uma série temporal é composta pela aglutinação de suas componentes intrínsecas, a saber, tendência, sazonalidade, ciclo e irregular, podendo ocorrer de forma aditiva ou multiplicativa. O objetivo do estudo de análise de séries temporais, visando uma predição, consiste em isolar cada componente.

2.3 – ANÁLISE DOS *OUTLIERS*

Por definição *outliers* são pontos pertencentes a uma série temporal que apresentam valores discrepantes, isto é, que se afastam ou “destoam” do padrão estatístico esperado. Os *outliers* podem advir de algum tipo de erro ou mesmo de situações reais atípicas. Em estimativas estatísticas é habitual eliminar os *outliers* para se evitar o viés sobre as estimativas. A presença de *outliers* pode distorcer significativamente análises baseadas em médias ou desvio padrão e levar a um viés, um erro de estimação, daí a necessidade de seu tratamento. A medida de acuracidade MPE (*Mean Percentage Error*) é uma medida de viés, que mostra o quanto a

previsão possui um viés em relação aos dados observados. Assim, se $MPE > 0$ indica que a previsão está superestimada em relação aos dados observados; se $MPE < 0$, significa que a previsão está subestimada e, finalmente, se $MPE = 0$ indica a ausência de viés.

De modo geral, uma justificativa estatística é o fator que deve nortear o procedimento de eliminação ou substituição de *outliers*. A obtenção dos *outliers* neste trabalho utilizou a técnica conhecida como *Boxplot*.

2.3.1 – *BOXPLOT*

O emprego desta técnica não depende da simetria ou distribuição dos dados. Tradicionalmente, um *boxplot* apresenta medida de tendência central não-paramétrica (mediana); medida de dispersão (quartis); medida de forma de distribuição ou simetria da amostra (valores pontuais mínimo e máximo). Assim, são considerados *outliers* os valores que se encontram acima do limite superior ou abaixo do inferior. A figura abaixo exibe um *boxplot* de disposição vertical.

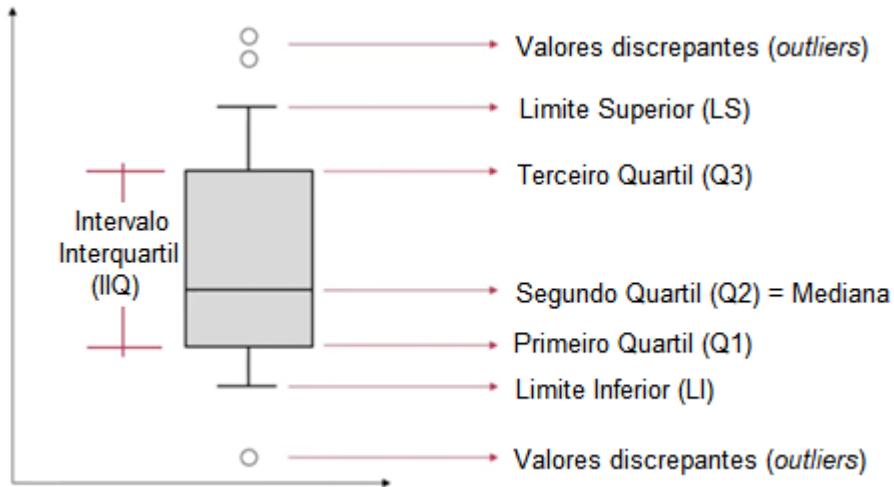


Figura 9 – Ilustração da estrutura básica de um *boxplot* vertical.

- O valor de cada quartil é obtido a partir da posição que esse ocupa na distribuição, a qual é determinada através da seguinte equação:

$$PosQ_i = \frac{i \cdot (n+1)}{4}, \quad i = 1, 2 \text{ e } 3$$

onde PosQi é a posição do quartil i e n o número de elementos da série.

- Intervalo Interquartil (IIQ) = $Q3 - Q1$
- Limite Superior (LS) = $Q3 + 1,5 * \text{IIQ}$
- Limite Inferior (LI) = $Q1 - 1,5 * \text{IIQ}$

Os gráficos abaixo apresentam uma visualização de alguns efeitos provocados por *outliers* numa série temporal. Trata-se de uma das séries temporais utilizadas neste trabalho (WS_m), sobre a qual foi realizado um tratamento de *outlier* utilizando-se a técnica acima. Neste caso, os efeitos visíveis se referem à distribuição dos valores da velocidade do vento propriamente dita, à dispersão desses valores e aos valores da função de autocorrelação, antes e depois do tratamento de *outliers*.

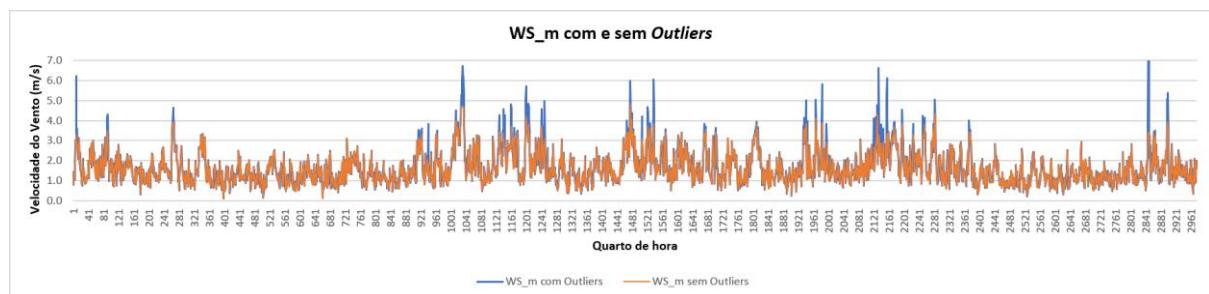


Figura 10 – Distribuição das velocidades do vento com e sem o tratamento de *outliers*.

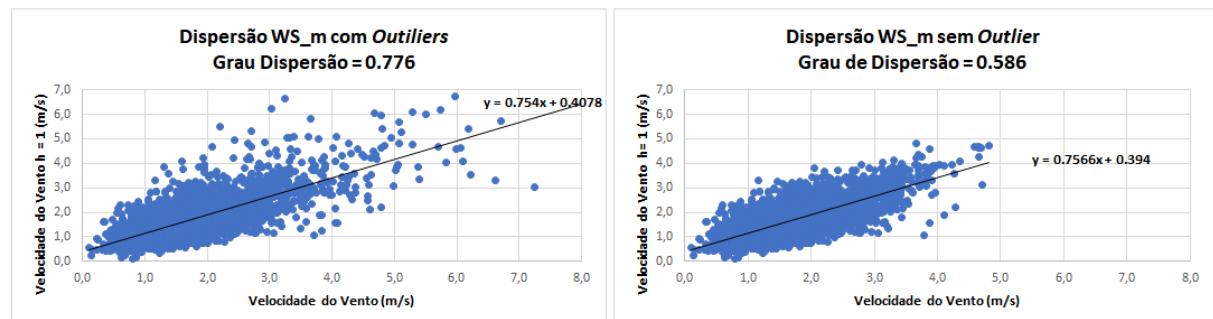


Figura 11 – Dispersão das velocidades do vento com e sem o tratamento de *outliers*.

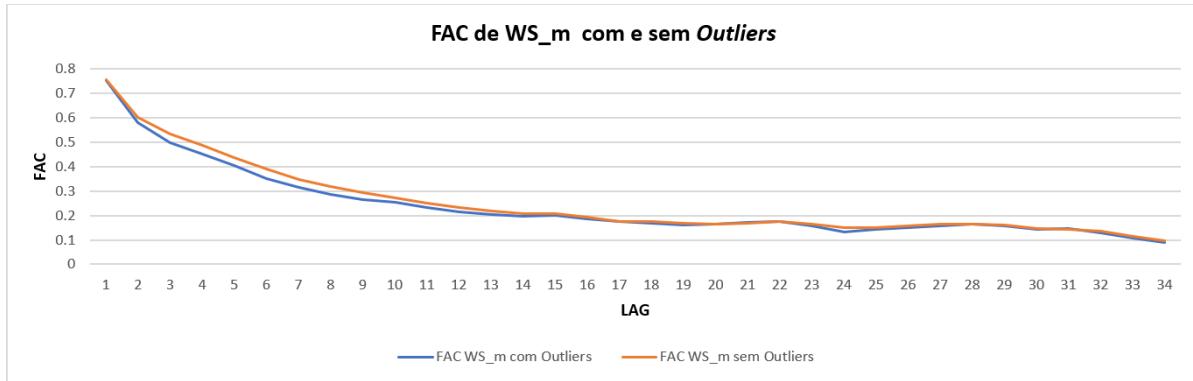


Figura 12 – Função de autocorrelação das velocidades do vento com e sem *outliers*.

2.4 – DIFERENCIAÇÃO

É uma metodologia simples de remoção de tendência e/ou efeitos sazonais. Uma série diferenciada é aquela que resulta da subtração entre observações consecutivas da série que lhe deu origem. Tanto no modelo aditivo quanto no multiplicativo, a diferenciação resulta no surgimento de uma nova série menor que a série original.

No modelo aditivo, a diferenciação tem o efeito de monitorar as mudanças ocorridas sobre a série original. Cada observação gerada na nova série diferenciada será o resultado de uma subtração entre duas observações da série original, a saber, a observação ocorrida num tempo posterior n e a observação no tempo atual. Este procedimento acontece de forma consecutiva até a última observação da série original. Assim, a nova série Y'_t pode ser definida como:

$$Y'_t = Y_{t+n} - Y_t \quad \text{onde: } n \text{ corresponde à quantidade desejada de observações à frente.}$$

No modelo multiplicativo, a diferenciação é aplicada sobre a função transformada pelo logaritmo. Isso é chamado de *log return*. Em outras palavras, o *log return* consiste em primeiro se aplicar o logaritmo sobre a série original e a seguir efetuar a subtração desejada, nessa ordem. A ordem da operação é importante porque ao aplicar o logaritmo está se tentando tornar a distribuição normal.

$$Y'_t = \log(Y_{t+n}) - \log(Y_t) \approx (Y_{t+n} - Y_t) / Y_t$$

A diferenciação pode ajudar a estabilizar a média de uma série temporal, removendo suas mudanças no nível e, portanto, eliminando (ou reduzindo) a tendência e a sazonalidade. Já uma transformação do tipo logarítmica, pode ajudar a estabilizar a variância da série temporal.

Quando a série diferenciada se torna um ruído branco, é normalmente chamada de passeio aleatório e o modelo pode ser escrito como $Y_{t+n} - Y_t = \varepsilon_t$, onde ε_t denota um ruído branco. Os modelos de passeio aleatório são amplamente utilizados para dados não estacionários, principalmente dados financeiros e econômicos. Passeios aleatórios normalmente têm:

- longos períodos de tendências aparentes para cima ou para baixo;
- mudanças repentinhas e imprevisíveis de direção.

As previsões de um modelo de passeio aleatório são iguais à última observação, pois os movimentos futuros são imprevisíveis e têm a mesma probabilidade de serem para cima ou para baixo. Assim, o modelo de caminho aleatório sustenta previsões ingênuas.

Aqui cabe observar a diferença entre diferenciação e defasagem. Ambas geram uma nova série com o número de observações inferior à original. As observações geradas pela diferenciação se originam de uma operação matemática de subtração e as observações geradas pela defasagem são idênticas às observações da série original, porém defasadas.

2.5 – TRANSFORMAÇÃO LOGARÍTMICA

Consiste em uma operação logarítmica (log neperiano ou normal) realizada sobre cada observação de uma série temporal. A aplicação deste tipo de transformação logarítmica é uma *tentativa* de transformar a distribuição original da série numa distribuição normal. Assim, trata-se de uma técnica largamente utilizada, com a finalidade de estabilizar a variância de uma série temporal.

2.6 – TRANSFORMAÇÃO BOX-COX

A transformação BOX-COX envolve logaritmo (natural) e transformação de potência. Sua dependência com o valor de λ é mostrada abaixo:

$$w_t = \begin{cases} \ln(y_t) & \text{Se } \lambda = 0; \\ \frac{(y_t^\lambda - 1)}{\lambda} & \text{senão.} \end{cases}$$

Se $\lambda = 0$, o logaritmo natural é usado e se $\lambda \neq 0$ uma transformação de potência é utilizada. Se $\lambda = 1$, então $w_t = y_t - 1$ e assim sendo os dados transformados são deslocados para baixo, sem influenciar na forma da série. Para os demais valores de λ haverá mudança da forma da série temporal.

2.7 – TESTES DE RAIZ UNITÁRIA

Uma maneira de determinar mais objetivamente se a diferenciação se faz necessária é usar um teste de raiz unitária. Estes são testes de hipóteses estatística de estacionariedade que são projetados para determinar se a diferenciação é necessária. Existem diversos testes de raiz assumindo diferentes hipóteses e suposições. Contudo há que ter cuidado ao empregar tais testes, isto é, deve-se analisar as condições correntes, pois segundo Hyndman *et al.* (2018) podem levar a respostas conflitantes.

2.8 – ESTACIONARIEDADE

Por definição, uma série temporal é considerada estacionária se suas propriedades estatísticas se mantêm razoavelmente estáveis ao longo do tempo, sobretudo no que diz respeito à média, variância e covariância. Resumidamente, pode-se dizer que uma série temporal estacionária é aquela cujas propriedades não dependem do tempo em que a série é observada.

Mais precisamente, se Y_t é uma série temporal estacionária no tempo, então para todos os termos s , a distribuição de Y_t, \dots, Y_{t+s} não depende de t .

Sem a estacionariedade não é possível apreender-se estatisticamente o que a série tem a dizer. Algumas técnicas estatísticas de previsão pressupõem tal característica, viabilizando assim a possibilidade de bons resultados preditivos. Caso contrário, pode ocorrer regressões espúrias ou resultados não confiáveis.

Em princípio, séries com tendência e sazonalidade não são estacionárias, mas eventualmente podem ser. A avaliação de estacionariedade pode ser realizada através de inspeção gráfica da série temporal e suas FAC e FACP, mas tal prática pode também não se apresentar explícita o suficiente; portanto, uma avaliação puramente gráfica pode levar a interpretações equivocadas, daí a necessidade dos testes estatísticos de estacionariedade. Nesse sentido vale registrar o comentário de Hyndman *et al.* (2018) sobre tal dificuldade: “Alguns casos podem ser confusos - uma série temporal com comportamento cíclico (mas sem tendência ou sazonalidade) é estacionária. Isso ocorre porque os ciclos não têm comprimento fixo, portanto, antes de observarmos a série, não podemos ter certeza de onde estarão os picos e vales dos ciclos.”

Conforme Nielsen (2021) alerta, “A estacionariedade é um conceito enganoso, principalmente quando aplicado a dados de séries temporais reais. ...é interessante observar que, mesmo a série não sendo estacionária, isso não significa que ela deva ter uma tendência. Um passeio aleatório é um bom exemplo de uma série temporal não estacionária que não tem uma tendência subjacente. ...supostamente, quando se trata de um determinado processo de série temporal amostrado por um passeio aleatório, vemos uma tendência. Isso motiva debates acalorados, ainda mais nas análises de séries temporais de preços de ações.”

Processos de diferenciação, entre outros, são técnicas para transformar uma série não estacionária em estacionária. Aqui vale lembrar que o modelo autorregressivo ARIMA possui potencial de realizar internamente processos de diferenciação (indicado pelo parâmetro d).

2.8.1 – TESTE DE ESTACIONARIEDADE DE RAIZ UNITÁRIA

Os testes para determinação de estacionariedade de uma série temporal são chamados de testes de hipóteses. Neste trabalho foi adotado o teste de raiz unitária de Dickey-Fuller Aumentado (ADF), que atualmente é um dos mais utilizados pela literatura, e os resultados obtidos foram confirmados através do teste de Kwiatkowski-Phillips-Schmidt-Shin (KPSS), que ultimamente também vem ganhando grande destaque.

O teste de Dickey-Fuller é um teste de significância estatística, isto é, envolve teste de hipóteses com uma hipótese nula (H_0) e uma hipótese alternativa (H_a). Ao final do teste é gerado um valor estatístico, que indicará a aceitação ou rejeição de cada uma das hipóteses. O teste de Dickey-Fuller (padrão) parte de um processo autorregressivo de primeira ordem:

$$Y_t = \mu + \alpha Y_{t-1} + \varepsilon_t$$

Se α for igual a 1 (um) a série é dita não-estacionária e o teste de raiz unitária consiste nessa verificação.

A hipótese nula (H_0) do teste de Dickey-Fuller (padrão e aumentado) parte do princípio da existência de uma raiz unitária e quando isso ocorre é dito, na terminologia estatística, não haver motivos para rejeitar a hipótese nula ou, em outras palavras, a série é dita não-estacionária e implica que esta possui alguma estrutura dependente do tempo e com variância não constante no tempo.

H_0 : $\alpha = 1$; possui raiz unitária; a série é não-estacionária em nível.

A hipótese alternativa (H_a) do teste de Dickey-Fuller (padrão e aumentado) é o oposto da hipótese nula (H_0), isto é, não possui raiz unitária e, portanto, a série é dita estacionária. Então, se a hipótese nula (H_0) for rejeitada, a série é dita estacionária em nível e também equivale dizer que a série é integrada de ordem zero ou $I(0)$, indicando que não é necessário um processo de diferenciação para torná-la estacionária, pois essa já é.

H_a : não possui raiz unitária; a série é estacionária em nível.

O teste Dickey-Fuller (padrão) é válido para modelos do tipo AR(1), ou seja, autorregressivo de primeira ordem, mas isso lhe confere um caráter um tanto restrito. Então para modelos autorregressivos de ordem superior a 1 (um), utilizamos o teste Dickey-Fuller Aumentado (ADF), que possui maior número de termos, de modo a contemplar a ordem p do modelo AR(p). O teste de Dickey-Fuller Aumentado pode ser executado considerando-se três alternativas:

- 1 - Sem intercepto e sem tendência;
- 2 - Com intercepto e sem tendência;
- 3 - Com intercepto e com tendência.

Além disso, as variáveis explicativas e o número de *lags* (defasagens) a serem consideradas no teste também podem ser definidos pelo usuário. Se o número de *lags* definido for igual a 0 (zero), trata-se do teste padrão (DF) e se maior que 0 (zero) do teste aumentado (ADF). O número de *lags* a ser definido para o teste deve ser suficiente para deixar os resíduos da regressão do teste sem qualquer autocorrelação serial. Assim, diante da percepção que os resíduos não são do tipo ruído branco, o número de *lags* deverá ser aumentado para deixar os resíduos sem autocorrelação.

No teste KPSS as definições de suas hipóteses possuem significados opostos ao teste DF e DFA. A hipótese nula do teste KPSS (H_0) assume que a série não possui raiz unitária, tratando-se então de uma série estacionária. Já a hipótese alternativa (H_a) presume que a série possui raiz unitária, sendo então não-estacionária.

2.9 – ANÁLISE DESCRIPTIVA DE SÉRIES TEMPORAIS

No processo de análise descritiva de séries temporais, procura-se entender o contexto do problema e da fonte de dados, a fim de realizar-se as transformações que se façam necessárias, por exemplo, filtragens, redução da assimetria dos dados e etc. A seguir, dá-se a verificação da estacionariedade da série, pois as propriedades estatísticas se aplicam a séries com essa característica. Caso se constate a não estacionariedade, aplica-se então as

transformações cabíveis para tentar torná-la estacionária. Uma vez alcançada a estacionariedade, passa-se à análise das funções de autocorrelação.

2.10 – DISPERSÃO, VARIÂNCIA, DESVIO PADRÃO, COVARIÂNCIA, CORRELAÇÃO, AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL

2.10.1 – DISPERSÃO

Entende-se por dispersão estatística (variabilidade ou espalhamento) o quanto uma distribuição pode se apresentar espalhada ou condensada. A dispersão pode ser medida através da variância, desvio padrão, covariância ou da amplitude interquartil.

2.10.2 – VARIÂNCIA

Dado um conjunto de dados, a variância mostra o quanto distante cada valor desse conjunto está da média. Quanto menor for a variância, mais próximos os valores do conjunto estarão da média e quanto maior a variância, mais distantes estarão da média. A variância é definida como a média do quadrado da distância de cada ponto até a média, ou seja, o desvio. A variância da variável aleatória de Y é geralmente designada por $\text{var}(Y)$ ou σ_Y^2 ou simplesmente σ^2 . Quando o conjunto das observações é uma população, sua variância recebe o nome de variância populacional e quando o conjunto das observações é somente uma amostra estatística, sua variância é conhecida como variância amostral. Assim, a variância (σ^2) da população Y_i onde $i = 1, 2, \dots, N$ é dada por:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \mu)^2 \quad \text{onde } \mu \text{ é a média da população.}$$

Sendo X_i uma amostra onde $i = 1, 2, \dots, n$, sua variância (s^2) é dada por:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{onde } \bar{X} \text{ é a média da amostragem.}$$

2.10.3 – DESVIO PADRÃO

O desvio padrão, populacional (σ) ou amostral (s), é outra métrica capaz avaliar a dispersão e o faz em torno da média de uma variável aleatória. A expressão matemática do desvio padrão, populacional ou amostral, equivale à raiz quadrada da variância correspondente. Quanto mais próximo de zero for o desvio padrão, mais uniforme serão os dados e quanto maior for o valor do desvio padrão, mais espalhados estarão os dados em relação à média. A vantagem de usar-se o desvio padrão ao invés da variância reside no fato do desvio padrão ser expresso na mesma unidade da variável aleatória e isso facilita comparações.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \mu)^2}$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

2.10.4 – COVARIÂNCIA

A covariância (ou variação conjunta) é utilizada para determinar a relação entre dois conjuntos de dados. É uma medida do grau de interdependência linear entre duas variáveis aleatórias, ou seja, traduz a proximidade dos valores observados na série com a proximidade do tempo, visto que uma característica intrínseca das séries temporais é que normalmente as observações adjacentes são dependentes. Então, quanto mais próximas no tempo, maior sua semelhança. A medida de defasagem h entre observações numa mesma série temporal é chamada de passo ou *lag*.

A covariância depende apenas de h e não especificamente do tempo t na série, pois estamos considerando séries estacionárias, cujas propriedades estatísticas (média, variância e covariância) não mudam com o tempo. Por definição, a covariância é o valor médio do somatório dos produtos dos desvios em relação à média da série original e da série defasada, para cada par de pontos, onde par de pontos se refere a uma observação em determinada posição da série original e o segundo ponto se refere à observação de mesma posição na série defasada de h . Trata-se da comparação entre o desvio da média na série original e o desvio da média na série defasada de h no tempo.

Os correlogramas (gráficos de barras da função de autocorrelação ou da função de autocorrelação parcial) são uma expansão do conceito da covariância.

$$\text{Cov}(X_t, X_{t+h}) = \frac{\sum(X_t - \bar{X}_t) \cdot (X_{t+h} - \bar{X}_{t+h})}{n}$$

onde: X_t - Ponto na série original no tempo t .

\bar{X}_t - Média da série original.

X_{t+h} - Ponto na série defasada de h no tempo t .

\bar{X}_{t+h} - Média da série defasada.

n - Número de pares considerados. Neste caso equivale ao número de observações da série defasada, uma vez que essa possui h elementos a menos que a série original.

Pela expressão acima, a covariância será positiva quando $X_t > \bar{X}_t$ e $X_{t+h} > \bar{X}_{t+h}$ ou quando $X_t < \bar{X}_t$ e $X_{t+h} < \bar{X}_{t+h}$ e, conforme mostrado pela Figura 13, isso indica que a série possuirá uma tendência crescente.

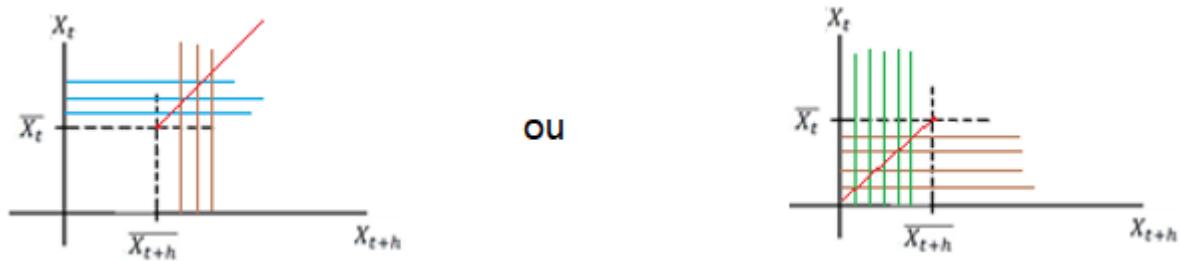


Figura 13 – Covariância positiva indicando tendência crescente.

A covariância será negativa quando $X_t > \bar{X}_t$ e $X_{t+h} < \bar{X}_{t+h}$ ou quando $X_t < \bar{X}_t$ e $X_{t+h} > \bar{X}_{t+h}$ e, conforme mostrado pela Figura 14, isso indica que a série possuirá uma tendência decrescente.



Figura 14 – Covariância negativa indicando tendência decrescente.

Finalmente, quando $X_t = \bar{X}_t$ ou $X_{t+h} = \bar{X}_{t+h}$, os pares considerados estarão igualmente distribuídos sobre os quadrantes e o valor da covariância será nulo, indicando ausência de tendência, conforme mostrado pela Figura 15.

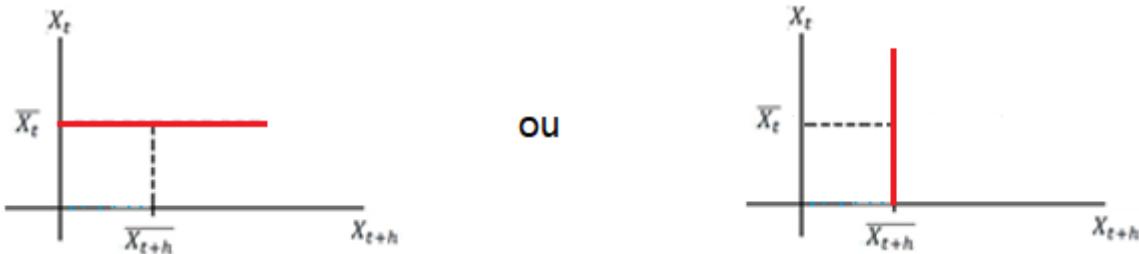


Figura 15 – Covariância nula indicando ausência de tendência.

2.10.5 – CORRELAÇÃO

Quando se deseja identificar a influência que um evento exerce sobre outro, intuitivamente surge a ideia de correlação entre eles e, logo após, o desejo de se conhecer um valor que meça quão correlacionados esses eventos estão.

Pode-se considerar que uma correlação se refere a qualquer tipo de relação entre valores médios e existem diferentes coeficientes de correlação (ρ ou r) capazes de medir o grau de correlação. Um dos mais conhecidos é o coeficiente de correlação de Pearson, obtido dividindo-se a covariância de duas variáveis pelo produto de seus desvios padrão; considerando-se a existência de uma relação de linearidade entre essas variáveis. Conforme descrito acima, uma característica importante das séries temporais é justamente a sua correlação serial, isto é, a associação linear entre duas variáveis na série. É essa característica que viabiliza estatisticamente a análise de séries temporais, pois variáveis próximas em relação ao tempo possuem proximidade de valor, são correlatas.

A correlação populacional trata da medida de direção (positiva ou negativa) e do grau que as variáveis X_t e X_{t+h} estão linearmente associadas a uma população. O coeficiente de correlação populacional $\rho_{X_t, X_{t+h}}$ entre as variáveis X_t e X_{t+h} é definido por:

$$\rho_{X_{t+h}, X_t} \equiv \text{Corr}(X_{t+h}, X_t) = \frac{\text{Cov}(X_{t+h}, X_t)}{\sigma_{X_{t+h}} \cdot \sigma_{X_t}} \quad \text{onde } \sigma \text{ é o desvio padrão}$$

Algumas propriedades:

- $-1 \leq \rho_{X_{t+h}, X_t} \leq 1$;
- $\rho_{X_{t+h}, X_t} = \rho_{X_t, X_{t+h}}$;
- $\rho_{X_{t+h}, X_t} = 0$ quando as duas variáveis não são correlacionadas.

O valor do coeficiente de correlação de Pearson varia de -1 a 1 para uma perfeita correlação linear negativa e positiva, respectivamente. Entende-se por uma perfeita correlação linear negativa quando o valor de uma variável aumenta e a outra diminui ou vice-versa e, perfeita correlação linear positiva, quando ambas aumentam ou diminuem. O valor do coeficiente de correlação de Pearson igual a 0 (zero) significa a inexistência de qualquer correlação, ou seja, as variáveis em questão são independentes. Assim, quanto mais próximo de 0 (zero) for o valor do coeficiente de correlação de Pearson, mais fraca será a correlação entre as variáveis e quanto mais próximo de -1 ou 1, mais forte será essa correlação.

A correlação amostral trata da medida de direção (positiva ou negativa) e do grau que as variáveis X_t e X_{t+h} estão linearmente associadas a uma amostra. O coeficiente de correlação amostral $r_{X_t, X_{t+h}}$ entre as variáveis X_t e X_{t+h} é definido por:

$$r_{X_t, X_{t+h}} \equiv \text{Corr}(X_{t+h}, X_t) = \frac{\text{Cov}(X_{t+h}, X_t)}{s_{X_{t+h}} \cdot s_{X_t}} \text{ onde } s \text{ é a variância}$$

2.10.6 – AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL

A autocorrelação é a correlação de uma variável com seus valores defasados em relação ao tempo. A correlação de uma variável com ela mesma, isto é, sem defasagem é sempre igual a 1 (um), pois trata-se de uma correlação perfeita.

A autocorrelação é uma medida numérica do grau de interdependência linear entre duas variáveis aleatórias e, nesse trabalho, sendo a primeira variável pertencente à série original (WS_m) e a segunda variável pertencente à série original defasada de h . Então, o valor da autocorrelação depende somente do intervalo h entre as variáveis. Se h for igual a zero, sem defasagem, isso indica a correlação entre uma variável e ela mesma, o que configura uma correlação perfeita ou total e, naturalmente, o valor numérico da autocorrelação será 1 (um).

Considerando uma série temporal X_t variando de $t=1,2,\dots,T$, o coeficiente de correlação entre X_t e X_{t+h} é chamado de autocorrelação de h -ésima ordem e é definido por:

$$\rho_h = \frac{Cov(X_{t+h}, X_t)}{\sqrt{Var(X_{t+h}) \cdot Var(X_t)}} = \frac{Cov(X_{t+h}, X_t)}{Var(X_t)}$$

onde $Var(X_t) = Var(X_{t+h})$, pois X_t é fracamente estacionária. Além disso tem-se:

- $\rho_0 = 1$;
- $\rho_m = \rho_{-m}$;
- $-1 \leq \rho_m \leq 1$.

Na expressão acima, ao se dividir as covariâncias de X_{t+h} e X_t pela raiz do produto entre as variâncias de X_{t+h} e X_t , a relação se torna adimensional.

Um conjunto de autocorrelações, ρ_h , é chamado de função de autocorrelação (FAC) de X_t e é representada através do gráfico da autocorrelação contra a defasagem. Para uma dada amostra de X_t e supondo que \bar{X} é a média amostral então, a autocorrelação amostral de primeira ordem de X_t é dada por:

$$\rho_I = \frac{\sum_{t=2}^T (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

que é um estimador consistente de ρ_I . De modo geral, a autocorrelação amostral de h -ésima ordem de X_t pode ser definida como:

$$\rho_h = \frac{\sum_{t=h+1}^T (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2} \quad \text{para } 0 \leq h \leq T-1$$

A função de autocorrelação (FAC) pode ser usada para visualizar-se como as defasagens da série podem impactar o valor atual, ou seja, se X_t é relacionado com X_{t+h} para $h \geq 1$. O valor de h indica a defasagem (passo ou *lag*) a ser aplicada sobre a série original para obtenção da segunda série (X_{t+h}), a série defasada. Em outras palavras, h indica a partir de onde (de qual

observação) à frente a série defasada se inicia. Então se $h = 1$, a primeira série a ser considerada no cálculo (X_t) é a série original e a segunda série (X_{t+1}) será aquela iniciando a partir do segundo elemento da série original. A autocorrelação pode ser representada esquematicamente através de gráficos de dispersão.

No caso de um modelo linear com apenas uma variável independente, o coeficiente de determinação R^2 equivale ao quadrado do coeficiente de correlação de Pearson e varia de 0 a 1, ou seja, entre 0 e 100%, onde 0% indica que o modelo não explica nada sobre a variabilidade dos dados de resposta ao redor de sua média e, 100%, indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média. Como ambos os coeficientes (Pearson e determinação) são adimensionais, isso permite comparar-se diferentes séries temporais através desses valores. Assim, o coeficiente de determinação R^2 é uma medida estatística de quão próximos os dados estão da linha de regressão ajustada. De modo geral, quanto maior for R^2 , melhor o modelo se ajusta aos seus dados e mais coesos estarão os pontos em torno da sua linha de regressão. Portanto, R^2 é um valor percentual que explica a variância do modelo de regressão e quanto mais variância for explicada pelo modelo de regressão, mais próximos os pontos estarão em relação à linha de regressão ajustada. Teoricamente, se um modelo pudesse explicar 100% da variância, os valores ajustados seriam iguais aos valores observados e, portanto, todos os pontos de dados estariam sobre a curva de regressão ajustada. Vale observar que R^2 não pode determinar se as estimativas e previsões dos coeficientes são tendenciosas, daí a importância dos gráficos de resíduos.

Com objetivo de apresentar algumas características descritas acima, novamente será utilizada a série *AirPassengers*, que registra mensalmente o número de passageiros aéreos entre os anos de 1949 e 1960. Através de uma análise visual do gráfico, é possível constatar-se uma possível presença de tendência e sazonalidade, indicando então a não estacionariedade da série.

A Figura 16, a mesma mostrada anteriormente na seção 2.2, representa a distribuição de passageiros no tempo.

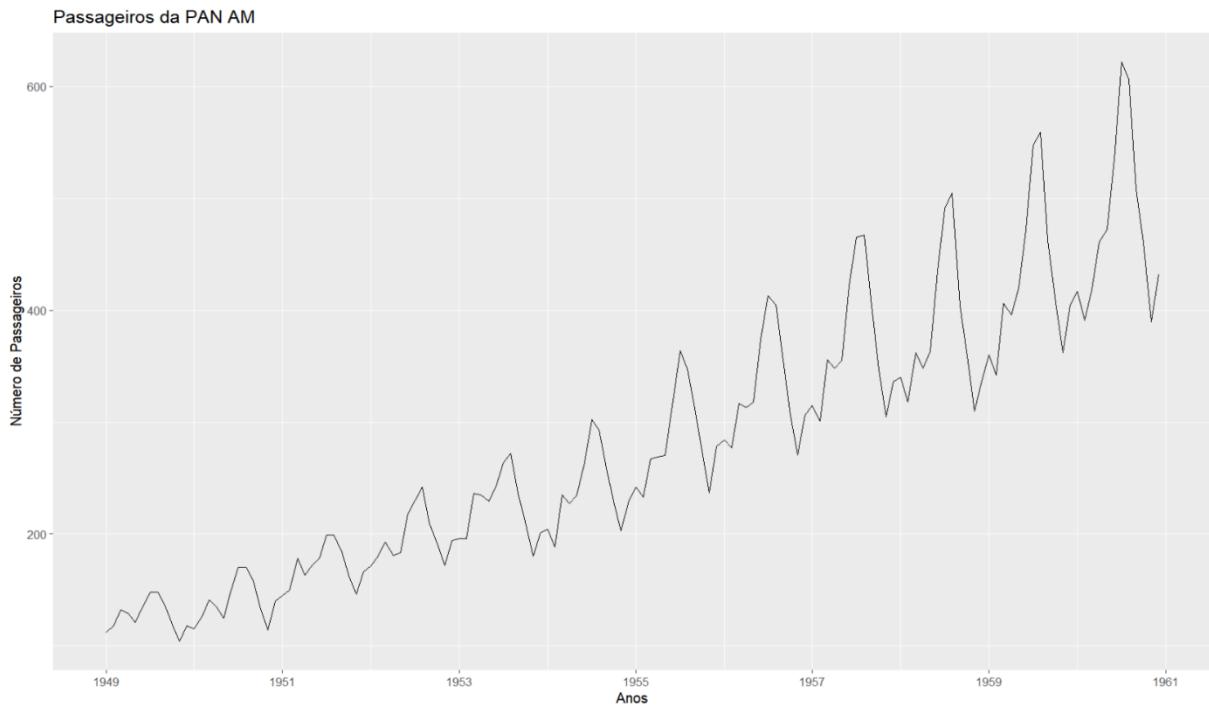


Figura 16 – Totais mensais de passageiros aéreos da PAN AM entre os anos 1949 e 1960.

Apesar do exame visual já sugerir a não estacionariedade da série, foi executado o teste Dickey-Fuller Aumentado (ADF), Tabela 1, para uma comprovação matemática que realmente demonstra sua não estacionariedade.

Tabela 1 – Teste de raiz unitária ADF para a série *AirPassengers* original.

Valor do Teste ADF	Valores Críticos		
	1%	5%	10%
-0.3524	-2.58	-1.95	-1.62

O primeiro passo na direção de tornar a série estacionária consiste em remover a sua sazonalidade. No caso da série *AirPassengers*, isso pode ser realizado através de uma diferenciação de passo (h) igual a 12 ($X_t - X_{t+12}$), pois a série apresenta sazonalidade em determinado mês durante o ano. A Figura 17 representa a série após a diferenciação.

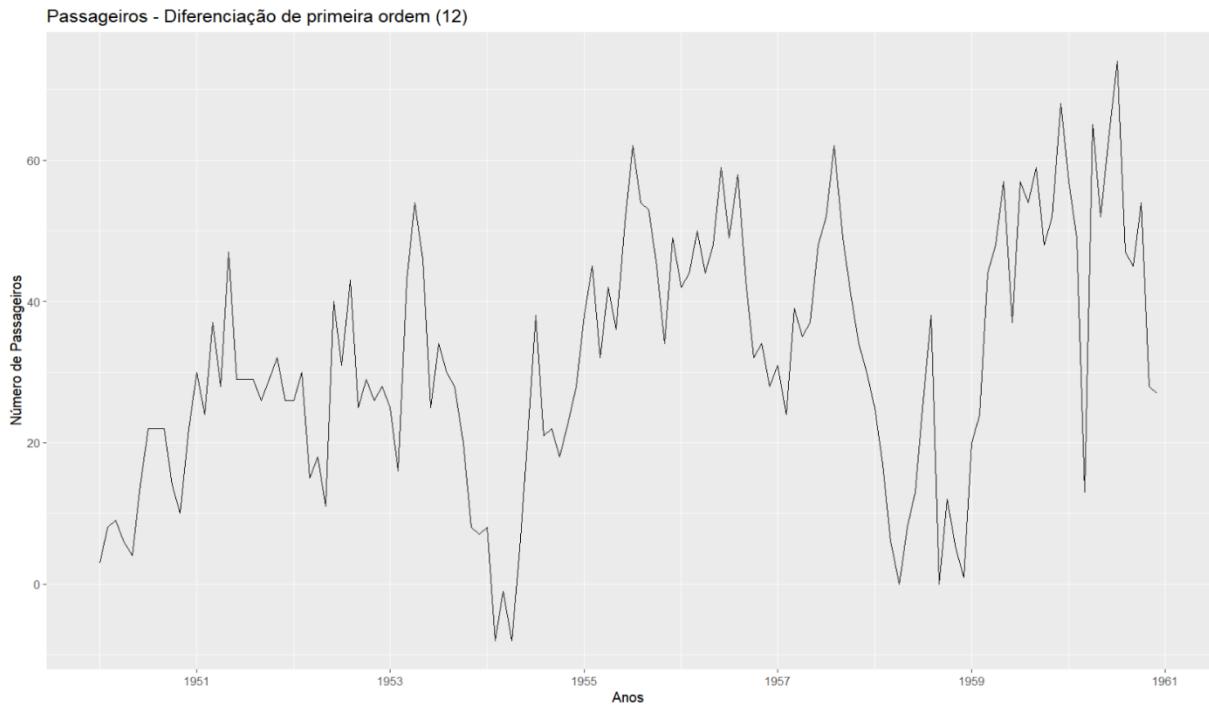


Figura 17 – Série *AirPassengers* após uma diferenciação de $h = 12$.

A sazonalidade foi amenizada, mas ainda pode-se constatar um traço de tendência. O teste ADF pode comprovar se a série já se tornou estacionária somente com a diferenciação de primeira ordem.

Tabela 2 – Teste de raiz unitária ADF para a série *AirPassengers* após a diferenciação de 12.

Valor do Teste ADF	Valores Críticos		
	1%	5%	10%
-1.3262	-2.58	-1.95	-1.62

Como previsto, o resultado do teste ADF mostrado na Tabela 2 sinaliza que a série ainda permanece não estacionária. Assim, o próximo procedimento no sentido de deixar a série estacionária se refere à remoção da tendência através de uma nova diferenciação, de segunda ordem, desta vez com $h = 1$ sobre a série anteriormente diferenciada com $h = 12$. O gráfico da Figura 18 esboça a série resultante após a segunda diferenciação.

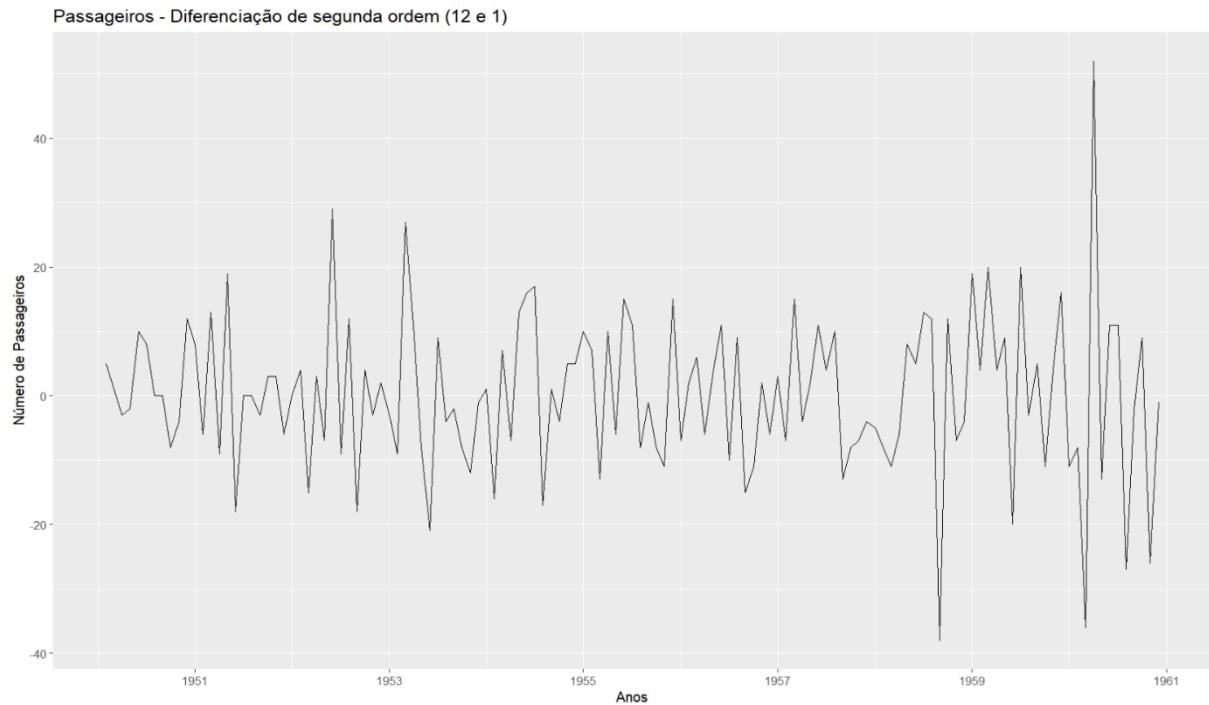


Figura 18 – Série *AirPassengers* após uma diferenciação de segunda ordem (12 e 1).

Visualmente, parece que a série resultante não apresenta mais nenhum traço de sazonalidade e tendência, mas essa confirmação acontecerá através do resultado obtido pelo teste ADF, mostrado na Tabela 3.

Tabela 3 – Teste ADF para a série *AirPassengers* após a diferenciação de segunda ordem (12 e 1).

Valor do Teste ADF	Valores Críticos		
	1%	5%	10%
-9.0108	-2.58	-1.95	-1.62

Como esperado após a diferenciação de segunda ordem, a série se tornou estacionária.

Agora pode-se passar ao exame da autocorrelação entre a série diferenciada de segunda ordem (série original tornada estacionária) e essa mesma série com tantas defasagens de meses quanto desejado. Para tanto, será utilizado um gráfico de dispersão envolvendo essas séries, por exemplo, será analisado a defasagem de um mês, isto é, $h = 1$, Figura 19.

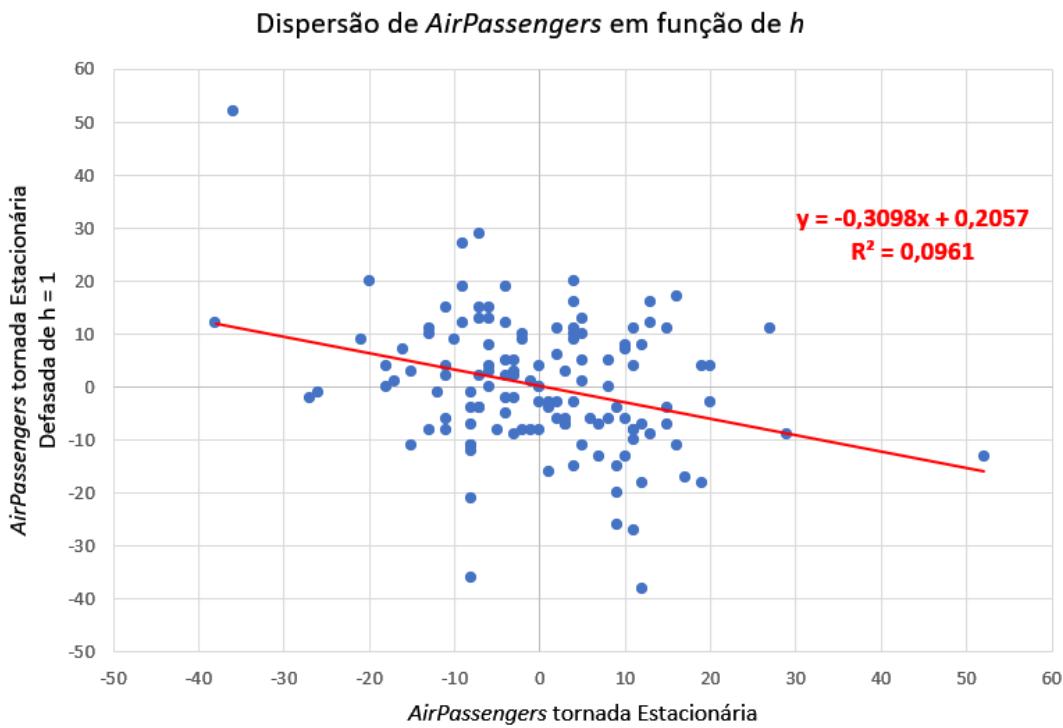


Figura 19 – Dispersão de *AirPassengers* em função de $h = 1$.

Através do gráfico de dispersão acima, pode-se observar a relação entre X_t (série original tornada estacionária) e X_{t+1} (série original tornada estacionária defasada de 1 mês). O valor do coeficiente de Pearson calculado através da fórmula acima foi: $\rho(1) = (R) \approx -0.310$ e o coeficiente de determinação (R^2) $\approx 0,0961$.

A reta vermelha representa a curva de regressão ajustada sobre os pontos, cuja equação e valor de R^2 são mostrados no gráfico. Caso todos os pontos estivessem sobre a reta ter-se-ia um valor de autocorrelação = -1 e, sendo esse valor negativo, significaria uma tendência decrescente. O valor de R (coeficiente de Pearson) calculado ($\approx -0,310$) está mais próximo de 0 (total ausência de correlação ou independência entre as variáveis) que 1 (correlação perfeita). A interpretação do coeficiente de determinação $R^2 = (-0.310)^2 \approx 0.0961 \approx 9,61\%$ significa que X_t explica 9,61% da variabilidade observada em X_{t+1} . Em outras palavras, conhecendo X_t tem-se 9,61% de probabilidade de explicar a variabilidade de valor um passo à frente (X_{t+1}), visto que o valor de h utilizado foi 1. Sendo a autocorrelação $\rho(h)$ uma função então, pode-se conhecer a relação de interdependência entre cada mês defasado h da série, que é a base para a modelagem.

A Tabela 4 apresenta as seguintes informações: os 21 primeiros valores da série *Airpassengers* tornada estacionária (X_t), os 21 valores de h e o valor da FAC calculada para cada h , isto é, o valor da FAC para cada um dos 21 meses à frente. Os valores de FAC foram calculados através do software estatístico livre Wessa (2017).

Tabela 4 – FAC calculado para defasagens de até 21 meses à frente sobre *AirPassengers*.

<i>AirPassengers</i> estacionária (X_t)	h	$\text{FAC}(h)$ ou (R)	(R^2)
5	0	1	1
1	1	-0.309815	0.09598
-3	2	0.095351	0.00909
-2	3	-0.096891	0.00938
10	4	-0.098995	0.00980
8	5	0.061001	0.00372
0	6	-0.000288	0.00000
0	7	-0.056108	0.003148
-8	8	-0.060966	0,003716
-4	9	0.175917	0.03094
12	10	-0.140279	0.01967
8	11	0.069735	0.004862
-6	12	-0.133673	0.01786
13	13	0.087177	0.00759
-9	14	0.002494	0.00000
19	15	0.065332	0.00426
-18	16	-0.109162	0.01191
0	17	-0.000338	0.00000
0	18	0.044028	0.00193
-3	19	-0.113945	0.01298
3	20	-0.091271	0.00833
3	21	0.041943	0.00175

A literatura recomenda utilizar-se apenas a parte inicial da função de autocorrelação, mais especificamente com o número máximo de h obedecendo a relação $h_{\text{máximo}} = 10 \cdot \text{Log}_{10}(N)$, onde N representa o número de observações da série. Esse critério foi seguido pelo *software* Wessa (2017), ao calcular as FACs mostradas acima sobre a série *Airpassengers* após sua transformação para torná-la estacionária. A série original *Airpassengers* possui 144 observações, mas depois de sua diferenciação de segunda ordem passou a contar com 131 observações, então, seu $h_{\text{máximo}}$ calculado foi $10 \cdot \text{Log}_{10}(131) = 21,17 \approx 21$. Assim como X_1 explica 9,60% da variabilidade observada em X_2 , do mesmo modo X_1 explica 1,19% da variabilidade observada em de X_{17} , ou seja, 16 meses à frente.

A Figura 20 apresenta o gráfico da FAC calculada para h variando de 0 a 21 e a Figura 21 mostra os seus correlogramas FAC e FACP, para a série estacionária *AirPassengers*.

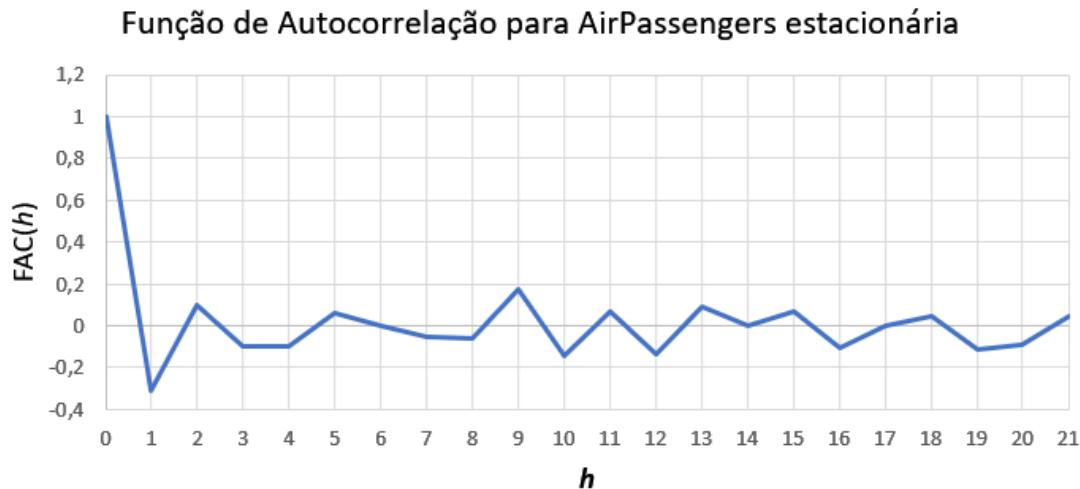


Figura 20 - FAC para cada uma das 21 defasagens da série *AirPassengers* estacionária.

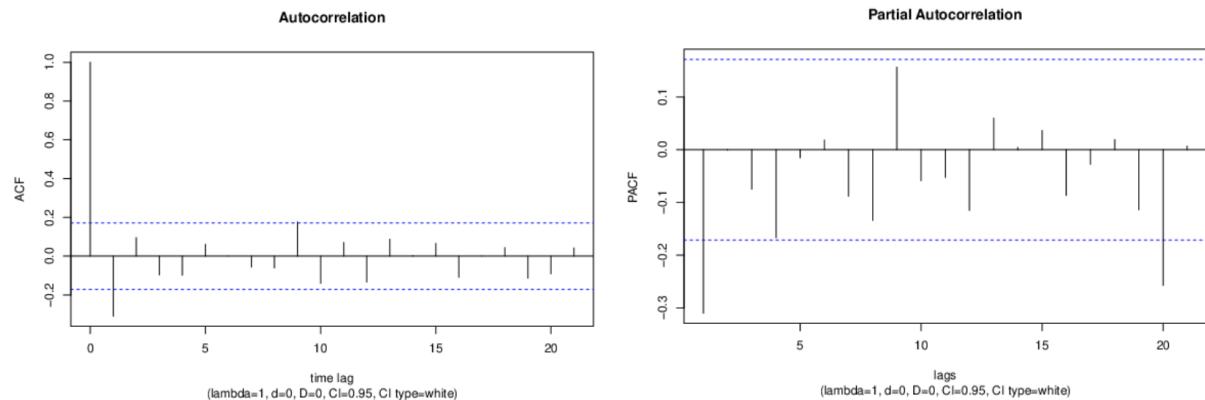


Figura 21 - Correlogramas para série *AirPassengers* estacionária. Fonte: Wessa (2017).

2.11 – CORRELOGRAMA

Um correlograma é uma representação gráfica com o objetivo de traçar os valores de autocorrelações (ou autocorrelações parciais) relativas a uma série com defasagens de h lags. Assim, a função de autocorrelação (FAC) é representada por um gráfico composto pelos valores de autocorrelação versus as defasagens consideradas.

Uma boa definição do gráfico da função de autocorrelação (FAC) e autocorrelação parcial (FACP) é dada por Hyndman *et al.* (2018) como segue: "...um gráfico FAC mostra as autocorrelações que medem a relação entre y_t e y_{t-k} para diferentes valores de k . Agora, se y_t e y_{t-1} estão correlacionados, então y_t e y_{t-2} também devem ser correlacionados, mas isso pode ser porque ambos estão correlacionados com y_{t-1} . Para evitar estas *interferências* é que se constrói a FACP. As autocorrelações parciais na FACP medem a relação entre y_t e y_{t-k} depois de remover os efeitos dos *lags* 1, 2, 3, ..., $(k-1)$. Portanto, a primeira autocorrelação parcial é idêntica à primeira autocorrelação, porque não há nada entre elas para remover. Em outras palavras, pode-se dizer que o gráfico da função de autocorrelação (FAC) mostra o valor da autocorrelação versus *lag* e o gráfico da função de autocorrelação parcial (FACP) mostra a correlação *pura* entre duas observações, isto é, eliminando as correlações implícitas com outros *lags*. A correlação pura se refere à correlação entre t e $t-k$, sem considerar as correlações intermediárias".

O correlograma é um gráfico de barras com os valores da FAC ou da FACP e apresenta uma região no eixo das abscissas chamada de intervalo de confiança. Geralmente esta região é representada por duas linhas pontilhadas paralelas ao eixo X e, normalmente, definidas por $\pm 1.96/\sqrt{N}$, onde N representa o número de elementos da série. Um valor que esteja dentro dos limites do intervalo de confiança é dito possuir 95% de probabilidade de não apresentar correlação com outro valor. Este é o comportamento esperado para uma série do tipo ruído branco, como pode ser observado na Figura 22, referente aos correlogramas (FAC e FACP) do ruído branco gerado na seção 2.12.1.

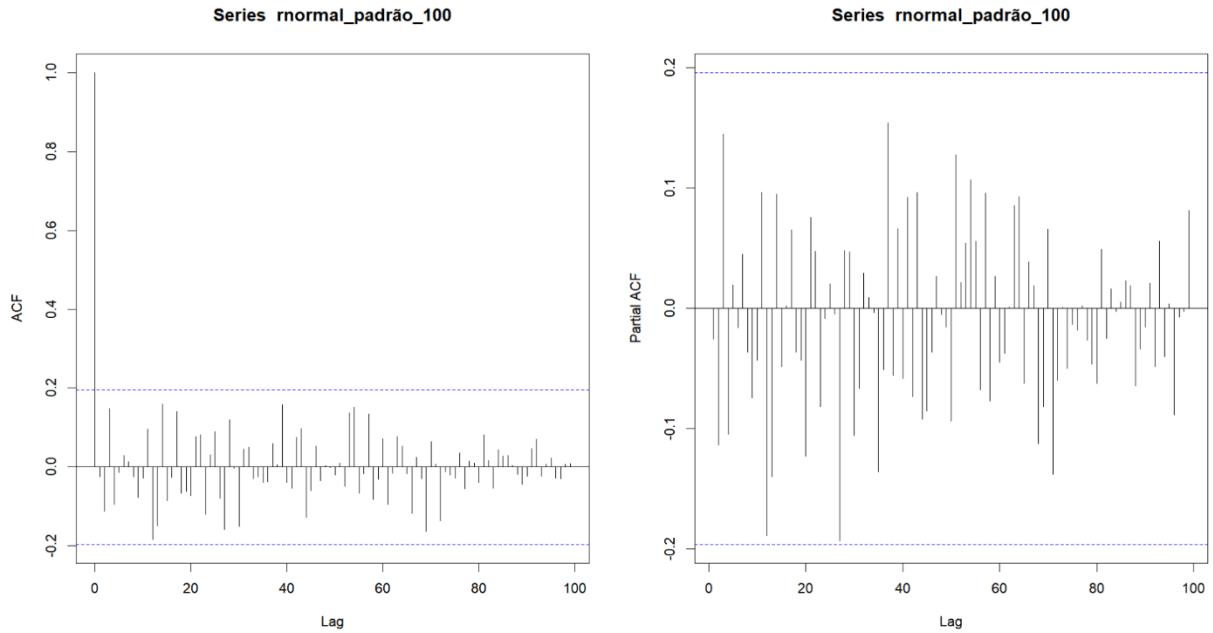


Figura 22 – Correlogramas do ruído branco gaussiano gerado.

Para qualquer série temporal estacionária, os coeficientes da função de autocorrelação $\rho(h)$ que estiverem no intervalo de confiança $\pm 1.96/\sqrt{N}$ serão considerados ZERO (sem nenhuma correlação) e são ditos independentes com 95% de probabilidade, enquanto os coeficientes que estiverem fora do intervalo serão considerados diferente de ZERO e indicarão uma correlação. Conforme esperado, todas os valores de FAC e FACp obtidos para o ruído branco gerado encontram-se dentro do intervalo de confiança, visto que por definição um ruído branco é iid.

2.12 – RUÍDO BRANCO

Originariamente o conceito de ruído branco aparece em processamento de sinais como sendo um sinal aleatório de mesma intensidade em diferentes frequências. O ruído branco se refere a um modelo estatístico para sinais e fontes de sinal, que em tempos discretos é um sinal discreto, cujas amostras se apresentam como uma sequência de variáveis aleatórias de média zero, sem autocorrelação e de variância finita.

Trazendo-se esse conceito para a análise de séries temporais, considera-se que quando um modelo estatístico é ajustado a uma série temporal estacionária Y_t , esse, comumente vem acompanhado de outra série temporal chamada ruído branco, que surge como uma série residual ε_t .

$$Y_t = f_{\text{ajuste}} + \varepsilon_t \quad \text{onde } Y_t \text{ é uma série estacionária e } \varepsilon_t \text{ o ruído branco}$$

Por definição uma série temporal ε_t é considerada um ruído branco se suas variáveis $\varepsilon_1, \varepsilon_2, \dots$ são iid, com média zero e valores aleatórios, tratando-se então de uma série temporal não previsível e isso implica em:

- 1 – Inexistência de autocorrelação e autocorrelação parcial significativa para qualquer *lag*;
- 2 – Variância constante ao longo da série;
- 3 – Todas as frequências estão em mesma proporção.

Assim, a representação mais simples de um ruído branco consiste em um conjunto de variáveis aleatórias com média zero e iid, isto é, espera-se que 95% dos picos do correlograma da FAC estejam no interior da região no eixo das abscissas chamada de intervalo de confiança e definida por $\pm 1.96/\sqrt{N}$, onde N representa o número de observações da série. Segundo Hyndman *et al.* (2018), caso o correlograma da FAC apresente alguns picos ultrapassando o intervalo de confiança ou se substancialmente mais de 5% dos picos estejam nessa situação, então a série provavelmente não será um ruído branco.

2.12.1 – RUÍDO BRANCO GAUSSIANO

Caso as variáveis aleatórias $\varepsilon_1, \varepsilon_2, \dots$ de um ruído branco sigam uma distribuição normal ou gaussiana, então a série é chamada ruído branco gaussiano.

A relevância da distribuição normal ou gaussiana é devido ao fato que diversos processos físicos e fenômenos naturais apresentam uma distribuição de probabilidade muito próxima à distribuição normal e, por conseguinte, é uma das distribuições de probabilidade mais utilizadas para modelagem de fenômenos naturais e inferências estatísticas.

Uma variável aleatória X assume uma distribuição Normal quando a sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad \text{onde } \mu \text{ é a média e } \sigma \text{ o desvio padrão}$$

A curva densidade de probabilidade é chamada de curva de Gauss ou curva em forma de sino, cuja notação é $X \sim N(\mu ; \sigma^2)$. Em estatística, o caractere \sim significa *tem distribuição*. A função $f(x)$ acima descreve uma curva simétrica, unimodal em forma de sino e quando os valores de μ e σ^2 são iguais a 0 e 1, respectivamente, a curva é denominada normal padrão ou normal reduzida, cuja notação é $X \sim N(0;1)$. Uma vantagem de usar-se uma distribuição normal padrão é que os seus valores de probabilidade (área sob a curva) já foram previamente calculados e encontram-se tabelados. Qualquer distribuição normal $X \sim N(\mu ; \sigma^2)$ pode ser padronizada, de modo que o processo de padronização da variável aleatória X fixe seus parâmetros em $\mu = 0$ e $\sigma = 1$. Essa transformação é representada pela definição de uma nova variável aleatória Z , chamada de variável aleatória normal padronizada e dada pela função linear Z abaixo.

$$Z = \frac{X - \mu}{\sigma} \quad \text{onde } X \text{ é a variável aleatória normal, } \mu \text{ a média e } \sigma \text{ o desvio padrão}$$

Neste trabalho, uma distribuição normal é verificada através de três técnicas. A primeira consiste na análise do gráfico Q-Q ou *quantile plot*, a segunda se refere ao seu histograma e a terceira técnica utiliza o teste de hipóteses de Shapiro-Winker.

O gráfico Q-Q ou *quantile plot* é uma ferramenta gráfica, que auxilia na avaliação de normalidade de um conjunto de dados e é usualmente utilizado em conjunto com o histograma. O gráfico Q-Q é somente uma verificação visual, não se configurando como uma prova completa e por isso faz-se necessária a utilização de outro teste estatístico para uma confirmação mais precisa. Um gráfico Q-Q basicamente realiza uma comparação dois a dois dos quantis teóricos de uma distribuição normal e os quantis do conjunto de dados em exame. Caso esse último seja uma distribuição normal, os pontos mostrados no gráfico se concentram em torno de uma reta. Cabe ressaltar que no gráfico Q-Q é comum que alguns pontos se afastem ligeiramente da reta mesmo quando o conjunto de dados é considerado uma distribuição normal, daí a importância de analisar este gráfico juntamente com um histograma.

O teste de hipóteses de Shapiro-Wilker parte da hipótese nula (H_0) considerando que o conjunto analisado segue uma distribuição normal e a hipótese alternativa (H_a) considera que o conjunto de dados não segue uma distribuição normal. Assim, dado um nível de significância, geralmente estabelecido em 5% (0.05), se o resultado do teste (*p-value*) for maior que o nível de significância, não haverá motivos para se rejeitar a hipótese nula (H_0) e concluir-se que a distribuição apresenta indícios de ser normal. Caso o resultado do teste (*p-value*) seja menor que o nível de significância, a hipótese nula (H_0) é rejeitada e a hipótese alternativa (H_a) aceita, isto é, considera-se que o conjunto analisado não segue uma distribuição Normal. No entanto, é sempre prudente contar com o suporte dos gráficos Q-Q e histograma.

A Figura 23 apresenta um gráfico contendo cem valores gerados randomicamente com uma distribuição normal padrão ($\mu = 0$ e $\sigma^2 = 1$). Tal distribuição foi obtida através da função *rnorm(n=100, mean=0, sd=1)* disponibilizada pelo *software R*, com o valor de semente igual a 123.

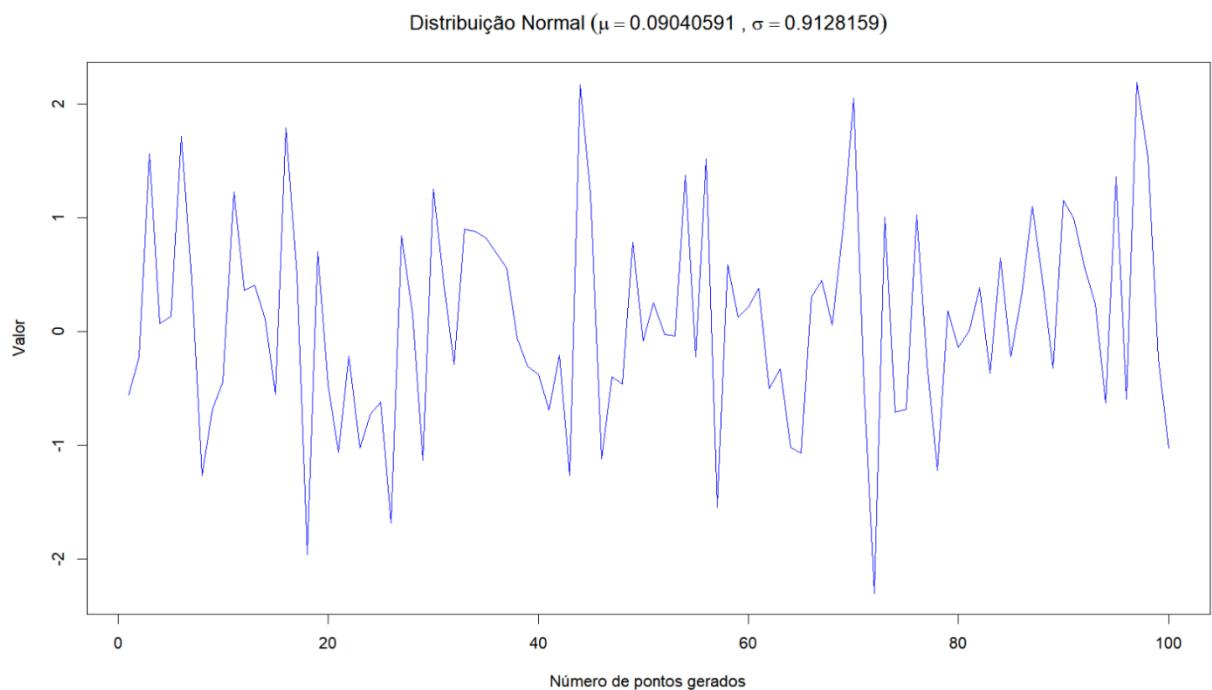


Figura 23 – Conjunto de 100 números randômicos gerados com distribuição normal padrão.

Observa-se que mesmo sendo definida com média = 0 e desvio padrão = 1, a distribuição aleatória gerada pela função *rnorm* produziu um conjunto de dados cuja média e desvio padrão não resultaram exatamente nesses valores. Isso ocorre devido à variação randômica produzida pelo comando *rnorm*. O mesmo pode acontecer com alguns valores de autocorrelação, que se espera igual a zero, porém, dependendo do *lag* considerado, isso também pode não suceder.

A Figura 24 mostra o histograma (azul) obtido pela distribuição gerada e sobre esse a curva normal padrão correspondente (vermelho). Para fins comparativos, foi escolhida a opção densidade no eixo das ordenadas do histograma.

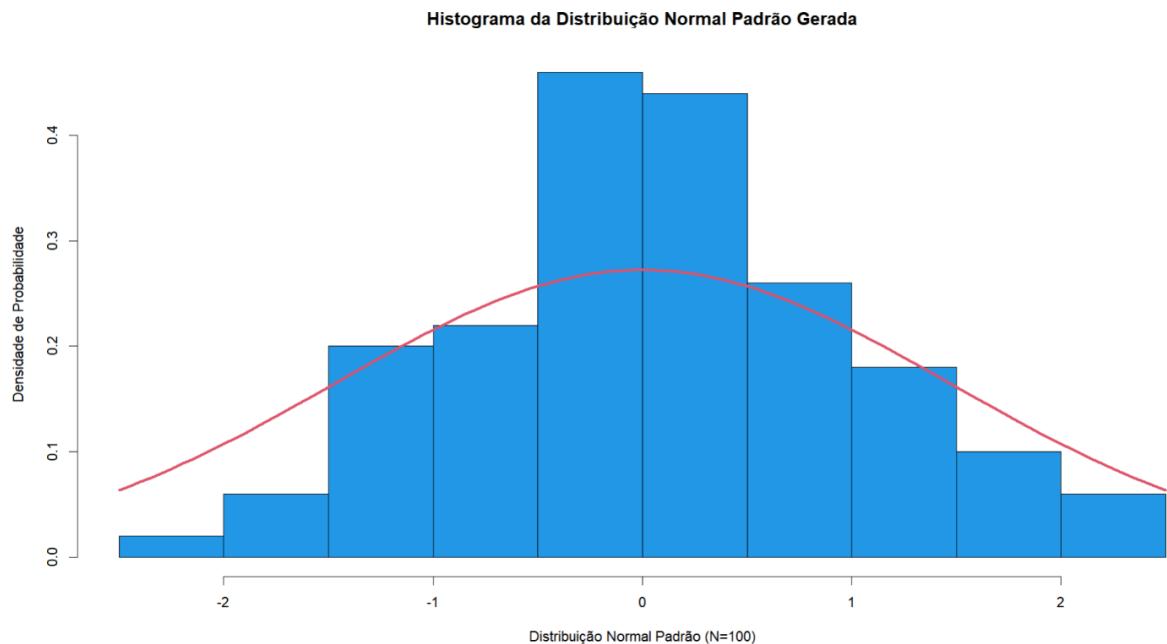


Figura 24 – Histograma e curva densidade de probabilidade do conjunto normal padrão gerado.

A Figura 25 esboça o gráfico Q-Q para esse conjunto normal padrão gerado, com a linha vermelha representando a distribuição normal padrão teórica.

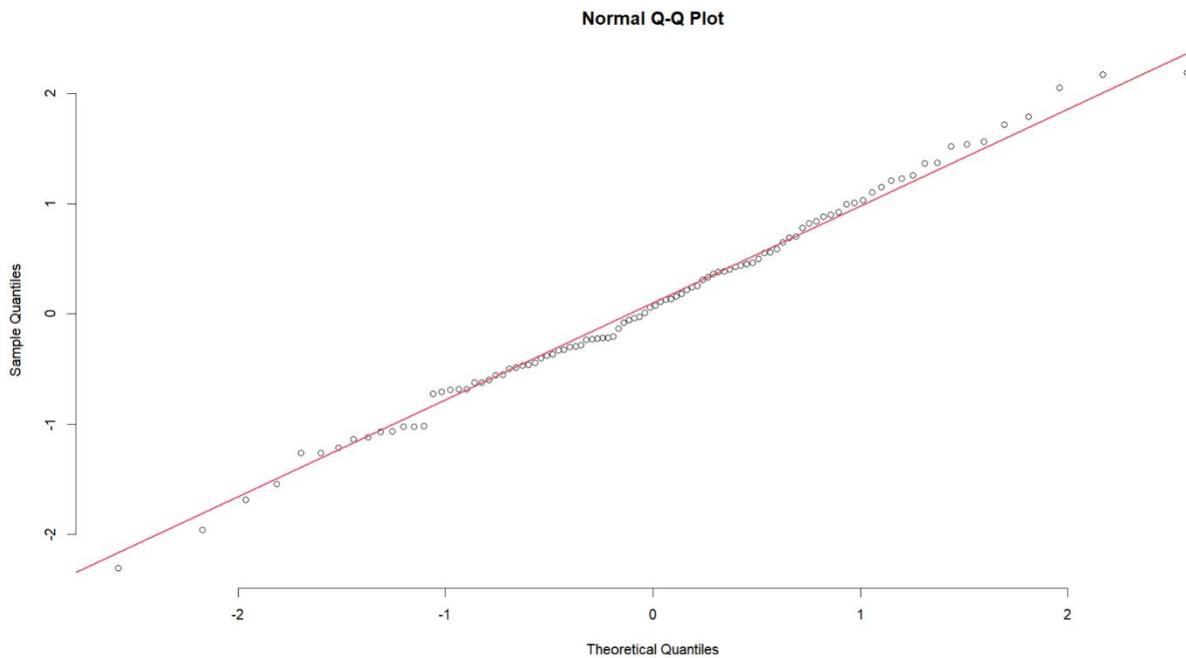


Figura 25 – Gráfico Q-Q para o conjunto normal padrão gerado.

A Figura 22 da seção 2.11 esboça os correogramas relativos à distribuição gerada e, conforme esperado, todos os *Lags* encontram-se no interior do intervalo de confiança. Em outras palavras, não há qualquer correlação entre os pontos do ruído branco, ou seja, são independentes. Cabe registrar que dependendo do valor da semente utilizada na geração randômica, pode resultar em algum *lag* ultrapassando o intervalo de confiança, mas tal seria considerado aceitável.

O teste de normalidade de Shapiro-Wilker executado sobre o conjunto normal padrão gerado acima apresenta o resultado registrado na Tabela 5.

Tabela 5 – Valor do teste de hipóteses de normalidade Shapiro-Wilker.

Valor do teste Shapiro-Wilker	Nível de Significância
0.9349	0.05

Interpretação do resultado:

Como o valor do teste ($p\text{-value} = 0.9349$) é muito maior que o valor crítico (0.05), então não há motivos para rejeitar-se a hipótese nula (H_0), significando assim que a distribuição do conjunto gerado pode ser considerada normal, como esperado.

2.13 - MODELAGEM

Após a análise descritiva da série temporal, o foco principal em termos de modelagem consiste na determinação da componente aleatória ou restante (*remainder*), isto é, o que resta da série após o ajuste do modelo. Pode-se garantir que uma série temporal seja a realização de uma variável aleatória quando pelo menos uma das características probabilísticas da série não muda com o tempo; e isso é a chamada fraca estacionariedade. As características probabilísticas são:

- 1 – Média μ constante;
- 2 – Variância σ^2 constante;
- 3 – Covariância $\text{Cov}(X_t, X_{t+h})$ dependendo apenas da diferença no tempo h entre duas medidas.

2.14 - RESÍDUOS

Hyndman *et al.* (2018) informam que para muitos modelos de séries temporais (não todos), os resíduos equivalem à diferença entre as observações e os valores correspondentes ajustados:

$$e_t = Y_t - \hat{Y}_t$$

Os resíduos possibilitam um exame minucioso e sistemático do quanto adequadamente um modelo capturou as informações contidas nos dados. Um bom modelo de previsão produzirá resíduos com as seguintes propriedades:

1 – Resíduos não correlacionados. Caso haja correlações entre os resíduos, isso significa que ainda existem informações pertencentes a outras componentes presentes. Em outras palavras, a existência de correlações entre os resíduos indica que eles ainda contêm *resquícios* de tendência e/ou sazonalidade, que deveriam ser adicionadas às suas componentes de direito para a melhoria do processo de previsão.

2 – O valor da média dos resíduos deve ser zero, caso contrário, a previsão possuirá um viés.

Assim, qualquer modelo de previsão que não satisfaça as propriedades acima pode ser aprimorado para fornecer uma previsão mais acurada. Isso, porém, não significa que um modelo que atenda essas propriedades não possa ser otimizado. É possível haver diversos modelos diferentes de previsão para o mesmo conjunto de dados, como será visto no decorrer deste trabalho. A checagem dessas propriedades é importante no senso da avaliação se um determinado modelo está usando todas as informações disponíveis.

2.15 – CRITÉRIOS DE INFORMAÇÃO

Os chamados critérios de informação são métricas utilizadas para mensurar a qualidade de um modelo estatístico. Os critérios de informações mais utilizados são o AIC (Akaike), BIC (bayesiano), HQ (Hannan & Quinn) e, mais recentemente, o AICc (Akaike corrigido). Diante de dois modelos candidatos a representar um fenômeno, a escolha do mais adequado é aquele que apresentar o menor valor do critério de informação.

2.16 – DECOMPOSIÇÃO DE SÉRIES TEMPORAIS - MODELO CLÁSSICO

A análise clássica de séries temporais foi uma das primeiras abordagens desta matéria, tendo sua origem na década de 1920 e amplamente utilizada até a década de 1950, motivada por um cenário de aplicações econométricas e num momento histórico que não estavam disponíveis os recursos computacionais e até mesmo alguns avanços matemáticos existentes atualmente. Contudo, o cerne da conceituação é bastante sólido e continua válido até hoje.

O modelo clássico considera que uma série temporal é formada por quatro componentes: Tendência (T), Ciclo (C), Sazonalidade (S) e Irregular (I). A tendência é uma componente de longo prazo e indica a propensão de evolução dos valores da série. O ciclo também é uma componente de longo prazo e mostra uma alternância sistemática da variável com ocorrência superior a um ano, em outras palavras, é a presença de períodos de alta e baixa com duração superior a um ano presentes no decorrer da série. A sazonalidade é uma componente de curto prazo, que se repete a cada ano na mesma época, permitindo uma avaliação do comportamento

sistemático da série durante o ano. Assim, ao contrário da componente ciclo, a sazonalidade retrata flutuações com duração inferior a um ano e com frequência anual. Se os elementos de uma série temporal são registrados com uma periodicidade igual ou superior a um ano, não haverá influência de sazonalidade, pois não será possível observar as flutuações repetitivas no interior dos anos. Então, somente uma série temporal registrada com periodicidade inferior a um ano oferece condições de identificação de sazonalidade. A componente irregular retrata fatos fortuitos capazes de influenciar significantemente a série, algo inesperado e sem a possibilidade de previsão, mas que pode estar presente no modelo.

Desta forma, no modelo clássico, o valor da série Y é o resultado da combinação destas quatro componentes e o processo de decomposição da série consiste na identificação de suas componentes.

$$Y = f(T, C, S, I)$$

As componentes tendência, ciclo, sazonalidade e irregular podem estar agrupadas de duas maneiras distintas, são os chamados modelos aditivos e multiplicativos.

Modelos Aditivos

Nessa categoria, a série temporal é formada através da soma de suas componentes e cada uma possui a mesma unidade da série.

$$Y = T + C + S + I \text{ ou}$$

$$Y = T + C + I \quad (\text{para séries anuais})$$

Modelos Multiplicativos

Neste tipo de modelo a série temporal é formada pelo produto de cada uma de suas componentes, porém somente a componente tendência possui a mesma unidade da série. As demais componentes são expressas sob a forma de índices, que modificam a tendência e possuem valores na vizinhança de 1 (um). Índices maiores e menores que 1 (um) aumentam e diminuem a tendência, respectivamente, e índices iguais a 1 (um) não influenciam.

$$Y = T \times C \times S \times I \text{ ou}$$

$$Y = T \times C \times I \quad (\text{para séries anuais})$$

A escolha mais adequada entre o modelo aditivo e multiplicativo depende das características inerentes das variáveis consideradas no estudo. No entanto, podemos avaliar uma escolha através de medidas de acuracidade baseadas na comparação entre os valores reais e os previstos por cada modelo, isto é, através de uma medida de erro obtida a partir da subtração entre o valor real e o previsto.

2.16.1 - OBTENÇÃO DA COMPONENTE TENDÊNCIA NO MODELO CLÁSSICO

A tendência é uma componente que mostra a evolução da variável analisada no longo prazo. A aferição da tendência pode ser obtida através de modelos de regressão ou médias móveis percentuais ou exponencialmente ponderada (ajuste exponencial). Uma vez identificada a tendência, essa componente poderá ser removida da série a fim de facilitar a identificação das demais componentes; ser aplicada no processo de recomposição e previsão e também apresentar uma estimativa do nível da série.

2.16.1.1 – OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE MODELOS DE REGRESSÃO

Nos modelos de regressão utilizados para obtenção da tendência, a variável independente será o tempo. Abaixo são apresentados alguns exemplos de modelos de regressão:

- Linear: $T = b \cdot t + a$
- Polinômio de Segundo Grau: $T = c^{te} \cdot t^2 + b \cdot t + a$
- Logarítmica: $T = b \cdot \ln(t) + a$
- Potência: $T = b \cdot t^a$
- Exponencial: $T = b \cdot e^{a \cdot t}$

Onde:

T é a tendência;

t é o valor de tempo (ou período) e

a e b são os coeficientes:

$$b = \frac{n \times \sum_{i=1}^n (t_i \times Y_i) - \sum_{i=1}^n t_i \times \sum_{i=1}^n Y_i}{n \times \sum_{i=1}^n (t_i^2) - \left(\sum_{i=1}^n t_i \right)^2}$$

$$a = \frac{\sum_{i=1}^n Y_i - b \times \sum_{i=1}^n t_i}{n}$$

sendo Y_i um elemento da série temporal, t_i o período (tempo) associado ao elemento e n o número de períodos da série.

A determinação de qual modelo de regressão é o mais adequado para representar a componente tendência será apontada após a avaliação das seguintes métricas: Erro Absoluto Médio (EAM), Erro Quadrático Médio (EQM), Erro Percentual Médio (EPM) e Erro Percentual Absoluto Médio (EPAM), todas embasadas sobre o erro (e_t) obtido através da diferença entre o valor real da série e o valor estimado pela regressão, conforme definidas abaixo:

$$\text{Erro: } e_t = Y_t - \hat{Y}_t$$

$$\text{Erro Absoluto Médio (EAM): } EAM = \frac{1}{n} \times \sum_{t=1}^n |e_t|$$

$$\text{Erro Quadrático Médio (EQM): } EQM = \frac{1}{n} \times \sum_{t=1}^n e_t^2$$

$$\text{Erro Percentual Médio (EPM): } EPM = \frac{1}{n} \times \sum_{t=1}^n \left[\left(\frac{e_t}{Y_t} \right) \times 100 \right]$$

$$\text{Erro Percentual Absoluto Médio (EPAM): } EPAM = \frac{1}{n} \times \sum_{t=1}^n \left| \left(\frac{e_t}{Y_t} \right) \times 100 \right|$$

2.16.1.2 – OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE MÉDIAS MÓVEIS CENTRADAS

Uma alternativa para a obtenção da tendência é através das médias móveis centradas. O primeiro passo para o cálculo da média móvel é a escolha do seu número de períodos n . Esse valor corresponde à quantidade de elementos da série temporal sobre a qual será calculada uma média simples. Em outras palavras, trata-se de uma *janela móvel* a percorrer toda a série. Assim, o primeiro conjunto de elementos envolvidos nessa média é composto pelos n primeiros elementos da série temporal. O resultado da média calculada é colocado numa posição central em relação aos elementos da série envolvidos no cálculo, pois se refere à média daqueles n

períodos. Os conjuntos seguintes serão compostos desprezando-se o primeiro valor do conjunto anterior e acrescentando-se o próximo elemento da série ainda não utilizado e, novamente, o resultado obtido é colocado na posição central em relação aos elementos do conjunto corrente. Este procedimento é repetido até que o último elemento da série seja incluído no conjunto. Essa dinâmica - colocação da média na posição central dos elementos envolvidos - é a responsável pela perda de valores de média correspondente aos períodos de início e fim da série.

Quando o número de períodos escolhido para o cálculo da média móvel centrada é ímpar, a posição central (posição mediana) encontra uma correspondência natural na série, porém quando esse número é par não existe tal correspondência. Isso implica na realização de um procedimento denominado centralização. A centralização corresponde ao cálculo de uma nova média em relação ao total móvel de dois períodos consecutivos previamente calculados, mas que agora encontra correspondência nos períodos da série temporal. Cabe ressaltar que o termo *periodicidade da série temporal* se refere à frequência com que esta é registrada e as expressões *periodicidade da média móvel* ou *ordem da média* ou *janela móvel* se referem ao número de períodos da série a ser utilizado no cálculo das médias.

A construção gráfica retratando uma série temporal costuma ser o primeiro passo de uma análise exploratória. Assim, a Figura 26 mostra a representação gráfica de uma série temporal hipotética de periodicidade igual a 21 dias (ou 21 períodos) e as curvas correspondentes às suas médias móveis centradas de 3, 5 e 7 períodos. A Tabela 6 apresenta a planilha com o posicionamento das médias móveis centradas de 3, 5 e 7 períodos e a Tabela 7 a planilha com o posicionamento da média móvel centrada de 4 períodos.

Pode-se constatar na Figura 26 que a curva que descreve a média móvel de sete períodos (vermelha) promove maior “suavização” da série que as curvas relativas às médias móveis de três (laranja) e cinco (roxa) períodos. Então, quanto maior o período da média móvel, mais atenuada ficará a linha de tendência que representa melhor o comportamento de longo prazo e, quanto menor for o período da média móvel, mais a tendência acompanhará os dados reais da série.

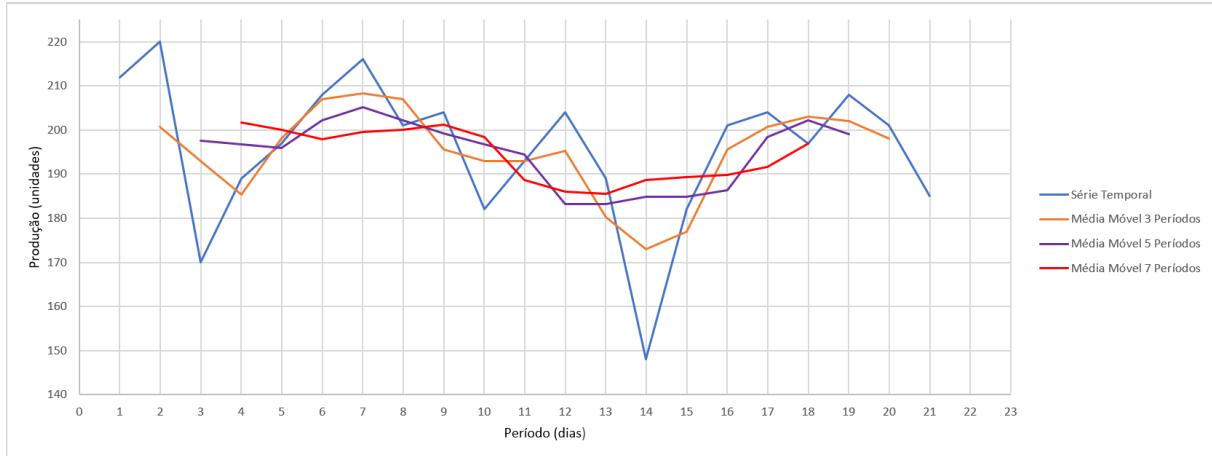


Figura 26 – Série Temporal hipotética com suas médias móveis de 3, 5 e 7 períodos.

Observando a figura acima, verifica-se uma perda de valores inicial e final das médias móveis correspondentes aos valores iniciais e finais da série, mais especificamente uma perda de dois valores (um inicial e um final) para a média móvel de três períodos; uma perda de quatro valores (dois iniciais e dois finais) para a de cinco períodos e uma perda de seis valores (três iniciais e três finais) para a média móvel de sete períodos. Isso acontece devido à própria definição de médias móveis, que exige que cada valor da média calculada ocupe a posição central (ou posição mediana) em relação aos valores da série utilizados no cálculo, como mostrado na Tabela 6.

Tabela 6 – Posicionamento das médias móveis de 3, 5 e 7 períodos.

	A	B	C	D	E	F	G	H
1	dia	Produção	Total Móvel 3 Períodos	Média Móvel 3 Períodos	Total Móvel 5 Períodos	Média Móvel 5 Períodos	Total Móvel 7 Períodos	Média Móvel 7 Períodos
2	1	212						
3	2	220	602	200,67				
4	3	170	579	193,00	988	197,6		
5	4	189	556	185,33	984	196,8	1412	201,71
6	5	197	594	198,00	980	196	1401	200,14
7	6	208	621	207,00	1011	202,2	1385	197,86
8	7	216	625	208,33	1026	205,2	1397	199,57
9	8	201	621	207,00	1011	202,2	1401	200,14
10	9	204	587	195,67	996	199,2	1408	201,14
11	10	182	579	193,00	984	196,8	1389	198,43
12	11	193	579	193,00	972	194,4	1321	188,71
13	12	204	586	195,33	916	183,2	1302	186,00
14	13	189	541	180,33	916	183,2	1299	185,57
15	14	148	519	173,00	924	184,8	1321	188,71
16	15	182	531	177,00	924	184,8	1325	189,29
17	16	201	587	195,67	932	186,4	1329	189,86
18	17	204	602	200,67	992	198,4	1341	191,57
19	18	197	609	203,00	1011	202,2	1378	196,86
20	19	208	594	198,00		199		
21	20	201						
22	21	185						

No cálculo da média móvel de período igual a 4 (Tabela 7), cada valor na coluna C se refere à soma de 4 elementos da série. Cada valor da coluna D corresponde à soma de dois

elementos da coluna C. O valor da média calculada na coluna E é obtido dividindo-se cada valor de D por oito, pois essa coluna (D) envolve a soma de 2 valores de quatro períodos da coluna C, totalizando assim oito períodos, como pode ser observado na barra de funções da planilha ($=D4/8$). Este é o artifício utilizado para a centralização correta quando o período da média móvel for par.

Tabela 7 – Posicionamento das médias móveis de 4 períodos.

	A	B	C	D	E
1	2	3	Total Móvel de 4 Períodos	Totais Móveis de 2 Períodos Centrados	Médias Móveis de 4 Períodos Centradas
2	1	212			
3	2	220			
4	3	170	791,00	1567,00	195,875
5	4	189	776,00	1540,00	192,5
6	5	197	764,00	1574,00	196,75
7	6	208	810,00	1632,00	204
8	7	216	822,00	1651,00	206,375
9	8	201	829,00	1632,00	204
10	9	204	803,00	1583,00	197,875
11	10	182	780,00	1563,00	195,375
12	11	193	783,00	1551,00	193,875
13	12	204	768,00	1502,00	187,75
14	13	189	734,00	1457,00	182,125
15	14	148	723,00	1443,00	180,375
16	15	182	720,00	1455,00	181,875
17	16	201	735,00	1519,00	189,875
18	17	204	784,00	1594,00	199,25
19	18	197	810,00	1620,00	202,5
20	19	208	810,00	1601,00	200,125
21	20	201	791,00		
22	21	185			
23					

Para obter-se a tendência através de médias móveis, numa série temporal de periodicidade inferior a um ano, devemos escolher como período da média móvel o mesmo valor da periodicidade da série. Por exemplo, para obtenção da tendência através de média móvel para uma série registrada com periodicidade trimestral, devemos escolher o valor 4 para a periodicidade da média móvel, pois existem 4 trimestres no ano. Este procedimento garante a obtenção da tendência sem influência da sazonalidade.

2.16.1.3 – OBTENÇÃO DA TENDÊNCIA ATRAVÉS DE AJUSTE EXPONENCIAL

Uma vantagem do método de ajuste exponencial em relação ao método de médias móveis é que no primeiro não há perdas de períodos iniciais e finais. Outra vantagem é que o ajuste exponencial leva em conta todos os períodos anteriores da série e a média móvel somente

aqueles pertencentes à janela móvel. O ajuste exponencial permite também uma previsão de curto prazo, mais especificamente, 1 (um) período imediatamente posterior ao último registrado na série.

O primeiro valor ajustado corresponde ao primeiro valor da série e cada valor subsequente considera a estimativa anterior e seu erro, ponderado por uma constante de suavização ou coeficiente de ajuste α . O valor de α mais eficiente é aquele que produz o menor Erro Quadrático Médio (MSE) dos valores calculados pelo ajuste exponencial.

Segundo MAKRIDAKIS *et al.* (1998) o ajuste exponencial pode ser obtido através da seguinte expressão:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

Onde:

\hat{Y}_{t+1} - Estimativa no período $t+1$

\hat{Y}_t - Estimativa no período t

Y_t - Valor real da série no período t

α - Coeficiente de ajuste $0 \leq \alpha \leq 1$

Para demonstração prática de alguns conceitos será utilizada uma série temporal registrada com periodicidade mensal e de abrangência igual a 15 anos. A Figura 27 apresenta a curva de evolução de vendas entre os anos 1990 e 2004.

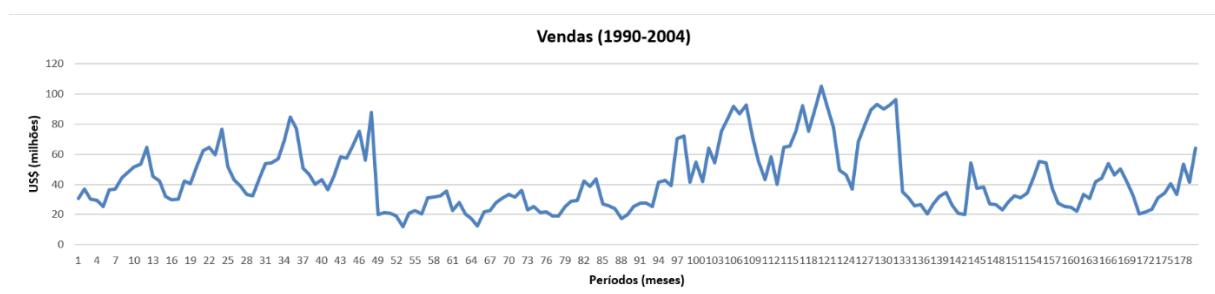


Figura 27 – Série temporal cobrindo 15 anos, com periodicidade mensal.

A Figura 28 apresenta o ajuste dos modelos de tendência linear, logarítmica, polinômio de segundo grau, potência e exponenciação.

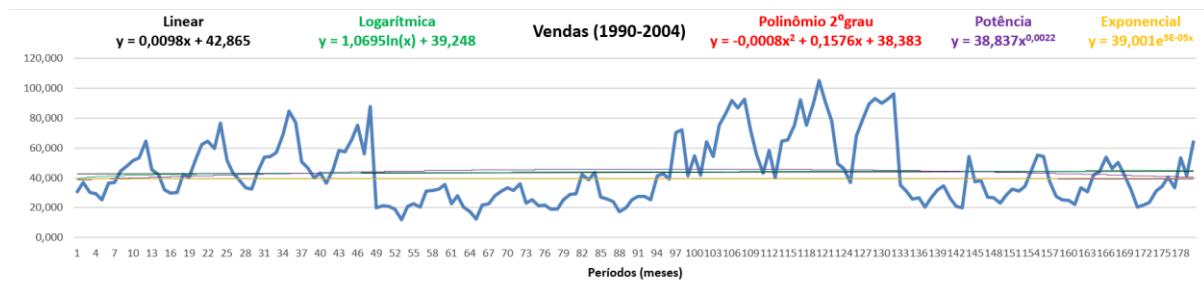


Figura 28 – Série temporal de vendas com 5 modelos de tendência.

2.16.2 – OBTENÇÃO DA COMPONENTE SAZONAL NO MODELO CLÁSSICO

Sazonalidade são variações de curto prazo que ocorrem sistematicamente no período de um ano. Sendo assim, séries temporais registradas com período igual ou superior a um ano não têm como apresentar tais variações. Séries temporais registradas com uma periodicidade inferior a um ano podem mostrar algum tipo de sazonalidade e essa componente é representada por índices associados a cada valor da série, tanto no modelo aditivo como no multiplicativo.

Os índices sazonais influenciam a tendência, seja adicionando (modelo aditivo), seja multiplicando (modelo multiplicativo). A regularidade sazonal, quando presente, é mais facilmente detectada em séries que possuam no mínimo cinco anos completos. Aqui, para a avaliação dos índices sazonais, será utilizado o método da razão para média móvel.

2.16.2.1 – MODELO ADITIVO

No modelo aditivo o elemento neutro é o 0 (zero), então podemos considerar que os índices com esse valor ou próximos não exercem influência sobre a série, ao passo que índices substancialmente diferentes de zero (positivos ou negativos) indicam influência da componente sazonal na série.

$$Y = T + C + S + I \rightarrow S = Y - (T + C + I)$$

A componente sazonal é o resultado do valor da série menos a soma das demais componentes. A média móvel centrada, quando definida com periodicidade igual a periodicidade da série, torna-se uma ferramenta útil para a avaliação das componentes de tendência, ciclo e irregular. Isto acontece porque a média móvel centrada amortiza as variações indicativas da componente irregular, acompanha a tendência da série e evidencia as alternâncias de longo prazo, que caracterizam uma possível presença de ciclos. A Figura 29 mostra a série vendas juntamente com a curva representativa de sua média móvel centrada de período igual a 12, o mesmo da série temporal.

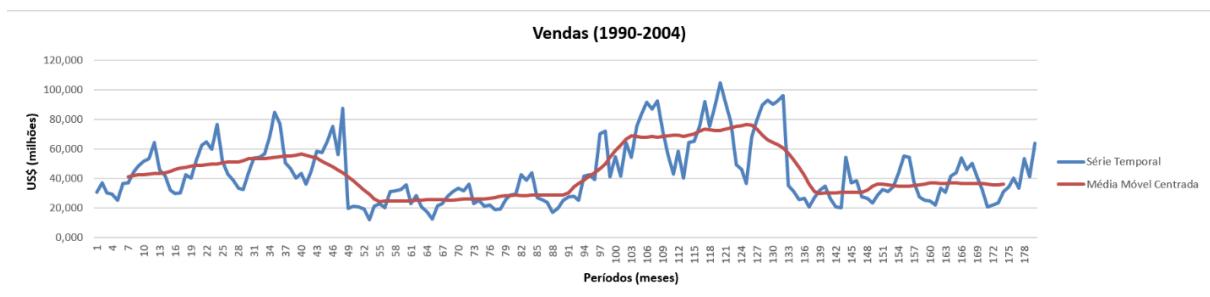


Figura 29 – Série temporal e média móvel centrada, ambas com periodicidade igual a 12.

Observando-se a curva de média móvel centrada acima, constata-se que essa segue a tendência da série, fornece indicação de possíveis ciclos (as alternâncias de alta e baixa no longo prazo) e promove uma suavização das possíveis componentes de irregularidade.

Assim, o índice sazonal aditivo associado a cada elemento da série (que possua média móvel centrada) é obtido subtraindo-se do valor desse elemento da média móvel centrada a ele associado.

$$\text{Índice Sazonal Aditivo} = Y - \text{Média Móvel Centrada} \quad (1)$$

Considerando que a série acumule valores relativos a n anos completos, a equação acima associará um índice sazonal a cada período da série. Então para cada período específico (aquele que se repete a cada ano) tem-se $n-1$ índices sazonais associados, visto que a média móvel centrada definida com mesma periodicidade da série perde um ano, ou seja, *não* associa valores a um número de períodos iniciais e finais, os quais totalizam um ano. Por exemplo, a série da figura acima, composta de 15 anos completos e periodicidade mensal (igual a 12), terá um total de 14 índices sazonais atribuídos a cada um dos 12 meses específicos do ano. Em outras

palavras, haverá um índice sazonal aditivo para cada mês dos 15 anos, porém, como há a perda de um ano (6 meses iniciais e 6 meses finais da série), haverá 14 índices sazonais aditivos para cada mês específico do ano. Faz-se então necessária a escolha de um representante do índice sazonal para cada período específico, isto é, um único representante para cada um dos 12 meses do ano. No processo de Recomposição e Previsão, estes índices serão repetidos para cada ano presente na série.

No modelo aditivo, o processo de escolha de um único índice sazonal associado a cada um dos períodos específicos é realizado através de uma média aritmética simples entre os $n-1$ valores fornecidos pela fórmula (1). Em seguida, são somados os valores obtidos para cada representante, isto é, são somados os valores obtidos pelo cálculo da média aritmética simples aplicada a cada período sazonal. Como 0 (zero) é o valor neutro da adição, espera-se que esse somatório produza o resultado 0 (zero), indicando, assim, que a sazonalidade está contida dentro do ano e os valores obtidos para cada índice sazonal estão corretos. Normalmente, esse resultado não é encontrado de imediato; geralmente observa-se uma diferença (positiva ou negativa), que deverá ser removida para que um novo somatório produza o valor 0 (zero). A diferença a ser removida é calculada através da razão entre o somatório e o valor da periodicidade. Uma vez obtida a diferença, essa é subtraída do valor de cada índice sazonal (aqueles obtidos através da média aritmética simples) e um novo somatório desses valores é produzido. Esse processo é repetido até que o somatório resulte no valor 0 (zero). A figura 30 mostra o fluxograma desse procedimento.

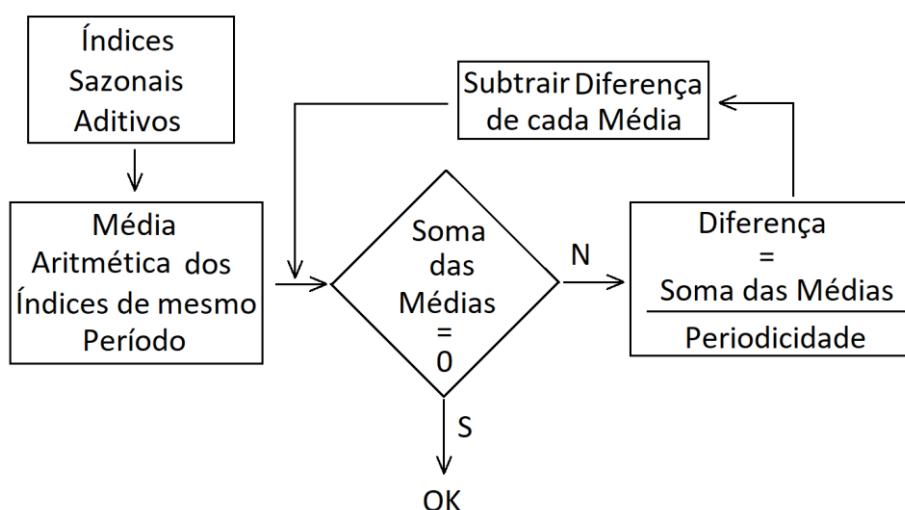


Figura 30 – Determinação dos Índices Sazonais Aditivos.

A tabela abaixo mostra os representantes dos Índices Sazonais Aditivos para cada mês do ano (cada período específico no ano). A coluna BD armazena todos os índices sazonais obtidos através da fórmula (1). Vale observar a ausência dos 6 primeiros valores devido à própria dinâmica do cálculo da média móvel centrada. A coluna BF apresenta a média para cada período específico (que se repete a cada ano) e a coluna BG mostra os valores dos índices após a correção. Apesar de não apresentado na tabela abaixo, os 6 últimos períodos correspondentes à coluna BD também não apresentam valores para o índice sazonal aditivo, visto que esses períodos não possuem média móvel centrada a eles atribuídos.

Tabela 8 – Representante de índices Sazonais Aditivos para cada um dos 12 períodos específicos.

BC	BD	BE	BF	BG
Mês/Ano	Índice Sazonal Aditivo	Meses	Médias Simples Índices Aditivos	Índices Sazonais Aditivos Corrigidos
jan/1990		Janeiro	0,504440476	0,501827877
fev/1990		Fevereiro	-3,552407738	-3,555020337
mar/1990		Março	-13,40061607	-13,40322867
abr/1990		Abri	-12,57394345	-12,57655605
mai/1990		Maio	-17,6445119	-17,6471245
jun/1990		Junho	-5,768693452	-5,771306052
Jul/1990	-4,34775	Julho	-2,77483631	-2,777448909
ago/1990	2,18075	Agosto	2,880892857	2,878280258
set/1990	5,817625	Setembro	7,295223214	7,292610615
out/1990	8,89325	Outubro	11,82223214	11,81961954
nov/1990	10,87333333	Novembro	12,1768631	12,1742505
dez/1990	21,264	Dezembro	21,06670833	21,06409573
jan/1991	2,012291667	Soma = 0,03135119		0,0000
fev/1991	-1,8305	Diferença = 0,002612599		
mar/1991	-13,0515			
abr/1991	-16,24579167			

2.16.2.2 – MODELO MULTIPLICATIVO

No modelo multiplicativo, somente a componente de tendência possui a mesma unidade da série, as demais componentes são índices que influenciam a tendência. Como o elemento neutro da multiplicação é 1 (um), então podemos considerar que os índices com esse valor ou próximos não exercem influência sobre a série. Ao passo que índices substancialmente diferentes de 1 (positivos ou negativos) indicam influência da componente sazonal na série.

$$Y = T \times C \times S \times I \rightarrow S = Y / (T \times C \times I)$$

Assim, a componente sazonal é o resultado do valor da série dividido pelo produto das demais componentes. Levando-se em conta as considerações acima sobre média móvel centrada, conclui-se que o índice sazonal multiplicativo associado a cada elemento da série (que possua média móvel centrada) é obtido dividindo-se o valor desse elemento pelo valor da média móvel centrada a ele associado.

$$\text{Índice Sazonal Multiplicativo} = Y / \text{Média Móvel Centrada} \quad (2)$$

De modo semelhante ao modelo aditivo, teremos associado a um mesmo período específico vários índices sazonais multiplicativos, sendo necessária a definição de um representante para cada um desses períodos. Como no modelo multiplicativo os índices multiplicam o valor da tendência, os extremos podem ser valores atípicos e influenciarem a tendência de forma incorreta. Nesses casos, para definir-se um representante, costuma-se utilizar a média interna desconsiderando os valores extremos (maior e menor). Uma alternativa plausível seria a utilização da mediana ao invés da média interna. Em seguida são somados todos os valores obtidos para cada representante, isto é, são somados os valores obtidos no cálculo da média interna ou mediana. Como 1 (um) é o valor neutro da multiplicação, espera-se que esse somatório produza exatamente o valor da periodicidade, indicando assim que a sazonalidade está contida dentro do ano e os valores obtidos para cada índice sazonal multiplicativo estão corretos. Normalmente, esse resultado também não é encontrado de imediato, geralmente observa-se uma diferença (positiva ou negativa) que deverá ser removida para que o novo somatório produza o valor da periodicidade. Essa diferença é obtida calculando-se primeiramente a diferença entre o somatório dos índices e o valor da periodicidade, para em seguida dividir o resultado pela periodicidade. Vale observar que no modelo aditivo não é realizado essa diferenciação (entre o somatório e a periodicidade), uma vez que o neutro da adição é zero, por isso divide-se direto o valor do somatório pela periodicidade.

No modelo multiplicativo, a correção dos índices acontece através do produto de um fator por cada índice multiplicativo. Esse fator é calculado subtraindo-se 1 (um) do valor da diferença anteriormente calculada. Uma vez obtido o fator, esse é multiplicado por cada índice sazonal (aqueles obtidos através da média interna ou mediana) e um novo somatório desses valores é produzido. Esse processo é repetido até a obtenção do valor da periodicidade como

resultado do somatório de cada índice sazonal. A Figura 31 ilustra o fluxograma desse procedimento.

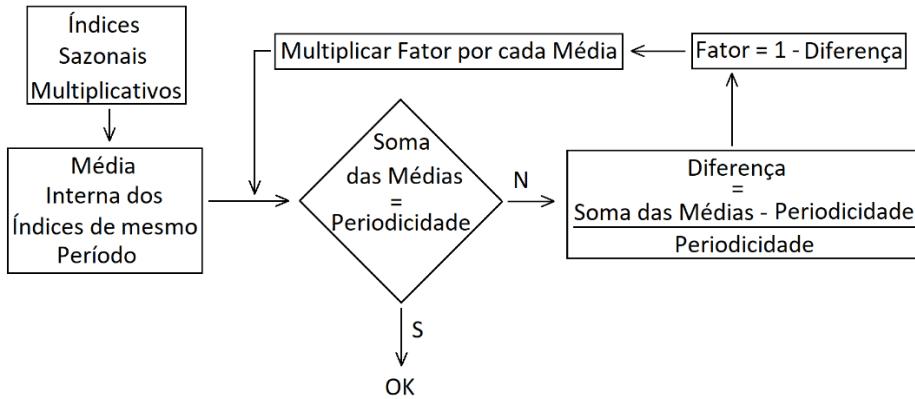


Figura 31 – Determinação dos Índices Sazonais Multiplicativos.

A Tabela 9 mostra os representantes dos Índices Sazonais Multiplicativos para cada mês do ano (cada período específico no ano). A coluna BJ armazena todos os índices sazonais obtidos através da fórmula (2). A coluna BL apresenta a média interna para cada período específico e a coluna BM mostra os valores dos índices após a correção. Cabe observar que o valor esperado para a soma dos índices multiplicativos corresponde à periodicidade, nesse caso exemplo igual a 12.

Tabela 9 – Representante de índices Sazonais Multiplicativos para cada um dos 12 períodos específicos.

BI	BJ	BK	BL	BM
Mês/Ano	Índice Sazonal Multiplicativo	Meses	Médias Internas Índices Multiplicativos	Índices Sazonais Multiplicativos Corrigidos
jan/1990		Janeiro	1,003629297	1,005300078
fev/1990		Fevereiro	0,915604832	0,917129076
mar/1990		Março	0,710515193	0,711698016
abr/1990		Abril	0,696537144	0,697696697
mai/1990		Maio	0,609148403	0,610162477
jun/1990		Junho	0,860084055	0,861515871
Jul/1990	0,894714219	Julho	0,932683545	0,93423622
ago/1990	1,051749812	Agosto	1,041728977	1,043463185
set/1990	1,137135769	Setembro	1,150567624	1,15248302
out/1990	1,209234479	Outubro	1,286717871	1,288859921
nov/1990	1,254557373	Novembro	1,281962934	1,284097069
dez/1990	1,492701237	Dezembro	1,490843247	1,493325113
jan/1991	1,046214205	SOMA 11,98002312		12,0000
fev/1991	0,958414267	DIFERENÇA -0,00166474		
mar/1991	0,709576208	FATOR 1,00166474		
abr/1991	0,647419209			

2.16.3 – OBTENÇÃO DAS COMPONENTES CICLO E IRREGULAR (CI)

A componente ciclo tem atuação no longo prazo, causada por alternâncias sistemáticas de valores de alta e baixa no decorrer da série e a componente irregular retrata alterações fortuitas e de previsão impossível. Costuma-se estudar as variações cíclicas e irregulares em conjunto, apesar de alguns autores optarem por estudar as variações cíclicas juntamente com a tendência.

No modelo aditivo, os índices de ciclo e irregular podem ser expressos como:

$$CI = Y - T - S \quad \text{ou} \quad (3)$$

$$CI = Y - T \quad \text{(para séries anuais)}$$

e no modelo multiplicativo como:

$$CI = Y / (T \times S) \quad \text{ou} \quad (4)$$

$$CI = Y / T \quad \text{(para séries anuais)}$$

A avaliação da influência da componente ciclo e irregular (CI) é realizada através da análise do gráfico gerado a partir dos índices de CI calculados pelas equações 3 ou 4 para cada período da série, seja no modelo aditivo ou multiplicativo. Haverá influência da componente CI (ciclo e irregular) sobre a série caso ocorra uma alternância sistemática de intervalos de alta e baixa com duração superior a um ano para cada um desses intervalos. No modelo aditivo, os padrões de alta ou baixa são aqueles com valores dos índices de CI acima ou abaixo de 0 (zero), respectivamente e no modelo multiplicativo esses padrões são caracterizados por valores dos índices de CI acima ou abaixo de 1 (um).

A Figura 32 apresenta o gráfico dos índices de CI calculados para o modelo aditivo da série vendas.

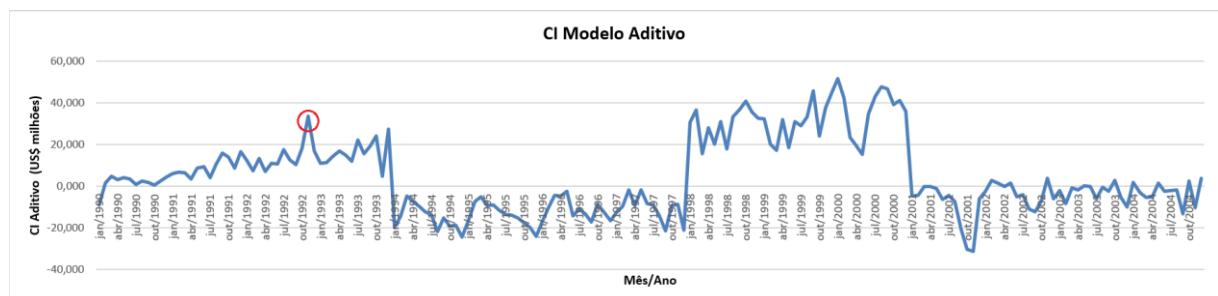


Figura 32 – Plotagem dos índices de CI calculados no modelo aditivo.

No gráfico acima, observa-se uma alternância de intervalos com valores maiores e menores que 0 (zero) e com duração superior a um ano, caracterizando assim a presença de ciclos na série. O primeiro intervalo observado, de alta, estende-se de 1990 a 1993 seguido por um intervalo de baixa que abrange os anos de 1994 a 1997. A seguir, observa-se um novo intervalo de alta acontecendo entre os anos 1998 e 2000 e novamente seguido por um outro de baixa entre os anos 2001 e 2004. Por tratar-se do modelo aditivo, a alternância ocorre em torno de 0 (zero), visto que zero é o elemento neutro na adição. Cabe ressaltar que o eixo das ordenadas apresenta a unidade da série, pois no modelo aditivo todas as componentes possuem a mesma unidade da série. Observa-se também a presença de valores discrepantes, tornando-os assim candidatos a componente irregular, como por exemplo o ponto circundado em vermelho.

A Figura 33 refere-se aos índices de CI calculados pelo modelo multiplicativo e, como pode ser observado, a alternância ocorre em torno de 1 (um). Vale observar também que nesse caso a ordenada não apresenta a unidade pois, no modelo multiplicativo, somente a componente tendência possui a mesma unidade da série.

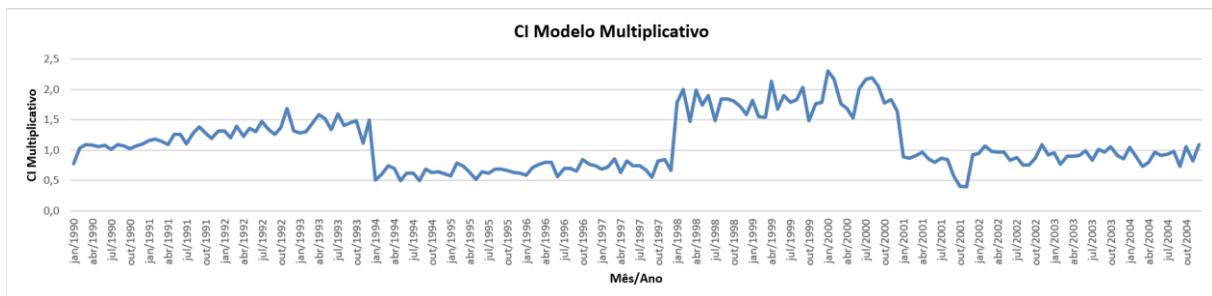


Figura 33 – Plotagem dos índices de CI calculados no modelo multiplicativo.

Na prática, a identificação de ciclos pode não ficar tão facilmente evidenciada como nos gráficos acima. Diversas situações podem levar a essa dificuldade, por exemplo, uma série que não contenha um número de anos suficiente para distinguir sua alternância (como foi constatado com a massa de dados disponibilizada para este trabalho).

Após o cálculo dos índices de CI (através das equações 3 ou 4) para cada período da série e após a identificação de alternância dos intervalos de alta e baixa (através do gráfico), faz-se necessário determinar índices cíclicos que representem tais intervalos. Nesses termos, ter-se-á

um único representante CI para todos os intervalos de alta e um único representante CI para todos os intervalos de baixa. O representante CI de alta e de baixa são calculados através da mediana de todos os índices de CI pertencentes a cada um desses intervalos. Esse processamento é realizado tanto para o modelo aditivo quanto para o multiplicativo, de modo que haverá um único representante CI de alta e outro de baixa em cada modelo. Os valores dos representantes CIs, assim calculados, serão posteriormente usados nos processos de recomposição. Para o processo de previsão, os representantes CIs de alta e baixa também serão calculados através da mediana, porém considerando-se apenas o último intervalo de alta e baixa presentes na série. O fundamento dessa restrição (somente o último intervalo) está amparado na suposição que esses exercem maior influência sobre os intervalos seguintes, se comparado à influência que os intervalos anteriores exercem sobre os intervalos futuros.

Sendo assim, tanto o modelo aditivo quanto o multiplicativo terão um representante CI de alta e outro de baixa para a recomposição e um representante CI de alta e outro de baixa para a previsão, totalizando quatro representantes CIs para cada modelo.

2.16.4 – PROCESSO DE RECOMPOSIÇÃO

Após a determinação dos valores representativos de cada componente da série original (Tendência, Ciclo, Sazonalidade e Irregular), a próxima etapa consiste na recomposição de uma nova série através desses valores. Esse processo é realizado para ambos os modelos, aditivo e multiplicativo, de forma que ao final haverá duas novas séries recompostas (uma pelo modelo aditivo e outra pelo multiplicativo). A determinação de qual modelo é o mais adequado para representar a série original será apontada utilizando-se as métricas de acuracidade definidas em 2.16.1.1.

2.16.4.1 – RECOMPOSIÇÃO PELO MODELO ADITIVO

A recomposição através do modelo aditivo é realizada somando-se os valores das componentes representativas calculadas para esse modelo, $= CJ2 + CK2 + CL2$, como mostrado na barra de funções da Tabela 11.

Para a componente Tendência, não há distinção entre os modelos aditivos e multiplicativos e sua determinação é dada através da métrica de acuracidade (entre as equações

de regressão), isto é, aquela que apresentar os valores mais próximos de zero. Já para os representantes de índices Sazonais e representantes CIs, verifica-se uma distinção entre os modelos. O conjunto de 12 representantes dos Índices Sazonais Aditivos (coluna BG da Tabela 8), atribuídos para série original, serão os mesmos usados no processo de Recomposição para cada ano. Na Tabela 10, observa-se a repetição dos representantes dos Índices Sazonais Aditivos (coluna BR) para cada período específico do ano.

Tabela 10 – Os 12 representantes do índice sazonal aditivo (coluna BR) repetidos para cada ano.

	BO	BP	BQ	BR
1	Mês/Ano	Vendas	T (exponencial)	12 índices Saz. Adt. Corr. (repetido p/ cada ano)
2	jan/1990	30,427	39,0029501	0,501827877
3	fev/1990	37,023	39,0049003	-3,555020337
4	mar/1990	30,281	39,00685059	-13,40322867
5	abr/1990	29,485	39,00880098	-12,57655605
6	mai/1990	25,317	39,01075147	-17,6471245
7	jun/1990	36,525	39,01270206	-5,771306052
8	jul/1990	36,947	39,01465274	-2,777448909
9	ago/1990	44,321	39,01660352	2,878280258
10	set/1990	48,240	39,0185544	7,292610615
11	out/1990	51,397	39,02050538	11,81961954
12	nov/1990	53,588	39,02245645	12,1742505
13	dez/1990	64,422	39,02440762	21,06409573
14	jan/1991	45,555	39,02635889	0,501827877
15	fev/1991	42,187	39,02831026	-3,555020337
16	mar/1991	31,888	39,03026172	-13,40322867
17	abr/1991	29,831	39,03221328	-12,57655605
18	mai/1991	30,033	39,03416494	-17,6471245
19	jun/1991	42,449	39,0361167	-5,771306052
20	jul/1991	40,256	39,03806855	-2,777448909
21	ago/1991	52,407	39,04002051	2,878280258
22	set/1991	62,282	39,04197256	7,292610615
23	out/1991	64,650	39,0439247	11,81961954
24	nov/1991	59,751	39,04587695	12,1742505
25	dez/1991	76,736	39,04782929	21,06409573
26	jan/1992	51,725	39,04978173	0,501827877
27	fev/1992	42,979	39,05173427	-3,555020337
28	mar/1992	38,855	39,05368691	-13,40322867

Os representantes CI de alta e baixa a serem usados na Recomposição serão aqueles correspondentes ao intervalo de recomposição corrente. A Tabela 11 apresenta os valores de CIs calculados para Recomposição e Previsão nos modelos aditivos e multiplicativos (coluna CG). Cabe observar que devido ao fato de os primeiros anos da série original pertencerem ao intervalo de alta, então, na recomposição desses anos, o valor do representante CI será o de alta (16,977) e não o de baixa (-7,840).

Tabela 11 – Componentes T, representantes S e CI calculadas para a recomposição no modelo aditivo.

	CF	CG	CH	CI	CJ		CK	CL	CM
	CI Aditivo			Mês/Ano	T (exponencial)	12 índices Saz. Adt. Corr. (repetido p/ cada ano)	CI Aditivo	Recomposição: Y Aditivo	
1	CI Aditivo			jan/1990	39,0029501	0,501827877	16,779	56,284	
2	Recomposição			fev/1990	39,0049003	-3,555020337	16,779	52,22893758	
3	Alta	16,779		mar/1990	39,00685059	-13,40322867	16,779	42,38267954	
4	Baixa	-7,840		abr/1990	39,00880098	-12,576555605	16,779	43,21130255	
5	Previsão			mai/1990	39,01075147	-17,6471245	16,779	38,14268459	
6	Alta	32,814		jun/1990	39,01270206	-5,771306052	16,779	50,02045363	
7	Baixa	-2,945		jul/1990	39,01465274	-2,777448909	16,779	53,01626145	
8				ago/1990	39,01660352	2,878280258	16,779	58,6739414	
9	CI Multiplicativo			set/1990	39,0185544	7,292610615	16,779	63,09022264	
11	Recomposição			out/1990	39,02050538	11,81961954	16,779	67,61918254	
12	Alta	1,473193		nov/1990	39,02245645	12,1742505	16,779	67,97576457	
13	Baixa	0,777217		dez/1990	39,02440762	21,06409573	16,779	76,86756098	
14	Previsão			jan/1991	39,02635889	0,501827877	16,779	56,30724439	
15	Alta	1,803137		fev/1991	39,02831026	-3,555020337	16,779	52,25234754	
16	Baixa	0,907066		mar/1991	39,03026172	-13,40322867	16,779	42,40609067	
17				abr/1991	39,03221328	-12,576555605	16,779	43,23471485	
18				mai/1991	39,03416494	-17,6471245	16,779	38,16609806	
19				jun/1991	39,0361167	-5,771306052	16,779	50,04386827	
20				jul/1991	39,03806855	-2,777448909	16,779	53,03967727	
21				ago/1991	39,04002051	2,878280258	16,779	58,69735839	
22				set/1991	39,04197256	7,292610615	16,779	63,11364079	
23				out/1991	39,0439247	11,81961954	16,779	67,64260187	
24				nov/1991	39,04587695	12,1742505	16,779	67,99918507	
25				dez/1991	39,04782929	21,06409573	16,779	76,89098265	
26				jan/1992	39,04978173	0,501827877	16,779	56,33066723	
27				fev/1992	39,05173427	-3,555020337	16,779	52,27577155	
28				mar/1992	39,05368691	-13,40322867	16,779	42,42951586	

2.16.4.2 – RECOMPOSIÇÃO PELO MODELO MULTIPLICATIVO

A recomposição através do modelo multiplicativo é realizada multiplicando-se os valores das componentes representativas para esse modelo. Os detalhes relativos às componentes descritas acima para o modelo aditivo aplicam-se igualmente ao modelo multiplicativo.

2.16.5 – PROCESSO DE PREVISÃO

A previsão de valores futuros é realizada sobre a nova série (modelo aditivo ou multiplicativo) identificada como a melhor representante da série original. Ao estabelecer-se um tempo futuro a ser previsto, ficam então definidos os períodos a serem considerados pela nova série. Assim, caso o modelo aditivo seja aquele que melhor represente a série original, então as suas componentes calculadas deverão ser somadas e, caso seja o multiplicativo, as suas componentes deverão ser multiplicadas.

A componente tendência usará esses novos períodos (alvo da previsão) no seu processamento. O valor do representante de índice de sazonalidade a ser utilizado será aquele que corresponda ao período a ser previsto e o valor do representante CI será aquele que corresponda ao intervalo (alta ou baixa) que o novo período se encaixe.

Por exemplo, a série da Figura 27 composta por 15 anos completos e registrada com periodicidade mensal, corresponde a 180 períodos. Considerando que o modelo multiplicativo foi o modelo escolhido para representação da série original e que se deseja fazer uma previsão de 6 meses, então, as componentes futuras a serem multiplicadas serão:

- *Tendência:* Dentre as equações analisadas, a exponencial ($y = 39,001 e^{5E-05X}$) foi a que melhor descreve a série, onde X representa o período. A última data registrada na série original foi dezembro de 2004, que corresponde ao período 180 (15x12). Logo, a primeira data futura, janeiro de 2005, corresponderá ao período 181 e a última data de previsão desejada (junho de 2005) corresponderá ao período 186.
- *Índices de Sazonalidade:* Foram estabelecidos 12 representantes de índices de sazonalidade para o modelo multiplicativo, um para cada mês específico existente no ano. Como a previsão desejada engloba 6 meses, os índices a serem usados serão aqueles referentes aos meses de janeiro a junho.
- *Representante CI:* A Figura 33 (gráfico de CI do modelo multiplicativo) mostra que existe a influência de ciclos na série e a expectativa é que o próximo intervalo, a ser iniciado em janeiro de 2005 e finalizado em dezembro de 2008, seja de alta. Como a previsão desejada envolve 6 meses de 2005, então o valor do representante de CI correspondente será o de alta.

2.17 – MÉTODOS DE SUAVIZAÇÃO EXPONENCIAL

A suavização exponencial foi proposta na década de 1950 por Brown e, no decorrer do tempo, novas abordagens e aprimoramentos foram acrescentados. Conforme descrito anteriormente, a análise de séries temporais visa o entendimento da dependência de seus elementos em relação ao tempo com o objetivo de realizar uma previsão de valores futuros. A maioria dos métodos e modelos de previsão são baseados no conceito que as observações passadas guardam informações sobre o padrão de comportamento da série. Então, o objetivo precípua dos diversos métodos consiste na identificação do padrão do ruído que porventura exista em suas observações.

Os modelos de suavização exponencial assumem que os valores extremos de uma série se apresentam de forma aleatória e, através da suavização destes, é possível a identificação do padrão básico de modo a utilizá-lo para fins de previsão. Segundo Hyndman *et al.* (2018) as previsões produzidas usando métodos de suavização exponencial são médias ponderadas de observações anteriores, com os pesos decaíndo exponencialmente à medida que as observações envelhecem. Ou seja, quanto mais recente a observação, maior o peso associado.

Os métodos de suavização exponencial podem ser classificados em função das características da série temporal a serem aplicados, a saber, séries que apresentam somente nível, séries que possuem nível e tendência e, finalmente, as séries que portam nível, tendência e sazonalidade. Assim sendo, o método de suavização simples (SES) é aplicado às séries que só possuem nível, isto é, não apresentam tendência nem sazonalidade. O segundo método é chamado de Suavização Exponencial Corrigido pela Tendência ou Suavização Exponencial Dupla ou ainda, como é mais conhecido, método de Holt, é utilizado para previsão de séries temporais que apresentam nível e tendência. O terceiro método, denominado Suavização Exponencial Corrigido pela Tendência e Sazonalidade ou Suavização Exponencial Tripla ou, como habitualmente chamado, método de Holt-Winters, é empregado para previsão de séries que apresentam nível, tendência e sazonalidade.

2.17.1 - MÉTODO DE SUAVIZAÇÃO EXPOENCIAL SIMPLES (SES)

Para utilização deste método, parte-se da premissa que a série não possui tendência nem sazonalidade, somente nível. A ausência de tendência inviabiliza qualquer suposição sobre qual possível direção que a série tomará numa observação futura. Nesse caso, a série Y_t pode ser descrita como:

$$Y_t = \mu_t + \varepsilon_t$$

Onde μ_t é a média do processo no tempo t e ε_t é uma perturbação (ou ruído) aleatória e independente com média zero e variância constante σ^2 , ou seja, $\varepsilon_t \sim \text{iid}$, cuja notação é: $\varepsilon_t \sim N(0 ; \sigma^2)$. Assume-se a_t como sendo uma estimativa de μ_t então, a_t será chamada de nível da série temporal. Sendo assim, o nível da série é uma estimativa da média, isto é, uma estimativa da média entre o tempo atual e o próximo a ser previsto.

O nível da série temporal a_t é definido como a média ponderada entre o valor atual da série Y_t e o valor anterior do nível a_{t-1} , da seguinte forma:

$$a_t = \alpha Y_t + (1 - \alpha)a_{t-1} \quad \text{onde: } 0 \leq \alpha \leq 1$$

Se α tende a 1 (um), então haverá pouca suavização e o nível a_t acompanhará “de perto” a série Y_t . Esse é o caso quando a variância do ruído σ^2 é muito menor que as mudanças na média da série, ou seja, a influência (ou peso) do ruído sobre a série como um todo será pequeno. Se α tende a 0 (zero), então haverá suavização mais incisiva ou, em outras palavras, pouca influência da observação atual na estimativa da média. Este é o caso quando a variância do ruído σ^2 é muito maior que as mudanças na média da série temporal. Desta forma, α determina o grau de suavização da série temporal e a tarefa de ajuste do método SES consiste em adequar o valor de α a fim de reduzir os erros.

Normalmente, atribui-se um valor inicial para α igual a 0.2 quando não se tem nenhuma informação sobre a série, pois considera-se que a maioria das séries possuem uma mudança na média entre $t-1$ e t menor que a variância do ruído σ^2 . Em outras palavras, a maioria das séries possui uma mudança no valor da média entre dois tempos consecutivos menor que o valor da variância do ruído σ^2 .

Em resumo, baixos valores de α resultam em um alisamento mais acentuado e valores mais elevados de α resultam em alisamentos mais suaves, isto é, valores mais elevados de α dão peso maior para as observações mais recentes. Desta forma, pesos próximos a zero são utilizados para suavizar componentes irregulares e ciclos indesejáveis. Assim, a previsão do h -ésimo valor futuro é obtida a partir da equação:

$$\hat{Y}_{t+h} = a_t \quad \text{onde } h = 1, 2, \dots$$

2.17.2 - MÉTODO DE HOLT

Em 1957, Holt estendeu o método SES a fim de permitir previsão em séries temporais que possuam componente de tendência. Esse método envolve a equação de previsão e outras duas equações, uma para o nível e a outra para a tendência. Dessa forma, o método introduz um novo parâmetro (β) com a finalidade de suavizar esta componente.

Uma série possuindo a componente tendência pode ser descrita como $Y_t = \mu_t + T_t + \varepsilon_t$. Onde μ_t é a média (ou nível) da série, T_t representa a componente tendência no tempo t e ε_t é uma perturbação (ou ruído) aleatória e independente, com média zero e variância σ^2 constante, ou seja, $\varepsilon_t \sim \text{iid}$, cuja notação é $\varepsilon_t \sim N(0 ; \sigma^2)$.

Equação da estimativa do nível:

$$a_t = \alpha Y_t + (1 - \alpha)(a_{t-1} + T_{t-1}) \quad \text{onde } 0 \leq \alpha \leq 1$$

Equação da tendência:

$$T_t = \beta(a_t - a_{t-1}) + (1 - \beta)T_{t-1} \quad \text{onde } 0 \leq \beta \leq 1$$

Equação da previsão:

$$\hat{Y}_{t+h} = a_t + hT_t \quad \text{onde } h = 1, 2, \dots$$

2.17.3 - MÉTODO DE HOLT-WINTERS

A fim de agregar a característica de sazonalidade, Winters aprimorou o método de Holt e esse é chamado de método de Holt-Winters. O método Holt-Winters pode ser classificado como aditivo ou multiplicativo. A opção aditiva é mais adequada quando as variações sazonais são aproximadamente constantes ao longo da série e a opção multiplicativa é preferível quando a série experimenta variações sazonais que mudam proporcionalmente ao nível da série.

O método de Holt-Winters, também chamado de suavização exponencial tripla, foi desenvolvido a partir do método de Holt e esse a partir do método de suavização exponencial simples (SES) sendo descrito como:

$$a_t = \alpha Y_t + (1 - \alpha) \cdot a_{t-1} \quad \text{com } 0 \leq \alpha \leq 1$$

O nível da série temporal a_t é definido como a média ponderada entre o valor atual da série Y_t e o valor anterior do nível da série a_{t-1} , desprezando-se a tendência e o efeito sazonal. Entretanto, o mais comum é que as séries possuam tendência e efeito sazonal. Para atender tal demanda, os métodos SES e Holt foram expandidos para o que ficou conhecido como modelo Holt-Winters, da seguinte forma:

$$a_t = \alpha \cdot (Y_t - S_{t-p}) + (1 - \alpha) \cdot (a_{t-1} + T_{t-1})$$

$$T_t = \beta \cdot (a_t - a_{t-1}) + (1 - \beta) \cdot T_{t-1}$$

$$S_t = \gamma \cdot (Y_t - a_t) + (1 - \gamma) \cdot S_{t-p}$$

Onde T é chamado de inclinação e S é o efeito sazonal. O índice p , que é associado ao efeito sazonal, indica o tamanho da estação ou período sazonal. Com essa consideração o modelo conta com três parâmetros de suavização (α , β e γ). Para utilização das equações acima faz-se necessária uma atribuição para seus valores iniciais e, na maioria das vezes, são escolhidos $a_1 = Y_1$, $T_1 = 0$ e S_1 igual à média das observações de uma estação.

A previsão com o método Holt-Winters para um horizonte de tempo k obedece à seguinte equação:

$$\hat{Y}_{n+k, 1:n} = a_n + k \cdot T_n + S_{n+k-p}$$

Então, a previsão \hat{Y}_{n+k} no tempo n é o nível corrente da série temporal com uma extração linear da sua tendência mais o termo apropriado de efeito sazonal. No software estatístico R os parâmetros desconhecidos são determinados através da minimização do erro quadrático da previsão.

2.18.1 – MODELO AUTORREGRESSIVO - AR(p)

Os modelos autorregressivos vivenciam uma grande evolução a partir dos anos de 1970 com os modelos propostos por Box e Jenkins, que revolucionaram a utilização prática e o estudo de séries temporais. Um modelo autorregressivo (AR), segundo Nielsen (2021), toma como base a intuição de que o passado prediz o futuro. Considerando a série y abaixo no período t ,

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \text{ onde } \varepsilon_t \text{ é a perturbação, } c \text{ e } \phi_n \text{ são constantes}$$

O modelo autorregressivo de ordem p AR(p) é baseado na combinação linear de observações passadas, isto é, y_t pode ser representado pelos valores passados do próprio y como $y_{t-1}, y_{t-2}, \dots, y_{t-p}$, daí a origem do termo autorregressivo. Em resumo, trata-se do passado de y explicando o y atual e p representando o número de valores passados a serem considerados na regressão. Então, a formulação mais natural de um modelo de séries temporais estacionárias é uma regressão linear dos valores passados da própria série. Assim, os modelos autorregressivos devem ser ajustados somente para séries temporais estacionárias.

O termo de perturbação ε_t na expressão acima tem origem num processo de ruído branco gaussiano iid, sendo também estocasticamente independente de $y_{t-1}, y_{t-2} \dots$ e ε_t possui a capacidade de direcionar a série em uma nova direção. Alterando-se os parâmetros ϕ_1, \dots, ϕ_p resulta em diferentes padrões de séries temporais. Por outro lado, a variação do termo de perturbação ε_t mudará somente a escala da série, não os padrões.

A Figura 34 mostra um modelo AR(1) e a Figura 36 um modelo AR(2), gerados através das expressões abaixo:

O modelo AR(1) foi criado com a estrutura: $y_t = 0.8 * y_{t-1} + \varepsilon_t$ onde: $y_1 = \varepsilon_1$.

O modelo AR(2) foi criado com a estrutura: $y_t = 8 + 0.8 * y_{t-1} - 0.1 * y_{t-2} + \varepsilon_t$ onde: $y_1 = \varepsilon_1$ e $y_2 = \varepsilon_2$.

Geração do modelo AR(1):

A série AR(1) utiliza *uma* observação passada (y_{t-1}) no cálculo dos seus elementos e a série AR(2) requer *duas* observações passadas (y_{t-1} e y_{t-2}). Contudo, no início da geração de AR seus valores passados ainda não são conhecidos, então, utiliza-se as primeiras observações do ruído branco para iniciar a série AR(p). Assim, se $p = 1$, o primeiro valor de AR(1) será a primeira observação do ruído branco ($y_1 = \varepsilon_1$) e para $p = 2$ os dois primeiros valores de AR(2) serão os dois primeiros valores do ruído branco ($y_1 = \varepsilon_1$ e $y_2 = \varepsilon_2$). O ruído branco ε_t utilizado para geração dos modelos AR acima é o mesmo gerado na seção 2.12 e trata-se de um ruído branco gaussiano padrão $\sim N(0,1)$, que apresenta média igual a 0.0904 (aproximadamente zero) e desvio padrão igual a 0.9128 (aproximadamente um).

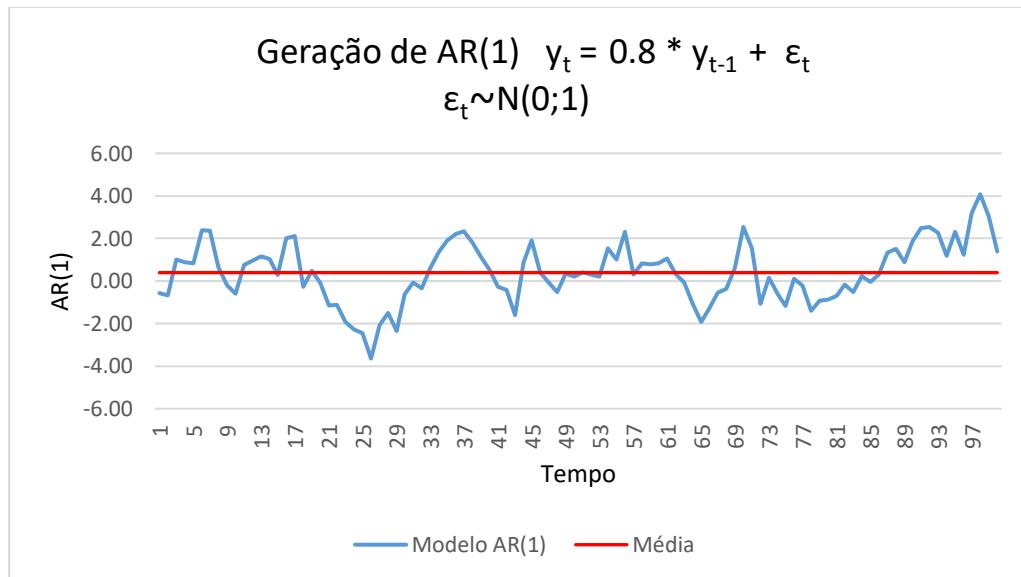


Figura 34 – Modelo AR(1) gerado.

Através do gráfico acima, observa-se que há uma variação cíclica positiva e negativa causada pela perturbação ε_t .

Correlogramas de AR(1):

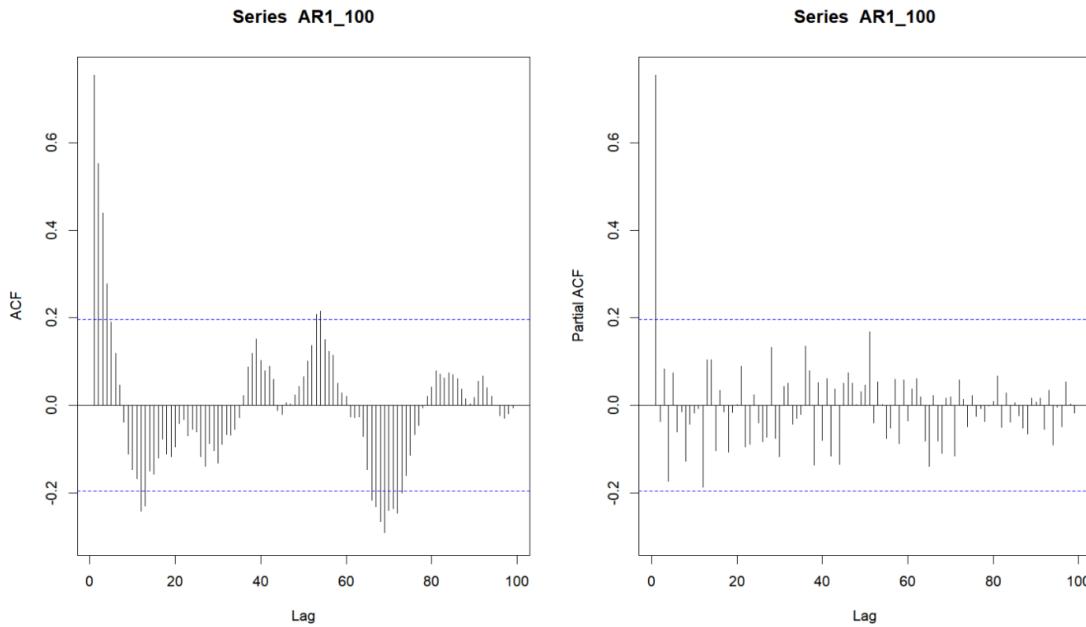


Figura 35 – Correlogramas da FAC e FACP do modelo AR(1) gerado.

Observa-se que no correlograma da FAC do modelo AR(1), Figura 35, o decaimento é exponencial, seguido de uma variação cíclica no interior do intervalo de confiança. Esse decaimento será importante na identificação de modelos (vide Tabela 12). Caso haja uma autorregressão mais complexa, com diversos valores de ϕ ($\phi_1, \phi_2, \phi_3 \dots$) então, o decaimento exponencial no correlograma da FAC não ocorrerá de forma exponencial tão precisa; mesmo assim será aproximadamente exponencial.

Para um modelo AR(1), Hyndman *et al.* (2018) apresenta a seguinte classificação:

- quando $\phi_1 = 0$, y_t é o ruído branco;
- quando $\phi_1 = 1$ e $c = 0$, y_t é um passeio aleatório;
- quando $\phi_1 = 1$ e $c \neq 0$, y_t é um passeio aleatório com *drift*;
- quando $\phi_1 < 0$, y_t tende a oscilar entre valores positivos e negativos.

Normalmente, restringe-se os modelos autorregressivos a dados estacionários e, nesse caso, algumas restrições nos valores dos parâmetros são necessárias.

- Para um modelo AR(1): $-1 < \phi_1 < 1$;
- Para um modelo AR(2): $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$.

Quando $p \geq 3$, as restrições são mais complicadas. Contudo, o *software* R cuida dessas restrições ao estimar o modelo.

Geração do modelo AR(2):

O modelo AR(2), gerado segundo a estrutura definida acima, e seus correlogramas são apresentados abaixo através da Figura 36 e Figura 37, respectivamente.

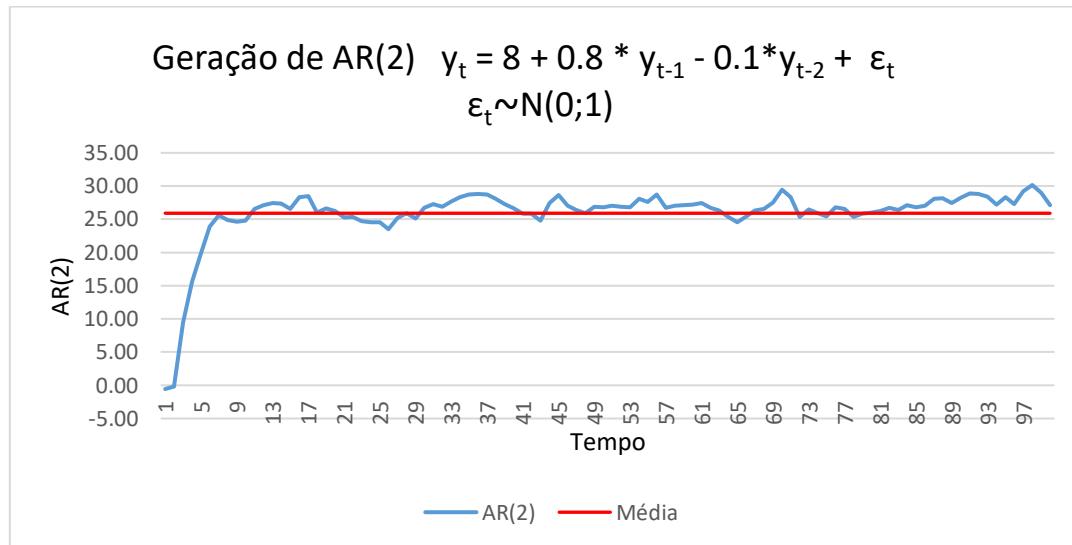


Figura 36 – Modelo AR(2) gerado.

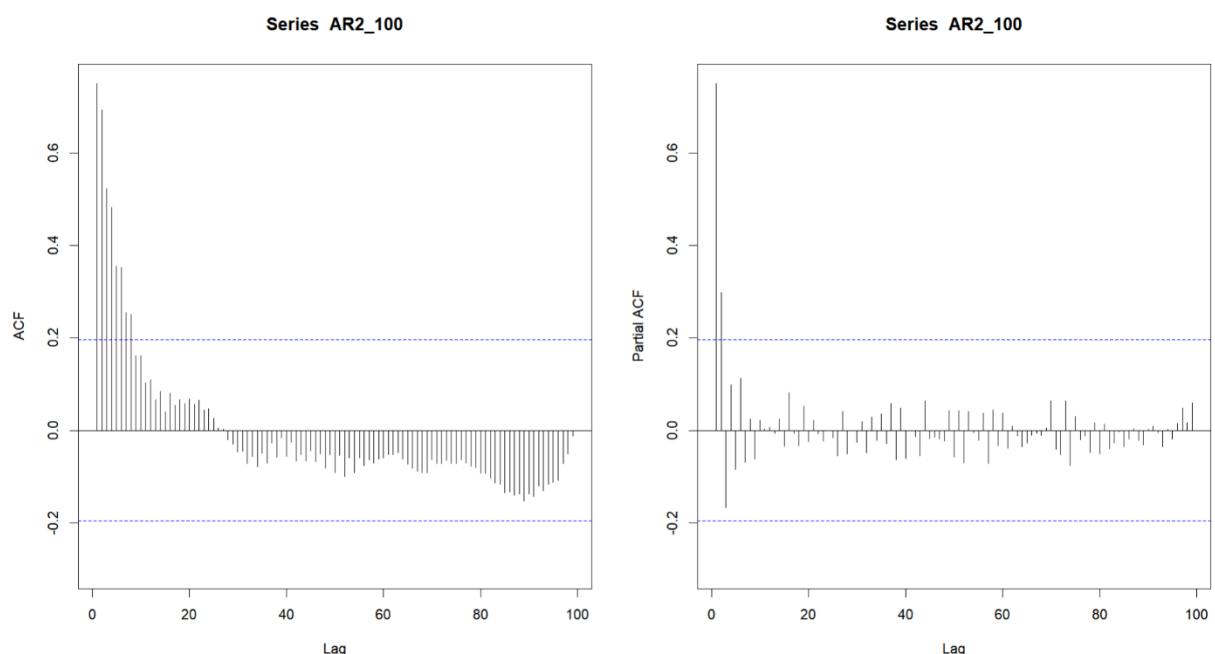


Figura 37 – Correlogramas da FAC e FACP do modelo AR(2) gerado.

2.18.2 – MODELOS DE MÉDIA MÓVEL - MA(q)

Um modelo de média móvel MA é uma combinação linear da perturbação corrente ε_t (ruído branco) adicionada às perturbações passadas $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$. O número de perturbações passadas é definido pela ordem q do modelo. Em outras palavras, é uma extensão do processo de ruído branco.

Cabe chamar a atenção no sentido de não se confundir o modelo autorregressivo de média móvel MA com a técnica de suavização de mesmo nome, o primeiro é usado para previsão de valores futuros e o segundo é utilizado, por exemplo, para determinação de tendência. O modelo autorregressivo MA tem a seguinte forma:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

A Figura 38 mostra um modelo MA(1) e a Figura 40 um modelo MA(2), gerados através das expressões abaixo e as figuras 39 e 41 apresentam seus respectivos correlogramas.

O modelo MA(1) foi criado com a estrutura: $y_t = \varepsilon_t + 0.8 * \varepsilon_{t-1}$ onde: $y_1 = \varepsilon_1$,

O modelo MA(2) foi criado com a estrutura: $y_t = 8 + \varepsilon_t + 0.8 * \varepsilon_{t-1} - 0.1 * \varepsilon_{t-2}$

onde: $y_1 = \varepsilon_1$ e $y_2 = \varepsilon_2$

A título de comparação nos exemplos de AR e MA, foi utilizado o mesmo ruído branco assim como as constantes c, ϕ e θ .

Geração do modelo MA(1):

A série MA(l) utiliza *uma* observação passada do ruído branco (ε_{t-1}) no cálculo dos seus elementos e a série MA(2) *duas* observações passadas do ruído branco (ε_{t-1} e ε_{t-2}). Contudo, no início da geração de MA, os valores passados do ruído branco (ε_{t-1} e ε_{t-2}) não são conhecidos, então, utiliza-se as primeiras observações do ruído branco para iniciar a série MA(q). Assim, se $q = 1$ o primeiro valor de MA(1) será a primeira observação do ruído branco

$(y_1 = \varepsilon_1)$ e para $q = 2$ os dois primeiros valores de MA(2) serão os dois primeiros valores do ruído branco ($y_1 = \varepsilon_1$ e $y_2 = \varepsilon_2$).

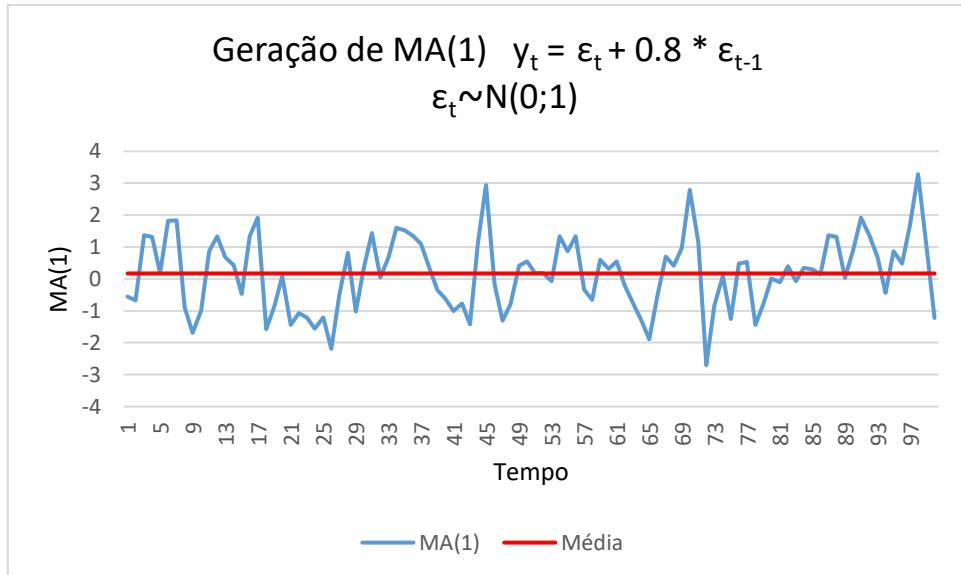


Figura 38 – Modelo MA(1) gerado.

Semelhante ao modelo AR, alterando-se os parâmetros $\theta_1, \dots, \theta_p$ resulta em diferentes padrões de séries temporais, por outro lado, a variação do termo de perturbação ε_t mudará somente a escala da série, não os padrões.

Correlogramas de MA(1):

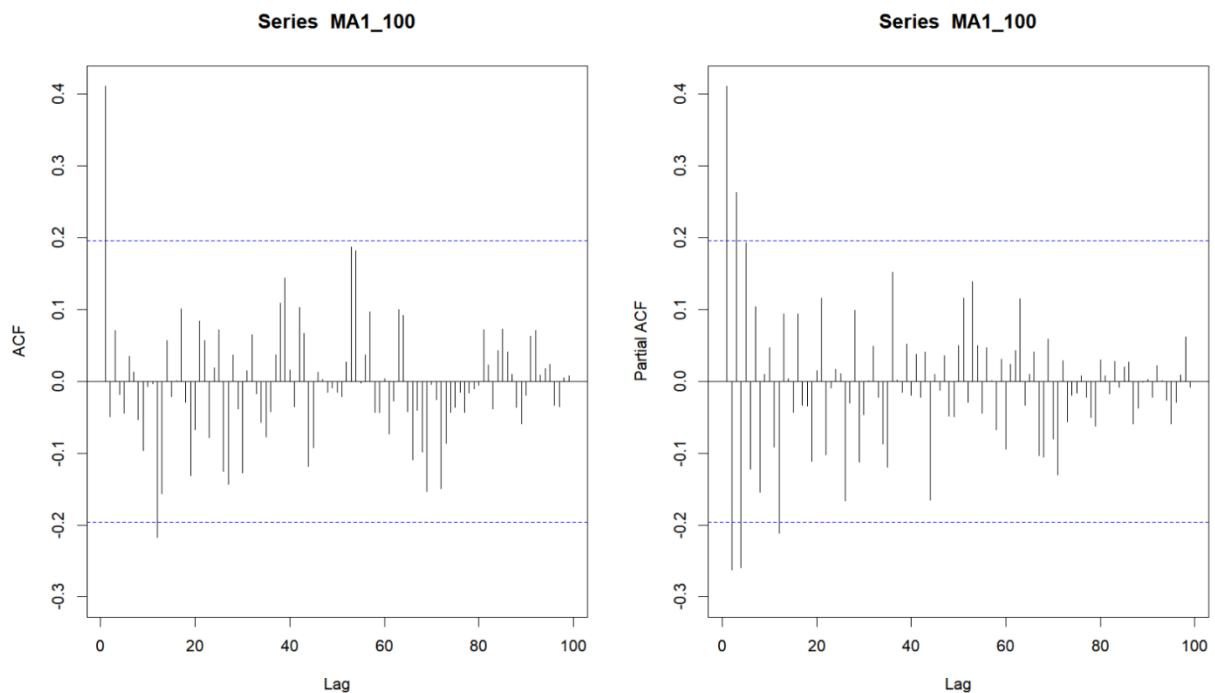


Figura 39 – Correlogramas da FAC e FACP do modelo MA(1) gerado.

Geração do modelo MA(2):

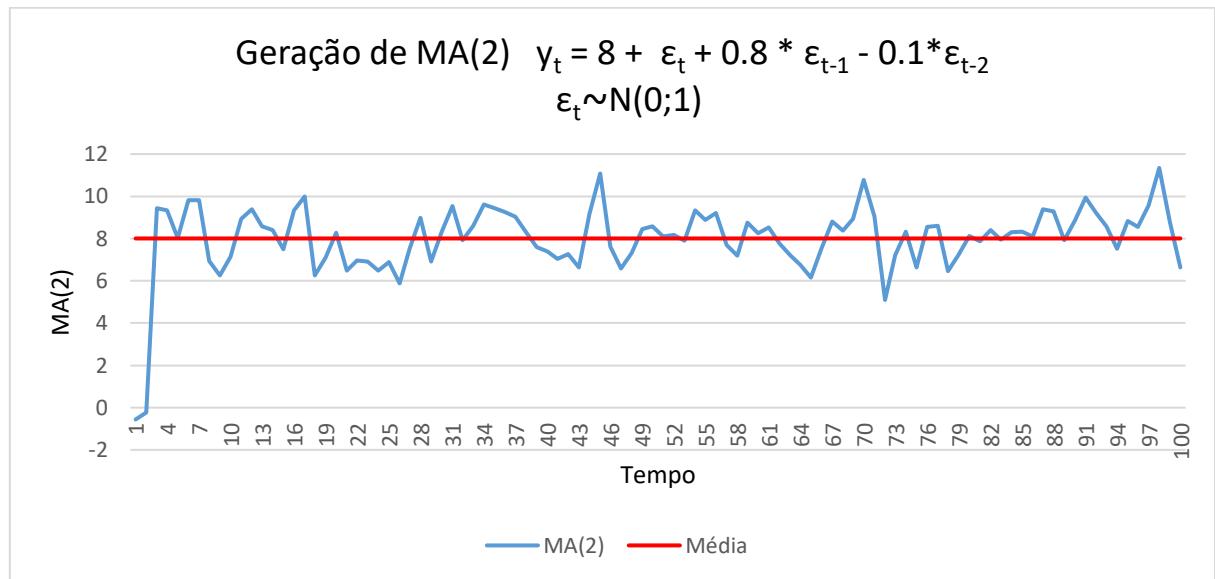


Figura 40 – Modelo MA(2) gerado.

Correlogramas de MA(2):

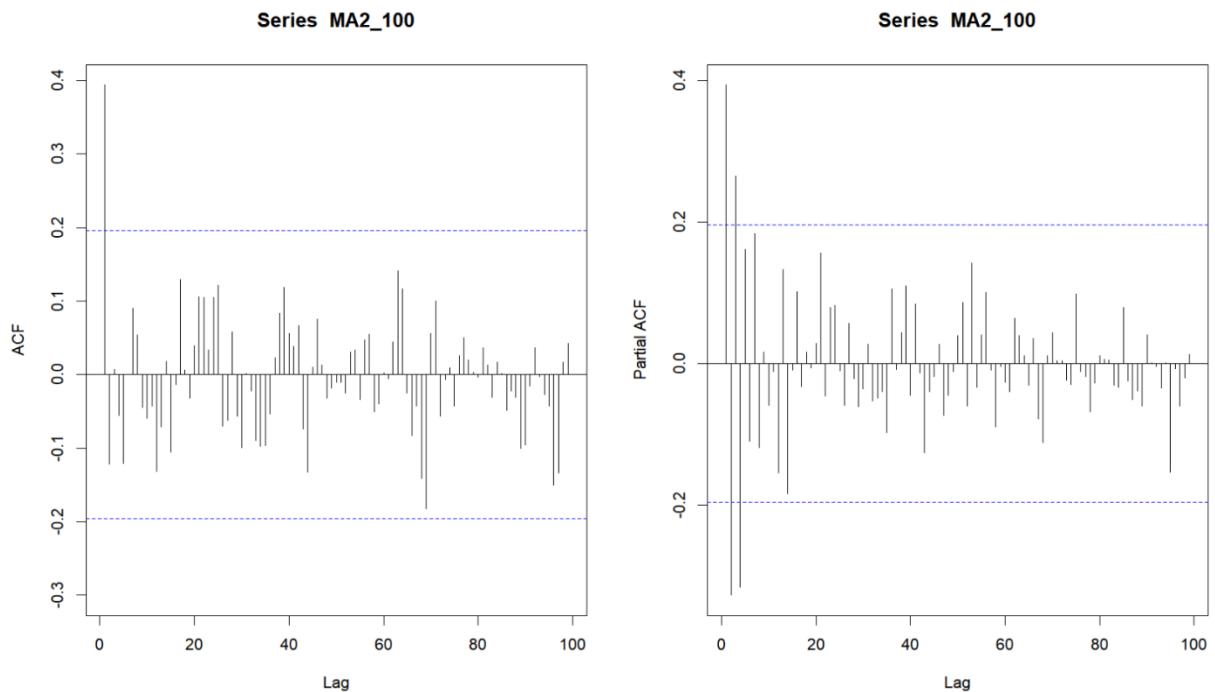


Figura 41 – Correlogramas da FAC e FACP do modelo MA(2) gerado.

Somente através da visualização dos gráficos temporais de AR e MA acima, constata-se que não é trivial inferir algumas características do processo tais como a estacionariedade, o valor de phi (ϕ) ou teta (θ), se é um processo AR ou MA e quais são as suas ordens p e q . Então, necessita-se de algo mais consistente para identificação e interpretação da série. Uma ideia inicial e amplamente divulgada na literatura para tentativa de determinação do modelo e suas ordens é mostrada na Tabela 12 abaixo, que se baseia na análise dos correogramas FAC e FACP, relacionando-os aos modelos AR e MA.

O decaimento exponencial citado na tabela significa que os valores da FAC ou FACP (eixo das ordenadas) vão diminuindo até zero à medida que os *lags* (eixo das abscissas) vão aumentando. O truncamento se refere aos valores dos *lags* na FAC ou FACP que estejam fora do intervalo de confiança, sendo então estes valores os candidatos às ordens p e q .

Tabela 12 – Associação dos modelos AR e MA às FAC e FACP.

Fonte: Bueno *et al.* (2015)

Modelo	FAC	FACP
AR(p)	Decai exponencialmente	Truncada na defasagem p
MA(q)	Truncada na defasagem q	Decai exponencialmente
ARMA(p, q)	Decai exponencialmente se $j > q$	Decai exponencialmente se $j > p$

Onde: $j = N/4$ e N é o tamanho da amostra. Bueno também sugere calcular-se a FACP até j *lags*.

O objetivo da tabela acima é orientar a identificação do modelo e as escolhas de seus respectivos dos valores ordem p ou q . Dessa forma, a tabela acima indica:

Para o modelo AR(p).

- A FAC apresentará um decaimento exponencial desde as defasagens mais recentes e diminuindo até desaparecer nas defasagens mais elevadas, ou seja, mais antigas.
- Na FACP, observa-se um padrão de truncamento na defasagem p . Assim, o foco se dá sobre esse truncamento e o valor do *lag* neste ponto indicará a ordem p .

Para o modelo $MA(q)$.

É o inverso das considerações acima para o modelo $AR(p)$

- A FAC truncando na defasagem q .
- A FACP decaindo exponencialmente.

Para o modelo $ARMA(p,q)$:

Tanto a FAC quanto a FACP apresentam um decaimento exponencial.

O gráfico da FAC mostra o valor da autocorrelação versus o *lag*, onde identifica-se a ordem q do modelo MA (que corresponde ao *lag* que apresenta a truncagem ou corte).

O gráfico da FACP mostra a correlação pura entre duas observações (eliminando as correlações implícitas), no qual identifica-se a ordem p do modelo AR (que corresponde ao *lag* que apresenta a truncagem ou corte). Em outras palavras, a correlação pura se refere à correlação entre t e $t-p$, sem considerar as correlações intermediárias – identificação de p para o modelo AR .

A tabela 12 acima, amplamente divulgada na literatura mais antiga, foi uma ideia inicial na tentativa de identificação e determinação das ordens dos modelos AR e MA , por isso o direcionamento indicado por ela deve ser visto com um *certo cuidado*, isto é, suas indicações devem ser *comprovadas* através de outras técnicas.

2.18.3 – MODELOS ARMA

O modelo conhecido por ARMA realiza a junção de um modelo autorregressivo (AR em azul) com um modelo de média móvel (MA em verde), conforme indicado pela expressão:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

O processo ARMA mais simples é definido como ARMA(1,1) e significa a existência de um processo do tipo AR(1) e de outro do tipo MA(1). Sendo a parte AR(1) definida por $y_t = c + \phi y_{t-1} + \varepsilon_t$ e possuindo apenas um termo defasado (ϕy_{t-1}) explicando y_t . O processo MA(1), definido por $y_t = c + \varepsilon_t + \theta \varepsilon_{t-1}$, também possui apenas um termo defasado de resíduo ($\theta \varepsilon_{t-1}$) explicando y_t . A parcela ε_t é um ruído branco.

Uma característica de elevada relevância de um processo autorregressivo ARMA(p,q) é que a série descrita por este seja estacionária.

2.18.3.1 – DETERMINAÇÃO DO MODELO ARMA

As recomendações abaixo eram os procedimentos mais comuns entre os anos de 1970 até o início dos anos 2000, quando ainda não havia as possibilidades computacionais e matemáticas atuais. Contudo, tais recomendações ainda oferecem contribuições para a avaliação de modelos. Normalmente, não é muito evidente avaliar os processos que estão agindo nas escalas de tendência e sazonalidade somente analisando a FAC e FACP, daí a necessidade de uma busca *empírica* alterando as ordens de AR e de MA na porção de tendência e sazonalidade.

Recomendações:

- Identificar a ordem do modelo AR e MA através da análise da FAC e FACP. Ressaltando que, dependendo da série, isso não é uma tarefa trivial;
- Empregar a estatística Q sobre os resíduos estimados;
- Caso não se confirme o modelo como adequado, isto é, se ainda existirem autocorrelações nos resíduos, então deve-se adicionar novas ordens ao modelo e repetir a verificação. Esse processo de tentativa e erro pode se tornar bastante exaustivo (devido ao número de combinações possíveis para as diversas ordens de AR e MA) até que se encontre o modelo satisfatório em termos de ausência de autocorrelação residual.

A fim de tentar mitigar essa dificuldade, foi criada a função *auto.arima*, desenvolvida por pesquisadores americanos e europeus, presente no pacote *forecast* desenvolvida por Hyndman e disponibilizada no *software* estatístico R. Essa função investiga diversas combinações e aponta uma *possível* solução, utilizando os critérios selecionados pelo usuário. Por tratar-se de um método automático, este deve ser utilizado com certa *parcimônia*, servindo apenas como uma *sugestão inicial*, semelhante às sugestões apontadas na tabela 12. O conhecimento do fenômeno descrito pela série temporal é de fundamental importância para a avaliação do resultado apresentado pela função, até porque os resultados variam significativamente, dependendo dos parâmetros selecionados pelo usuário. Por exemplo, a série WS_m sabidamente apresenta sazonalidade, contudo o resultado do *auto.arima* não aponta nessa direção. A Figura 42 mostra o roteiro sugerido por Hyndman *et al.* (2018) para estimação de um modelo ARMA ou ARIMA.

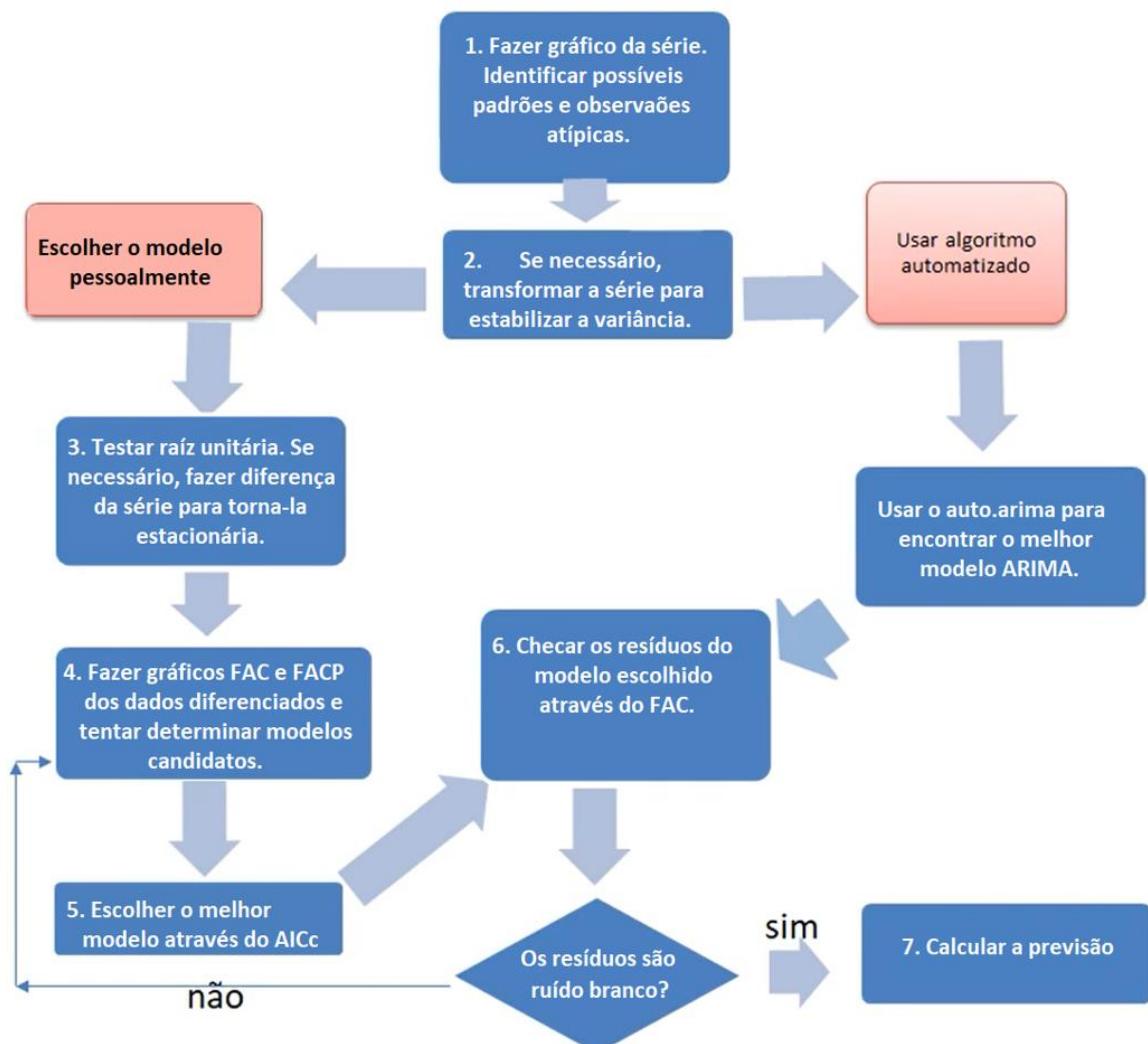


Figura 42 – Roteiro para escolha de um modelo.

Fonte: Hyndman *et al.* (2018)

O lado esquerdo do roteiro indica o procedimento utilizado pelos pesquisadores até aproximadamente os primeiros anos de 2000, que pode ser considerado como manual. À direita, uma sugestão valendo-se da atual evolução computacional e de um algoritmo de automação, que executa o fluxo da esquerda inúmeras vezes de forma mais rápida, culminando com uma *sugestão de possível melhor solução*.

Os procedimentos mais recentes muitas vezes ignoram um pouco algumas características dos correlogramas, contudo, ainda há motivos para examiná-los, pois são úteis para investigação dos padrões das séries temporais.

2.18.4 – MODELOS ARIMA

Outra abordagem para fins de previsão é o modelo ARIMA que, juntamente com a suavização exponencial, são as duas técnicas mais utilizadas para previsão e oferecem abordagens complementares. Os modelos de suavização exponencial estão baseados na tendência e sazonalidade e o ARIMA visa descrever as autocorrelações.

Ao combinar-se diferenciação ao modelo ARMA, obtém-se o modelo ARIMA não sazonal. ARIMA é um acrônimo para *AutoRegressive Integrated Moving Average* e, neste contexto, integração é o inverso de diferenciação. Assim, o modelo ARIMA pode ser escrito como:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Onde: y'_t é a série diferenciada, que pode ser de primeira ordem ou superior.

Da expressão acima, observa-se que os componentes do lado direito incluem valores defasados de y_t assim como os resíduos retardados ε_{t-1} a ε_{t-q} . Esse modelo é conhecido como ARIMA(p, d, q), onde p se refere à parte autorregressiva, d ao grau de diferenciação e q à ordem relativa à média móvel. Pode-se também observar que alguns modelos são um caso particular do modelo ARIMA, conforme apontado pela Tabela 13.

Tabela 13 – Casos particulares do modelo ARIMA.

Fonte: Hyndman *et al.* (2018)

ruído branco	ARIMA(0,0,0)
Caminhada aleatória	ARIMA(0,1,0) sem constante
Passeio aleatório com deriva	ARIMA(0,1,0) com uma constante
Autoregressão	ARIMA($p, 0, 0$)
Média móvel	ARIMA(0,0, q)

Basicamente, a diferença entre o modelo ARMA e o ARIMA é que no primeiro assume-se que não há sazonalidade e a série é estacionária. Por este motivo, não existe a parte I (relativa à diferenciação) no modelo ARMA. O modelo ARIMA possui a capacidade de realizar processos de diferenciação sobre uma série a fim de deixá-la estacionária e tal indicação acontece na posição d do modelo.

2.18.5 – MODELOS SARIMA

SARIMA é o acrônimo de modelos Autorregressivos Integrados de Média Móvel com Sazonalidade. A introdução de uma parte sazonal é o que distingue este modelo na classe dos autorregressivos. Um modelo SARIMA pode ser descrito como $ARIMA(p, d, q)(P, D, Q)_m$, onde o primeiro conteúdo entre parênteses se refere à parte não sazonal do modelo, o segundo à parte sazonal e m corresponde ao número de períodos sazonais.

$$\begin{array}{ccc}
 \text{ARIMA} & \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{parte não sazonal} \\ \text{do modelo}}} & \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{parte sazonal} \\ \text{do modelo}}}
 \end{array}$$

Os modelos ARIMA exploram a autocorrelação entre as observações de uma série em instantes consecutivos, porém as observações também podem apresentar autocorrelação para determinados períodos sazonais e os modelos que contemplam esta característica são conhecidos como SARIMA. Uma vantagem do modelo SARIMA é oferecer rapidez e conveniência na previsão de séries temporais.

2.19 – MÉTODOS DE PREVISÃO PARA SÉRIES TEMPORAIS DE ALTA FREQUÊNCIA

Séries temporais que possuem alta frequência de aquisição de dados, como aquelas registradas por hora ou fração, são conhecidas como séries intradiárias e exibem padrões sazonais mais complicados. Assim sendo, as ditas séries intradiárias podem apresentar um padrão diário, semanal, quinzenal, mensal, trimestral, semestral e anual. Essas séries são comumente desafiadoras no tocante à previsão, pois a maioria dos métodos anteriormente citados não lidam eficientemente com períodos sazonais elevados. Apesar destas séries terem a possibilidade de serem regidas por vários padrões sazonais, isso não implica necessariamente na obrigatoriedade de utilização de todos os padrões sazonais existentes, mas somente daqueles que provavelmente estarão presentes nos dados correntemente analisados.

As duas séries temporais utilizadas nesse trabalho se enquadram na classificação de séries intradiárias, pois a primeira (WS_h) possui frequência de aquisição horária e a segunda (WS_m) frequência de aquisição de quinze minutos. Nesses casos, são sugeridas algumas abordagens especiais como a utilização da Decomposição de Tendência-Sazonal usando Loess STL (*Seasonal and Trend Decomposition using Loess*), que lida igualmente bem com sazonalidade múltipla e com alta frequência de aquisição. O STL é um método versátil e robusto para decomposição de séries temporais, enquanto Loess é um método para estimar relacionamentos não lineares.

A Decomposição STL, além de retornar as componentes de tendência e restante, também retorna as diversas componentes sazonais, que porventura existam. Por exemplo, a Figura 43 mostra a decomposição STL realizada sobre a série *calls* (disponibilizada no *software* estatístico R), que apresenta as componentes relativas aos padrões sazonais diário de valor 169 (*Seasonal169*) e semanal de valor 845 (*Seasonal845*), assim como as componentes de tendência e restante.

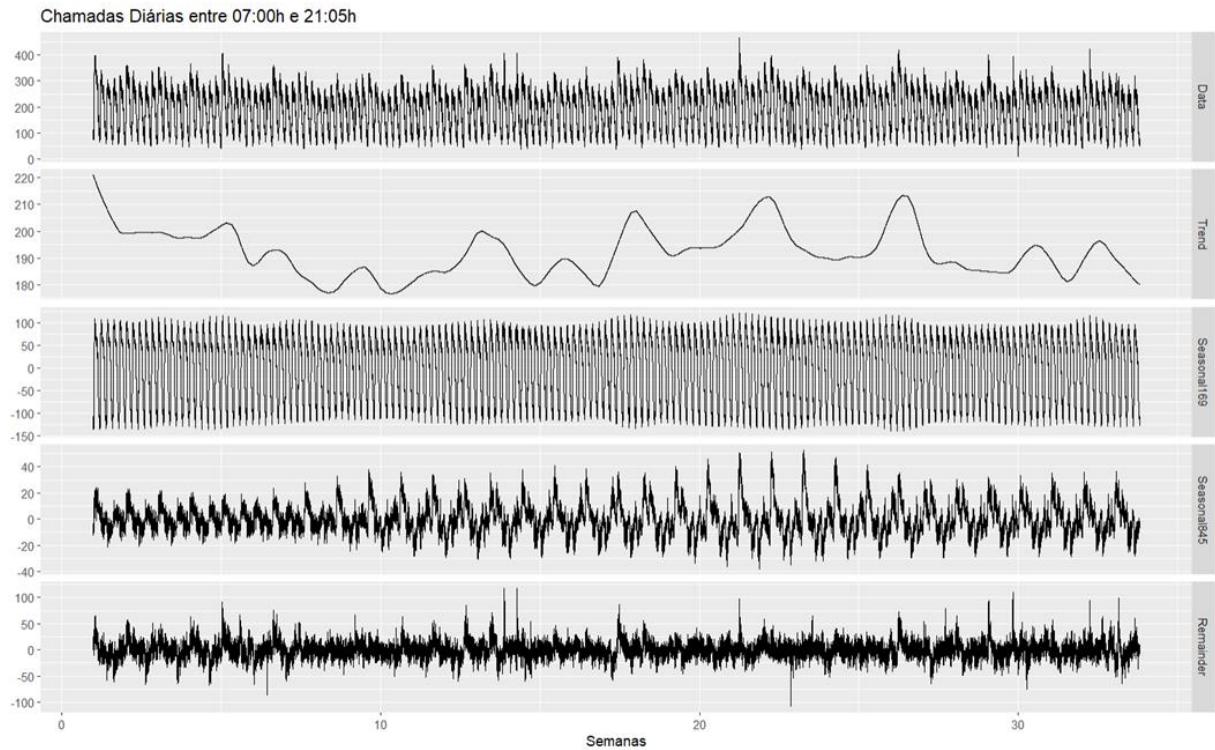


Figura 43 – Decomposição STL sobre a série *calls* mostrando sua múltipla sazonalidade.

A segunda abordagem sugerida para trabalhar-se com séries intradiárias é a Regressão Harmônica Dinâmica com Vários Períodos Sazonais, que utiliza os termos de Fourier (em função de seno e cosseno) para cada período sazonal existente. Outra sugestão se refere ao chamado modelo TBATS, desenvolvido por De Livera *et al.* (2011), que utiliza uma combinação de termos de Fourier com um modelo de espaço de estados de suavização exponencial e uma transformação Box-Cox.

O STL pode lidar com fenômenos de sazonalidades múltiplas, sazonalidades de alta frequência (como nas séries intradiárias) e ciclos. Caso a série alvo desse estudo (WS_m) considerasse um ou mais anos, certamente apresentaria um comportamento multissazonal e, nesse caso, a escolha do modelo TBATS seria mais uma opção.

Segundo Hyndman *et al.* (2018) a decomposição STL também pode ser usada na previsão, com cada uma das componentes sazonais previstas (caso exista mais de uma), empregando um método sazonal ingênuo e a previsão de dados ajustados sazonalmente usando modelos de espaço de estado de suavização exponencial ou outro método especificado pelo usuário. Tal estratégia foi utilizada neste trabalho para a previsão das componentes obtidas pela

decomposição STL, seja para a componente sazonal (método ingênuo), seja para a componente tendência (método SES) e para a componente restante (modelo ARMA).

2.19.1 – DECOMPOSIÇÃO DE TENDÊNCIA-SAZONAL USANDO LOESS – STL

A Decomposição de Tendência-Sazonal usando Loess (STL) é um método versátil e robusto de decomposição de séries temporais, elaborado por Cleveland *et al.* (1990), utilizando o método Loess de estimação de relacionamentos não lineares e frequentemente usado em análises econômicas e ambientais. A Decomposição STL é uma técnica de suavização fundamentada em regressões ajustadas localmente, que consiste em um método não paramétrico que estima curvas e superfícies utilizando suavização. Com o desenvolvimento das ferramentas computacionais, surgiram diversas publicações como Stones (1980) e o método ganhou grande projeção, inclusive com a criação de Cleveland (1990) do *software* de chamado Lowess. Esse programa foi implementado em diversos pacotes estatísticos, incluso o R, aqui utilizado.

Segundo Marquetti *et al.* (2014), o cerne do método pode ser entendido observando-se um simples modelo de regressão, onde a variável independente y e a variável dependente x são relacionadas por:

$$y_i = g(x_i) + \varepsilon_i$$

onde ε_i representa o termo iid com distribuição normal, média zero e variância constante ($\sim N(0, \sigma^2)$). Diferente dos métodos paramétricos, que estimam a função de forma global, a regressão local avalia a função g na vizinhança de cada ponto de interesse $x = x_0$. A estimativa de uma função local é realizada através de uma média ponderada das observações que se encontram na vizinhança do ponto desejado x_0 . Para tal, define-se o tamanho da vizinhança h do ponto $x = x_0$ e, a seguir, escolhe-se uma função K que pondere o conjunto de pontos na vizinhança de x_0 , conforme inicialmente proposto por Nadaraya (1964) e Watson (1964). Tal função recebe o nome de *Kernel* e h é chamada de bando ou parâmetro de suavização. Assim, a equação para a média local ponderada por K tem a seguinte forma:

$$\hat{g}(x_0) = \frac{\sum_{i=1}^n K_h(x_i - x_0) y_i}{\sum_{i=1}^n K_h(x_i - x_0)}$$

Algumas limitações, como viés nas regiões de fronteiras e no interior, caso a variável independente não seja uniforme e caso a função de regressão tenha curvatura, foram resolvidas através de regressão local linear ponderada, proposta por Stone (1977) e Cleveland (1979). Então ao estimar uma linha reta localmente ao invés de uma constante, fica resolvido o problema de viés de primeira ordem.

A modelagem de um processo através de regressão local implica nas seguintes escolhas: uma função Kernel, o parâmetro de suavização h (*span* ou *bandwidth*) e o grau polinomial. Caso o parâmetro de suavização escolhido seja pequeno, a estimativa terá um viés reduzido, porém uma variabilidade elevada. Do contrário, caso h seja grande, ocorrerá o oposto, isto é, um viés elevado e uma pequena variabilidade. Loader (1995) compara diferentes procedimentos para escolha da banda e os classifica em dois grupos. O primeiro está calcado em métodos clássicos, que representam extensões de procedimentos empregados em regressões paramétricas, como validação cruzada, critério de informação de Akaike e Cp de Mallows. O segundo grupo está apoiado em inserções (*plug-ins*), que consiste em escrever a função inicialmente estimada \hat{g} , como uma função g desconhecida e aproximada por uma expansão de Taylor ou outra expansão assintótica. Uma estimativa é então inserida (*plugged-in*) para derivar uma estimativa da tendenciosidade e uma estimativa da aderência, tal como o erro quadrático médio integrado (*mean integrated square error*). O próprio Loader (1995) afirma que os métodos clássicos apresentam melhores resultados práticos e com grande abrangência de casos.

A escolha do grau do polinômio também interfere na relação entre variância e viés, de modo que quanto maior o grau da polinomial, menor será o viés e maior a variância para um mesmo parâmetro de suavização. Normalmente, polinomiais de baixa ordem já oferecem excelentes estimativas, usualmente são utilizadas polinomiais com graus variando entre zero e três. A função *kernel* pondera as observações na vizinhança do ponto de interesse x_0 e segundo Cleveland *et al.* (1996) e Loader (1999) esta deve ser contínua, simétrica, com maior peso em torno de x_0 .

Características da Decomposição STL:

- Componente sazonal dinâmica, isto é, a sazonalidade pode sofrer alterações no decorrer do tempo e essas podem ser controladas pelo usuário;
- Capacidade de lidar com *outliers*;

- Aplicável nativamente só para decomposições aditivas. $Y_t = T_t + S_t + E_t$.

Essa última característica não implica necessariamente numa limitação de uso, visto que a decomposição multiplicativa pode ser transformada em aditiva através de uma transformação logarítmica. Assim, a decomposição multiplicativa $Y_t = T_t \times S_t \times E_t$ após uma transformação logarítmica, transforma-se em $\log Y_t = \log T_t + \log S_t + \log E_t$, isto é, uma decomposição aditiva.

2.19.2 – DECOMPOSIÇÃO STL NO SOFTWARE ESTATÍSTICO R

A função *stl()* presente no pacote *forecast* e disponível no software estatístico R, foi desenvolvida por Ripley B. D. a partir do código fortran disponibilizado em Cleveland *et al.* (1990). Um parâmetro obrigatório da função *stl* é o *s.window*, que permite ao usuário definir o tamanho da janela de ajuste para sazonalidade, de forma que menores valores resultam em mudanças de sazonalidade mais rápida, mas produzem mais ruído sobre esta componente. Este parâmetro controla a rapidez que a componente sazonal pode mudar e, utilizando-se o valor "periodic", força-se essa componente ser periódica, ou seja, idêntica para todos os dias, como exigido no caso da série WS_m utilizada neste trabalho.

A função *stl()* remove os valores sazonais e a componente restante é suavizada para encontrar a tendência. O nível geral é removido da componente sazonal e adicionado à componente de tendência. Esse processo é repetido algumas vezes. A componente restante são os resíduos do ajuste sazonal mais a tendência. Cleveland também recomenda que o valor de *s.window* seja ímpar e no mínimo igual a 7. Em resumo, valores menores fazem com que a componente sazonal acompanhe mais rapidamente os movimentos da série e valores maiores a tornam mais suavizada, porém não acompanha tão bem a série. Conforme aumenta-se o valor de *s.window*, a sazonalidade tende a tornar-se constante. Como de praxe, a ideia é observar os resíduos na tentativa de deixá-los mais próximos a um ruído branco. O parâmetro *t.window* permite ao usuário a definição do tamanho da janela de ajuste para tendência e, novamente, valores menores produzirão mudanças de tendências mais rápidas. Se o valor do parâmetro *robust* for definido como *true*, a influência de *outliers* sobre a série é mitigada.

Como mencionado acima, a decomposição também pode ser usada na previsão, combinando o STL a métodos mais simples como os modelos de espaço de estado de suavização exponencial utilizando os dados de tendência STL e então realizando previsões adicionando a última temporada de dados sazonais. A função `stlf()` realiza essa tarefa automaticamente. A função `mstl()` é uma variação da `stl()` e é designada para lidar com sazonalidade múltipla, conforme mostrada na decomposição da série `calls` apresentada acima.

2.20 – TESTE Q DE LJUNG-BOX

A qualidade, estabilidade e adequação de um modelo de previsão pode ser mensurada através da análise de seus resíduos. Nesse sentido, uma das técnicas disponíveis é o teste Q de Ljung-Box, que avalia os coeficientes de autocorrelação em conjunto, avaliando a sua significância e verificando a existência de valores diferentes de zero. Por exemplo, pode haver algumas pequenas autocorrelações específicas, porém o conjunto destas pode ser significantemente grande.

Assim, o que se deseja é constatar se alguma autocorrelação é diferente de zero até a defasagem k e a hipótese nula (H_0) é que todas as ρ_k até a ordem k são iguais a zero, ou seja, a série é estacionária. A expressão da estatística de teste é:

$$Q = T(T+2) \sum_{j=1}^k \frac{\rho_j^2}{T-j}$$

Então:

$$H_0 : \sum_{j=1}^n \rho_j = 0$$

$$H_a : \sum_{j=1}^n \rho_j \neq 0$$

Se uma das autocorrelações for diferente de zero, há evidência de existência de um modelo ARMA(p,q) ou ARIMA(p,d,q).

3 - ANÁLISE DOS DADOS COLETADOS PELAS TORRES METEOROLÓGICAS

3.1 - AQUISIÇÃO DOS DADOS METEOROLÓGICOS

As grandezas físicas velocidade e direção do vento, presentes nas séries temporais aqui utilizadas, foram obtidas através dos dados coletados pelas torres meteorológicas localizadas nas imediações das usinas nucleares de Angra dos Reis. Neste trabalho foram considerados os dados aquisitados pelas torres A, B C e D, no período de 2013 a 2020, e a Figura 44 apresenta a localização dessas torres no referido sítio. O intervalo considerado, (2013 a 2020), foi o período disponibilizado para esse trabalho. Uma vez adquiridos, os dados precisam ser tabulados e, para tal, foi utilizada uma planilha eletrônica (Excel).

Os dados meteorológicos coletados pelas torres são armazenados e enviados à central a cada 15 minutos (ou quarto de hora), isto é, nos quartos “00”, “15”, “30” e “45”. A torre A possui três níveis de altura (10, 60 e 100 m) com a capacidade de adquirir 40 dados meteorológicos, conforme apresentado na Tabela 14. As demais torres (B, C e D) possuem apenas um nível de altura (10 m) para coleta de 6 dados meteorológicos, como mostrado na Tabela 15. Portanto, o número total de dados coletados a cada hora pela torre A é igual a 160 e os dados coletados por hora pelas outras torres é 24, totalizando 3840 valores diários adquiridos pela torre A e 576 valores diários para cada uma das demais torres.

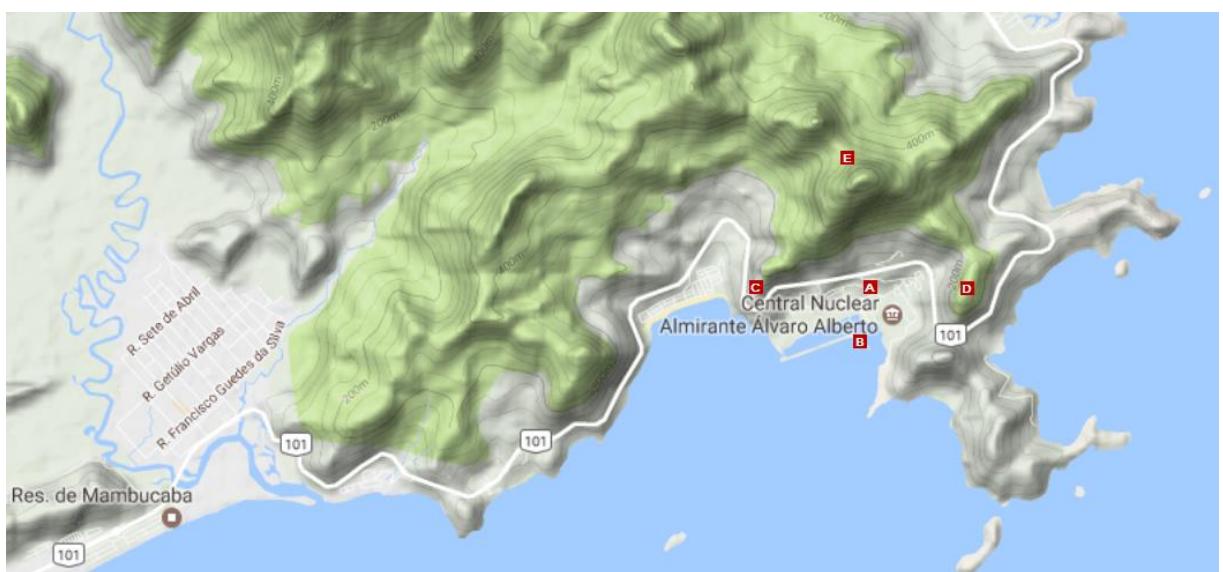


Figura 44 – Disposição das torres meteorológicas no sítio das usinas nucleares de Angra.

Fonte: Laboratório de Monitoração de Processos - LMP (2022).

3.2 - ESTRUTURA DOS ARQUIVOS TRANSMITIDOS PELAS TORRES

Os arquivos originais contendo os dados coletados pelas torres a cada 15 minutos são nomeados obedecendo à seguinte regra de formação QTMIMF onde:

Q – Quarto: indica que os dados são coletados de 15 em 15 minutos;

T – Torre: indica a torre responsável pela coleta (A, B, C ou D);

M – Mês: representação do mês de coleta, variando de A (janeiro) à L (dezembro);

I – Dia de Início da coleta;

M – Mês: representação do mês de coleta, variando de A (janeiro) à L (dezembro);

F – Dia de Fim da coleta

Devido ao maior número de dados coletados e enviados pela torre A, Tabela 14, esses são registrados em dois arquivos, aqui chamados de Torre A1 e Torre A2. A Tabela 15 apresenta a estrutura dos dados coletados e enviados pelas torres B, C e D. Os arquivos recebidos originalmente estão no formato .DBF, portanto, utilizou-se o *software* chamado DBFViewer para transformá-los em planilhas Excel. A seguir, foram executadas macros especialmente desenvolvidas para analisar e posicionar os campos meteorológicos no formato utilizado pelas séries temporais utilizadas neste trabalho (WS_h e WS_m).

Tabela 14 – Dados meteorológicos coletados pela Torre A.

Fonte: Central Nuclear Almirante Álvaro Alberto - CNAAA (2022).

Campo	Descrição	Unidade	Ref. do Arquivo .DBF
WS100M	Velocidade Vento - Torre de 100m ()	m/s	1
WD100M	Direção do Vento - Torre de 100m ()	graus	2
SIG100	Sigma - Torre de 100m		3
WG100	Gust - Torre de 100m		18
AT100	Temperatura - Torre de 100m (Cº)	Cº	10
TMX100	Máxima Temperatura - Torre de 100m (Cº)	Cº	24
TMN100	Mínima Temperatura - Torre de 100m (Cº)	Cº	21
WS60M	Velocidade Vento - Torre de 60m (m/s)	m/s	4
WD60M	Direção do Vento - Torre de 60m (graus)	graus	5
SIG60	Sigma - Torre de 60m		6
WG60	Gust - Torre de 60m		19
AT60	Temperatura - Torre de 60m (Cº)	Cº	11
TMAX60	Máxima Temperatura - Torre de 60m (Cº)	Cº	25
TMIN60	Mínima Temperatura - Torre de 60m (Cº)	Cº	22
WS10M	Velocidade Vento - Torre de 10m (m/s)	m/s	7
WD10M	Direção do Vento - Torre de 10m (graus)	graus	8

SIG10	Sigma - Torre de 10m		9
WG10	Gust - Torre de 10m		20
AT10	Temperatura - Torre de 10m (Cº)	Cº	12
TMAX10	Máxima Temperatura - Torre de 10m (Cº)	Cº	26
TMIN10	Mínima Temperatura - Torre de 10m (Cº)	Cº	23
DT1	Diferencial de Temperatura entre 10 e 100m	Cº	15
DT1MIN	Diferencial de Temperatura Mínima entre 10 e 100m	Cº	27
DT1MAX	Diferencial de Temperatura Máxima entre 10 e 100m	Cº	30
DT2	Diferencial de Temperatura entre 10 e 60m	Cº	16
DT2MIN	Diferencial de Temperatura Mínima entre 10 e 60m	Cº	28
DT2MAX	Diferencial de Temperatura Máxima entre 10 e 60m	Cº	31
DT3	Diferencial de Temperatura entre 60 e 100m	Cº	17
DT3MIN	Diferencial de Temperatura Mínima entre 60 e 100m	Cº	29
DT3MAX	Diferencial de Temperatura Máxima entre 60 e 100m	Cº	32
PASQH	Pasquil Horizontal		38
PASQV	Pasquil Vertical		39
SR	Radiação Solar (cal/cm2 min.)	cal/cm2 min.	14
SRMIN	Mínima - Radiação Solar ()	cal/cm2 min.	35
SRMAX	Máxima - Radiação Solar (cal/cm2 min.)	cal/cm2 min.	36
RH	Humidade Relativa (%)	%	13
RHMIN	Mínima Humidade Relativa (%)	%	33
RHMAX	Máxima Humidade Relativa (%)	%	34
RN	Índice Pluviométrico (mm)	mm	37
BATT	Bateria	Volts	40

Tabela 15 – Dados meteorológicos coletados pelas Torres B, C e D.

Fonte: Central Nuclear Almirante Álvaro Alberto - CNAAA (2022).

Campo	Descrição	Unidade	Referência do Arquivo .DBF
WS	Velocidade Vento	m/s	1
WD	Direção do Vento	graus	2
SIG	Sigma		3
WG	Gust		4
PASQ	Pasquil		5
BATT	Bateria	Volts	6

3.3 - ADEQUAÇÃO DOS ARQUIVOS À ESTRUTURA DE UMA SÉRIE TEMPORAL

As torres meteorológicas disponibilizam as informações coletadas em intervalos de 15 minutos (ou quarto de hora), de modo que a planilha original armazena esses valores nos intervalos de 0, 15, 30 e 45 minutos de cada hora e utiliza uma linha para cada dado meteorológico aquisitado. Através desses quartos de hora, a planilha original considera desde a hora 0 (zero) até a hora 24 (vinte e quatro), sendo que a hora 0 registra apenas os dados coletados nos quartos 15, 30 e 45 e a hora 24 somente os dados coletados no quarto 0. As demais horas do dia (1 a 23) consideram todos os quartos de hora, isto é, os instantes de coleta 0, 15, 30 e 45.

A fim de trabalhar-se com uma série temporal de período igual a uma hora, fez-se necessário transformar a periodicidade de quinze minutos da planilha original para esse novo período, de modo que cada linha na nova planilha represente uma hora do dia e suas colunas os campos meteorológicos coletados no quarto escolhido para essa representação horária. Assim, cada linha na nova planilha corresponderá a 40 linhas da planilha original para a torre A e a 6 linhas das planilhas originais advindas das torres B, C e D. Então, independente da torre e quarto considerados, para cada dia na nova planilha foram geradas 24 linhas, que correspondem às horas diárias e suas colunas representam os respectivos dados meteorológicos coletados.

A alteração de periodicidade de quarto de hora para hora nos permite escolher qual quarto será o mais adequado para construir a nova planilha. Para tanto, foi realizada uma análise sobre todos os registros armazenados em cada quarto de hora de todas as planilhas de cada torre no período correspondente aos anos entre 2013 e 2020, totalizando 15.821.272 registros analisados. Tal análise tem o objetivo de identificar anormalidades nos dados registrados e qual torre e quarto de hora possuem o maior número de dados isentos dessas inconformidades.

3.4 - ANÁLISE DE CONSISTÊNCIA DOS REGISTROS ORIGINAIS

Os dados de velocidade e direção do vento são obtidos através de anemômetros instalados nas torres, que após sua aquisição são transmitidos para uma central. Tanto no processo de aquisição quanto de transmissão podem ocorrer erros, fazendo-se então mister uma conferência rigorosa sobre todos os dados recebidos. Como esperado, a análise dos registros originariamente armazenados apontou algumas irregularidades no tocante à ausência, invalidade e duplicidade de informação. Tal análise foi realizada através de macros especialmente desenvolvidas para

esse fim e as Tabelas 16 e 17 apresentam, como exemplo, os resultados obtidos para a planilha QCD01D30, que corresponde aos dados enviados pela torre C a cada 15 minutos durante o mês de abril de 2013.

Tabela 16 – Parte superior dos resultados da análise realizada sobre a planilha QCD01C30.

C	D	E	F	G	H	I	J	K	L	M	N
1	NAME	CH	Planilha	Regs. Existentes	Regs. Esperados	Tot. Regs. Faltantes					
2	WS	1.147	QCD01D30	17190	17280	90					
3	WD	281.63					3	1	30	6	
4	SIGMA	31.739					3	1	45	6	
5	GUST	1.537					3	2	0 a 45	24	
6	PASQ	1	Tot. Reg. Faltantes	para cada Quarto			3	3	0 a 45	24	
7	BATT	12.448	Quarto 00	Quarto 15	Quarto 30	Quarto 45	3	4	0 a 45	24	
8	WS	0.57613	24	18	24	24	3	5	0	6	
9	WD	191.81									
10	SIGMA	49.232	Dia Inicial	1	Dia Final	30					
11	GUST	1.1771	Dias Faltantes	0							
12	PASQ	1									
13	BATT	12.422	Dados Meteorológicos	Inválidos por Quarto							
14	WS	1.0878	Quarto 00	Quarto 15	Quarto 30	Quarto 45	Dado Meteorológico				
15	WD	160.47	3	3	2	2	WS				
16	SIGMA	11.469	3	3	2	2	WD				
17	GUST	1.2671	3	3	2	4	SIGMA				
18	PASQ	4	2	3	3	3	GUST				
19	BATT	12.428	2	3	3	3	PASQ				
20	WS	0.81279	2	3	3	3	BATT				
21	WD	211.61									
22	SIGMA	31.735	15	18	15	17	65				
23	GUST	1.1371									
24	PASQ	1	Planilha Nova com	periodo igual a 1 hora							
25	BATT	12.428			Linhas Esperadas	Linhas Existentes					
26	WS	1.0589	Opção 1 (1 a 24h)	Quarto 00	720	716					
27	WD	303.05	Opção 4 (0 a 23h)	Quarto 15	720	717					
28	SIGMA	13.14		Quarto 30	720	716					
29	GUST	1.4705		Quarto 45	720	716					
30	PASQ	8									
31	BATT	12.422	Registros Repetidos	Dado Met. Repetido							
32	WS	1.0782									
33	WD	298.66	Resumo da Análise:	QCD01D30	2013	Abri					
34	SIGMA	6.5125	Dia Inicial	1	Dia Final	30					
35	GUST	1.337									
36	PASQ	6		Quarto 00	Quarto 15	Quarto 30	Quarto 45				
37	BATT	12.429	Regs. Inválidos	15	18	15	17				
38	WS	1.1165	Regs. Faltantes	24	18	24	24				
39	WD	157.03	Regs. Repetidos	0	Dias Faltantes	0					
40	SIGMA	34.001									
41	GUST	2.0095									
42	PASQ	1	Planilha	Regs. Existentes	Regs. Inválidos	Regs. Faltantes					
43	BATT	12.449	QCD01D30	17190	65	90					
44	WS	1.5772									
45	WD	165.18									
46	SIGMA	5.8711									
47	GUST	1.0780									

Tabela 17 – Parte inferior dos resultados da análise realizada sobre a planilha QCD01C30.

C	D	E	F	G	H	I	J	K	L	M	N
22	SIGMA	31.735	15	18	15	17	65				
23	GUST	1.1371									
24	PASQ	1	Planilha Nova com	periodo igual a 1 hora							
25	BATT	12.428			Linhas Esperadas	Linhas Existentes					
26	WS	1.0589	Opção 1 (1 a 24h)	Quarto 00	720	716					
27	WD	303.05	Opção 4 (0 a 23h)	Quarto 15	720	717					
28	SIGMA	13.14		Quarto 30	720	716					
29	GUST	1.4705		Quarto 45	720	716					
30	PASQ	8									
31	BATT	12.422	Registros Repetidos	Dado Met. Repetido							
32	WS	1.0782									
33	WD	298.66	Resumo da Análise:	QCD01D30	2013	Abri					
34	SIGMA	6.5125	Dia Inicial	1	Dia Final	30					
35	GUST	1.337									
36	PASQ	6		Quarto 00	Quarto 15	Quarto 30	Quarto 45				
37	BATT	12.429	Regs. Inválidos	15	18	15	17				
38	WS	1.1165	Regs. Faltantes	24	18	24	24				
39	WD	157.03	Regs. Repetidos	0	Dias Faltantes	0					
40	SIGMA	34.001									
41	GUST	2.0095									
42	PASQ	1	Planilha	Regs. Existentes	Regs. Inválidos	Regs. Faltantes					
43	BATT	12.449	QCD01D30	17190	65	90					
44	WS	1.5772									
45	WD	165.18									
46	SIGMA	5.8711									
47	GUST	1.0780									

Na parte superior desta análise (Tabela 16) são registrados o nome da planilha, o número de registros existentes, número de registros esperados e o total de registros faltantes, assim como o dia, hora, quarto de hora e o número de registros faltantes especificamente nesse quarto. Em seguida é apresentado o total de registros faltantes por quarto, o dia de início e fim da coleta e o número de dias faltantes. As próximas informações se referem aos registros presentes, porém inválidos, agrupados por quarto e relativo a cada dado meteorológico aquisitado.

A parte inferior da análise (Tabela 17) apresenta os registros repetidos em termos de data e hora juntamente com seus campos meteorológicos e, finalmente, o número de linhas esperadas e existentes para a nova planilha a ser gerada. As últimas linhas da análise fazem um resumo dos valores relevantes, que serão utilizados em outras planilhas (da Tabela 18 a Tabela 26 e da Tabela 27 a Tabela 30) para computação dos valores totais de registros faltantes e inválidos de cada torre e o valor total de registros inválidos por quarto, respectivamente.

3.5 - ESCOLHA DA TORRE E QUARTO DE HORA MAIS ADEQUADOS

Como uma das séries temporais objeto deste trabalho possui periodicidade de uma hora (WS_h), existe a possibilidade de escolha do quarto a ser utilizado para a sua representação, assim como qual torre apresenta o menor número de registros inválidos e faltantes.

A outra série aqui manipulada (WS_m) utiliza todos os quatro quartos de hora (00, 15, 30 e 45) e se refere somente ao mês de maio de 2016. A série WS_h foi utilizada no estudo da decomposição clássica e a série WS_m nos demais modelos analisados.

3.5.1 ESCOLHA DA TORRE PARA REPRESENTAÇÃO DA SÉRIE TEMPORAL WS_h

As Tabelas de número 18 a 25 apresentam os valores parciais e totais de registros faltantes e inválidos para cada uma das 480 planilhas originais relativas aos anos entre 2013 e 2020, permitindo assim a determinação de qual torre apresenta a menor quantidade dessas irregularidades. A Tabela 26 mostra os totais.

Tabela 18 – Número de registros faltantes e inválidos para cada torre no ano de 2013.

Tabela 19 – Número de registros faltantes e inválidos para cada torre no ano de 2014.

		Torre A1	Planilha	Exist.	Inv.	Falt.	Torre A2	Planilha	Exist.	Inv.	Falt.	Torre B	Planilha	Exist.	Inv.	Falt.	Torre C	Planilha	Exist.	Inv.	Falt.	Torre D	Planilha	Exist.	Inv.	Falt.
21	2014																									
22	janeiro	QAA01A15	57520	6001	80		QAA16A31	61440	6710	0		QBA01A31	17838	1127	18		QCA01A31	17844	38	12		QDA01A31	17838	0	18	
23	fevereiro	QAB01B15	57600	6154	0		QAB16B28	49280	5222	640		QBB01B28	16038	484	90		QCB01B28	16032	14	96		QDB01B28	16026	15	102	
24	março	QAC01C15	57600	6039	0		QAC16C31	61440	6631	0		QBC01C31	17856	0	0		QCC01C31	17856	0	0		QDC01C31	17856	9	0	
25	abril	QAD01D16	55000	5877	2600		QAD16D30	57560	6114	40		QBD01D30	17172	2	108		QCD01D30	16026	143	1254		QDD01D30	17196	23	84	
26	maio	QAE01E15	57560	6416	40		QAE16E31	61440	6296	0		QBE01E31	17592	5	264		QCE01E31	17598	401	258		QDE01E31	17604	28	252	
27	junho	QAF01F15	57600	6104	0		QAF16F30	57600	6426	0		QBF01F30	17274	28	6		QCF01F30	17280	469	0		QDF01F30	17280	31	0	
28	julho	QAG01G15	57600	6143	0		QAG16G31	61360	6611	80		QBG01G31	17844	291	12		QCG01G31	17850	244	6		QDG01G31	17850	48	6	
29	agosto	QAH01H15	57440	6144	160		QAH16H31	61440	6759	0		QBH01H31	17826	712	30		QCH01A31	17832	284	24		QDH01H31	17838	28	18	
30	setembro	QAI01I15	57520	5967	80		QAI16I30	57600	5967	0		QBI01I30	17256	46	24		QCI01I30	17256	15	24		QDI01I30	17082	27	198	
31	outubro	QAJ01J15	57400	5954	200		QAJ16J31	61280	6315	160		QBJ01J31	17838	173	18		QCJ01J31	17826	40	30		QDJ01J31	17814	31	42	
32	novembro	QAK01K17	57560	5893	40		QAK16K30	57600	6053	40		QBK01K30	17280	130	0		QCK01K30	17280	41	0		QDK01K30	17274	48	6	
33	dezembro	QAL01L15	57560	6030	40		QAL16L31	61280	6460	160		QBL01L31	17844	427	12		QCL01L31	17844	49	12		QDL01L31	17844	37	12	
34	Total			687960	72722	3240			709280	75564	1120			209658	3425	582			208524	1738	1716			209502	325	738
35	Tot. Exist.			2024924																						
36	Tot. Inv.			153774																						
37	Tot. Falt.			7396																						

Tabela 20 – Número de registros faltantes e inválidos para cada torre no ano de 2015.

		Torre A1	Planilha	Exist.	Inv.	Falt.	Torre A2	Planilha	Exist.	Inv.	Falt.	Torre B	Planilha	Exist.	Inv.	Falt.	Torre C	Planilha	Exist.	Inv.	Falt.	Torre D	Planilha	Exist.	Inv.	Falt.
40	2015																									
41	janeiro	QAA01A15	57560	6049	40		QAA16A31	61600	6281	1280		QBA01A31	17604	463	252		QCA01A31	17670	109	186		QDA01A31	17652	81	204	
42	fevereiro	QAB01B15	57120	5842	480		QAB16B28	49920	5247	0		QBB01B28	16116	24	12		QCB01B28	16116	321	12		QDB01B28	16128	114	0	
43	março	QAC01C15	57560	6363	40		QAC16C31	61200	6507	240		QBC01C31	17802	0	54		QCC01C31	17850	261	6		QDC01C31	17790	110	66	
44	abril	QAD01D15	57560	5883	40		QAD16D30	57560	5760	40		QBD01D30	17262	0	18		QCD01D30	17280	1	0		QDD01D30	17274	85	6	
45	maio	QAE01E15	57560	5778	0		QAE16E31	60960	6187	480		QBE01E31	17802	0	54		QCE01E31	17820	0	36		QDE01E31	17814	132	42	
46	junho	QAF01F15	57520	5931	80		QAF16F30	57440	5781	160		QBF01F30	17250	0	30		QCF01F30	17250	0	30		QDF01F30	17250	81	30	
47	julho	QAG01G15	57400	5866	200		QAG16G31	61240	6420	200		QBG01G31	17778	0	78		QCG01G31	17790	1	66		QDG01G31	17790	178	66	
48	agosto	QAH01H15	57600	6186	0		QAH16H31	61440	6418	0		QBH01H31	17856	28	0		QCH01A31	17844	1	12		QDH01H31	17856	59	0	
49	setembro	QAI01I10	38400	3868	0		QAI11I30	75200	7635	1480		QBI01I30	17064	157	216		QCI01I30	17034	3	246		QDI01I30	17034	64	246	
50	outubro	QAI01J13	49920	5010	0		QAI18J31	68040	6853	1080		QBJ01J31	17832	0	24		QCI01J31	17832	3	24		QDJ01J31	17814	68	42	
51	novembro	QAK01K17	30360	3026	34920		QAK18K30	49920	5054	0		QBK01K30	11874	1	5406		QCK01K30	17280	0	0		QDK01K30	17280	166	0	
52	dezembro	QAL01L17	65280	6606	0		QAL18L31	53760	5559	0		QBL01L31	17856	0	0		QCL01L31	17856	0	0		QDL01L31	17856	187	0	
53	Total			643880	66408	35800			719690	73702	4960			204096	673	6144			209622	700	618			209538	1325	702
54	Tot. Exist.			1984096																						
55	Tot. Inv.			142808																						
56	Tot. Falt.			48224																						

Tabela 21 – Número de registros faltantes e inválidos para cada torre no ano de 2016.

		Torre A1	Planilha	Exist.	Inv.	Falt.	Torre A2	Planilha	Exist.	Inv.	Falt.	Torre B	Planilha	Exist.	Inv.	Falt.	Torre C	Planilha	Exist.	Inv.	Falt.	Torre D	Planilha	Exist.	Inv.	Falt.
59	2016																									
60	janeiro	QAA01A17	65280	7418	0		QAA18A31	53760	5787	0		QBA01A31	17856	696	0		QCA01A31	17856	15	0		QDA01A31	17856	97	0	
61	fevereiro	ARQ, ERRO					QAB18B28	46000	5038	80		QBB01B28	16692	181	12		QCB01B29	16692	81	12		QDB01B29	16692	69	12	
62	março	QAC01C17	65280	7514	0		QAC18C31	53640	6214	120		QBC01C31	17850	218	6		QCC01C31	17838	0	18		QDC01C31	17856	78	0	
63	abril	QAD01D17	65200	8107	80		QAD18D30	49920	5910	0		QBD01D30	17280	653	0		QCD01D30	17274	3	6		QDD01D30	17280	121	0	
64	maio	QAE01E17	65280	7389	0		QAE18E31	53760	6136	0		QBE01E31	17856	503	0		QCE01E31	17856	0	0		QDE01E31	17856	84	0	
65	junho	QAF01F17	65280	7287	80		QAF18F30	49760	6722	160		QBF01F30	16368	581	912		QCG01G31	17844	117	12		QDG01G31	17850	131	6	
66	agosto	QAH01H17	65280	8289	0		QAH18H31	53760	6134	0		QBH01H31	17850	912	6		QCH01H31	17424	1045	432		QDH01H31	17856	80	0	
67	setembro	QAI01I17	65280	8411	0		QAI18I30	49920	5453	0		QBI01I30	17280	835	0		QCI01I30	17280	9	0		QDI01I30	17280	162	0	
68	outubro	QAJ01J17	65240	7135	160		QAJ18J31	53760	6290	0		QBJ01J31	17832	705	24		QCI01J31	17832	8	24		QDJ01J31	17832	160	24	
69	novembro	QAK01K17	65280	7286	0		QAK18K30	49920	5850	0		QBK01K30	17280	1706	0		QCK01K30	17280	55	0		QDK01K30	17280	165	0	
70	dezembro	QAL01L17	65280	7916	0		QAL18L31	53760	6085	0		QBL01L31	17856	2395	0		QCL01L31	17856	45	0		QDL01L31	17856	152	0	
71	Total			717800	86629	280			621880	72203	200			210750	10717	66	</									

Tabela 23 – Número de registros faltantes e inválidos para cada torre no ano de 2018.

		Torre A1			Torre A2			Torre B			Torre C			Torre D				
		Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	
100	2018	QAA01A17	65240	6884	40	QAA18A31	53760	5697	0	QBA01A31	17850	1319	6	QCA01A31	17862	93	0	
101	janeiro	QAB01B17	65240	6856	40	QAB18B28	42240	5111	0	QBB01B28	16128	556	0	QCB01B28	15546	97	582	
102	fevereiro	QAC01C17	65240	7402	40	QAC18C31	53720	6301	40	QBC01C31	17850	594	6	QCC01C31	17856	1475	0	
103	março	QAD01D17	65280	7492	0	QAD18D30	49920	6193	0	QBD01D30	17280	15	0	QCD01D30	17280	4002	0	
104	abril	QAE01E17	65120	7957	160	QAE18E31	53760	6889	0	QBE01E31	17844	54	12	QCE01E31	17838	4417	18	
105	maio	QAF01F17	65280	8000	0	QAF18F30	49920	5287	0	QBF01F30	17280	161	0	QDE01E31	17820	4	36	
106	junho	QAG01G17	65280	6602	0	QAG18G31	53760	5474	0	QBG01G31	17856	173	0	QCG01G31	17856	114	0	
107	julho	QAH01H17	65280	6547	0	QAH18H31	53760	5371	0	QBH01H31	17856	320	0	QCH01H31	17856	19	0	
108	agosto	QAI01I17	65280	6555	0	QAI18I30	49920	5045	0	QBI01I30	16902	130	378	QCI01I30	17280	63	0	
109	setembro	QAJ01J17	65280	6399	0	QAJ18J31	53760	5381	0	QBJ01J31	17796	156	60	QCI01J31	17856	55	0	
110	outubro	QAK01K17	65160	6473	120	QAK18K30	49920	4952	0	QBK01K30	17262	124	18	QCK01K30	17256	119	24	
111	novembro	QAL01L17	65120	6547	160	QAL18L31	53760	5466	0	QBL01L31	17856	110	0	QCL01L31	17814	2075	42	
112	dezembro	Total	782800	83714	560		618200	67167	40		209760	3712	480		209580	15602	666	210150
113		Tot. Exist.	2030490															261
114		Tot. Inv.	170456															90
115		Tot. Falt.	1836															
116																		

Tabela 24 – Número de registros faltantes e inválidos para cada torre no ano de 2019.

		Torre A1			Torre A2			Torre B			Torre C			Torre D				
		Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	
119	2019	QAA01A17	65000	6717	280	QAA18A31	53760	5520	0	QBA01A31	17856	271	0	QCA01A31	17826	287	30	
120	janeiro	QAB01B17	65280	6409	0	QAB18B28	42240	4242	0	QBB01B28	16116	93	12	QCB01B28	16128	25	0	
121	fevereiro	QAC01C17	65200	6444	80	QAC18C31	53360	5612	400	QBC01C31	17346	176	510	QCC01C31	17814	179	42	
122	março	QAD01D17	65240	6823	40	QAD18D30	49920	5328	0	QBD01D30	17274	422	6	QCD01D30	17280	401	0	
123	abril	QAE01E17	65280	6913	0	QAE18E31	24000	2774	29760	QBE01E31	17856	495	0	QCE01E31	17400	533	456	
124	maio	QAF01F17	11400	1161	53880	QAF18F30	43680	5112	6240	QBF01F30	16680	2177	600	QCF01F30	17280	103	0	
125	junho	QAG01G17	65280	7085	0	QAG18G31	47880	4954	5880	QBG01G31	17856	2395	0	QCG01G31	17790	118	66	
126	julho	QAH01H17	14560	1605	50720	QAH18H31	40720	4222	13040	QBH01H31	17850	1012	6	QCH01H31	17766	160	90	
127	agosto	QAI01I17	14560	14560	50720	QAI18I30	45880	4572	4040	QBI01I30	17268	801	12	QCI01I30	17274	257	6	
128	setembro	QAI01J17	65080	6761	200	QAI18J31	53760	5419	0	QBI01J31	17856	3622	0	QCI01J31	17856	0	0	
129	outubro	QAK01K17	65080	6549	200	QAK18K30	49920	5065	0	QBK01K30	17280	3	0	QCK01K30	17280	0	0	
130	novembro	QAL01L17	65280	6705	0	QAL18L31	53760	5466	0	QBL01L31	17856	6	0	QCL01L31	17808	0	48	
131	dezembro	Total	677920	69898	105440		558880	58286	59360		209094	11473	1146		209550	2124	690	209970
132		Tot. Exist.	1865414															1800
133		Tot. Inv.	143581															270
134		Tot. Falt.	166906															
135																		

Tabela 25 – Número de registros faltantes e inválidos para cada torre no ano de 2020.

		Torre A1			Torre A2			Torre B			Torre C			Torre D				
		Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	Planilha	Exist.	Inv.	Falt.	
138	2020	QAA01A17	65280	6556	0	QAA18A31	53760	5697	0	QBA01A31	17856	13	0	QCA01A31	17856	0	0	
139	janeiro	QAB01B17	65280	6668	0	QAB18B29	46080	4768	0	QBB01B29	16704	1688	0	QCB01B29	16704	0	0	
140	fevereiro	QAC01C17	65280	6763	0	QAC18C31	53720	6301	40	QBC01C31	17856	2313	0	QCC01C31	17856	0	0	
141	março	QAD01D17	65240	6842	40	QAD18D30	49920	6193	0	QBD01D30	17274	1954	6	QCD01D30	17274	0	6	
142	abril	QAE01E17	65280	6771	0	QAE18E31	53760	6889	0	QBE01E31	17850	400	6	QCE01E31	17856	2	0	
143	maio	QAF01F17	65240	6816	40	QAF18F30	49920	5287	0	QBF01F30	17280	1179	73	QCF01F30	17280	4	0	
144	junho	QAG01G17	65040	6665	240	QAG18G31	53760	5474	0	QBG01G31	17856	309	6	QCG01G31	17838	0	18	
145	julho	QAH01H17	65040	6903	0	QAH18H31	53760	5538	0	QBH01H31	17856	660	0	QCH01H31	17856	2	0	
146	agosto	QAI01I17	65280	6994	0	QAI18I30	49920	5021	0	QBI01I30	17280	411	0	QCI01I30	17244	0	36	
147	setembro	QAJ01J17	65280	6627	0	QAJ18J31	53760	5432	0	QBJ01J31	17856	234	0	QCI01J31	17850	0	6	
148	outubro	QAK01K17	65280	6689	0	QAK18K30	49920	4995	0	QBK01K30	17280	282	0	QCK01K30	17280	43	0	
149	novembro	QAL01L17	65280	6502	0	QAL18L31	53760	5443	0	QBL01L31	17856	264	0	QCL01L31	17856	9	0	
150	dezembro	Total	783040	80796	320		622040	67038	40		210798	9707	18		210750	129	66	207546
151		Tot. Exist.	2034174															3270
152		Tot. Inv.	157682															
153		Tot. Falt.	3714															
154																		

Tabela 26 – Total geral de registros faltantes e inválidos para cada torre de 2013 a 2020.

		Tot. Geral
157	Tot. Exist.	15821272
158	Tot. Inv.	1254365
159	Tot. Falt.	387566
160		
162	Geral por Torre	Torre A1
163	Tot. Exist.	5694960
164	Tot. Inv.	612983
165	Tot. Falt.	218640
166		
167	Inv. + Falt.	831623
168		
169		
170		
171		
172		
173		
174		
175		
176		
177		
178		
179		
180		
181		
182		
183		
184		
185		
186		
187		
188		
189		
190		
191		
192		
193		
194		
195		
196		
197		

Como pode ser observado na tabela acima, a torre que apresenta o menor número de registros faltantes e inválidos foi a torre D. A partir desta constatação, resta determinar o quarto que apresentará o menor número de registros inválidos nessa torre.

3.5.2 - QUARTO DE HORA PARA REPRESENTAÇÃO SÉRIE TEMPORAL

Uma vez estabelecido que os dados meteorológicos adquiridos pela torre D se mostraram os mais confiáveis, cabe então a escolha do quarto de hora que apresenta o menor número de valores inválidos adquiridos por essa torre. O quarto de hora escolhido será o representante horário.

As Tabelas de número 27 a 30 (abrangendo 2 anos por tabela) apresentam o total de valores inválidos, por quarto de hora, obtidos na aquisição da velocidade (WS) e direção (WD) do vento para cada mês entre os anos 2013 e 2020, realizada pela torre D. A Tabela 31 mostra os totais.

Tabela 27 – Valores mensais inválidos de WS e WD, por quarto, para os anos 2013 e 2014.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
2013	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45		2014	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45	
1															
2															
3															
4	janeiro	QDA01A31	WS	12	14	14	18		janeiro	QDA01A31	WS	0	0	0	
5			WD	12	14	14	18			WD	0	0	0	0	
6	fevereiro	QDB01B28	WS	9	10	6	10		fevereiro	QDB01B28	WS	2	1	0	2
7			WD	9	10	6	10			WD	2	1	0	2	
8	março	QDC01A31	WS	14	18	14	14		março	QDC01A31	WS	1	0	0	1
9			WD	14	18	14	14			WD	1	0	0	1	
10	abril	QDD01D30	WS	0	0	0	4		abril	QDD01D30	WS	1	0	5	1
11			WD	0	0	0	4			WD	1	0	5	1	
12	maio	QDE01E31	WS	0	0	0	0		maio	QDE01E31	WS	3	3	2	1
13			WD	0	0	0	0			WD	3	3	2	1	
14	junho	QDF01F30	WS	0	0	0	0		junho	QDF01F30	WS	3	2	4	1
15			WD	0	0	0	0			WD	3	2	4	1	
16	julho	QDG0131	WS	0	0	0	0		julho	QDG0131	WS	1	4	9	2
17			WD	0	0	0	0			WD	1	4	9	2	
18	agosto	QDH01H31	WS	0	0	0	0		agosto	QDH01H31	WS	2	2	3	2
19			WD	0	0	0	0			WD	2	2	3	2	
20	setembro	QDI01I30	WS	0	0	0	0		setembro	QDI01I30	WS	4	2	1	2
21			WD	0	0	0	0			WD	4	2	1	2	
22	outubro	QDJ01J31	WS	0	0	0	0		outubro	QDJ01J31	WS	0	3	4	3
23			WD	0	0	0	0			WD	0	3	4	3	
24	novembro	QDK01K30	WS	1	0	0	1		novembro	QDK01K30	WS	7	2	3	3
25			WD	1	0	0	1			WD	7	2	3	3	
26	dezembro	QDL01L31	WS	1	1	1	1		dezembro	QDL01L31	WS	2	4	0	6
27			WD	1	1	1	1			WD	2	4	0	6	
28	Total 2013			74	86	79	86		Total 2014			52	46	62	48

Tabela 28 – Valores mensais inválidos de WS e WD, por quarto, para os anos 2015 e 2016.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
31	2015	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45		2016	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45
32	janeiro	QDA01A31	WS	7	5	8	7		janeiro	QDA01A31	WS	8	8	9	6
33			WD	7	5	8	7			WD	8	8	9	6	
34	fevereiro	QDB01B28	WS	8	11	6	12		fevereiro	QDB01B28	WS	7	2	5	8
35			WD	8	11	6	12			WD	7	2	5	8	
36	março	QDC01A31	WS	4	12	14	6		março	QDC01A31	WS	7	8	3	7
37			WD	4	12	14	6			WD	7	8	3	7	
38	abril	QDD01D30	WS	6	12	6	3		abril	QDD01D30	WS	10	8	8	14
39			WD	6	12	6	3			WD	10	8	8	14	
40	maio	QDE01E31	WS	8	11	12	12		maio	QDE01E31	WS	4	10	7	6
41			WD	8	11	12	12			WD	4	10	7	6	
42	junho	QDF01F30	WS	6	5	5	11		junho	QDF01F30	WS	6	7	9	9
43			WD	6	5	5	11			WD	6	7	9	9	
44	julho	QDG0131	WS	15	9	15	19		julho	QDG0131	WS	11	13	7	11
45			WD	15	9	15	19			WD	11	13	7	11	
46	agosto	QDH01H31	WS	5	4	8	2		agosto	QDH01H31	WS	2	12	5	6
47			WD	5	4	8	2			WD	2	12	5	6	
48	setembro	QDI01I30	WS	7	7	5	2		setembro	QDI01I30	WS	15	14	12	12
49			WD	7	7	5	2			WD	15	14	12	12	
50	outubro	QDJ01J31	WS	4	4	6	8		outubro	QDJ01J31	WS	13	19	8	12
51			WD	4	4	6	8			WD	13	19	8	12	
52	novembro	QDK01K30	WS	13	20	8	13		novembro	QDK01K30	WS	12	17	14	11
53			WD	13	20	8	13			WD	12	17	14	11	
54	dezembro	QDL01L31	WS	16	16	13	16		dezembro	QDL01L31	WS	14	12	13	11
55			WD	16	16	13	16			WD	14	12	13	11	
56	Total 2015			108	122	112	122		Total 2016			218	260	200	226

Tabela 29 – Valores mensais inválidos de WS e WD, por quarto, para os anos 2017 e 2018.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
59	2017	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45		2018	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45
60	janeiro	QDA01A31	WS	8	8	9	6		janeiro	QDA01A31	WS	2	2	2	4
61			WD	8	8	9	6			WD	2	2	2	4	
62	fevereiro	QDB01B28	WS	9	9	7	4		fevereiro	QDB01B28	WS	0	0	0	0
63			WD	9	9	7	4			WD	0	0	0	0	
64	março	QDC01A31	WS	4	5	4	2		março	QDC01A31	WS	0	0	0	1
65			WD	4	5	4	2			WD	0	0	0	1	
66	abril	QDD01D30	WS	7	20	17	8		abril	QDD01D30	WS	0	0	0	0
67			WD	7	20	17	8			WD	0	0	0	0	
68	maio	QDE01E31	WS	9	12	13	8		maio	QDE01E31	WS	0	0	0	0
69			WD	9	12	13	8			WD	0	0	0	0	
70	junho	QDF01F30	WS	5	1	0	2		junho	QDF01F30	WS	0	0	0	0
71			WD	5	1	0	2			WD	0	0	0	0	
72	julho	QDG0131	WS	7	5	8	8		julho	QDG0131	WS	0	0	0	0
73			WD	7	5	8	8			WD	0	0	0	0	
74	agosto	QDH01H31	WS	0	0	0	0		agosto	QDH01H31	WS	0	0	2	0
75			WD	0	0	0	0			WD	0	0	2	0	
76	setembro	QDI01I30	WS	0	0	0	0		setembro	QDI01I30	WS	4	6	6	7
77			WD	0	0	0	0			WD	4	6	6	7	
78	outubro	QDJ01J31	WS	0	0	0	0		outubro	QDJ01J31	WS	2	6	3	5
79			WD	0	0	0	0			WD	2	6	3	5	
80	novembro	QDK01K30	WS	0	0	0	0		novembro	QDK01K30	WS	4	2	2	3
81			WD	0	0	0	0			WD	4	2	2	3	
82	dezembro	QDL01L31	WS	0	0	0	0		dezembro	QDL01L31	WS	0	0	0	0
83			WD	0	0	0	0			WD	0	0	0	0	
84	Total 2017			88	120	116	76		Total 2018			214	222	200	200

Tabela 30 – Valores mensais inválidos de WS e WD, por quarto, para os anos 2019 e 2020.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
87	2019	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45		2020	Planilha	Dado Met.	Qrt 00	Qrt 15	Qrt 30	Qrt 45	
88	janeiro	QDA01A31	WS	1	3	3	0		janeiro	QDA01A31	WS	0	0	0	0	
89			WD	1	3	3	0			WD	0	0	0	0	0	
90	fevereiro	QDB01B28	WS	1	1	4	2		fevereiro	QDB01B28	WS	0	0	0	0	0
91			WD	1	1	4	2			WD	0	0	0	0	0	
92	março	QDC01A31	WS	5	3	4	6		março	QDC01A31	WS	0	0	0	0	0
93			WD	5	3	4	6			WD	0	0	0	0	0	
94	abril	QDD01D30	WS	16	10	10	9		abril	QDD01D30	WS	0	0	1	0	0
95			WD	16	10	10	9			WD	0	0	1	0	0	
96	maio	QDE01E31	WS	30	28	28	28		maio	QDE01E31	WS	0	0	0	0	0
97			WD	30	28	28	28			WD	0	0	0	0	0	
98	junho	QDF01F30	WS	0	0	0	0		junho	QDF01F30	WS	0	0	0	1	0
99			WD	0	0	0	0			WD	0	0	0	1	0	
100	julho	QDG0131	WS	0	0	0	0		julho	QDG0131	WS	0	0	0	0	0
101			WD	0	0	0	0			WD	0	0	0	0	0	
102	agosto	QDH01H31	WS	1	0	0	0		agosto	QDH01H31	WS	0	0	0	0	0
103			WD	1	0	0	0			WD	0	0	0	0	0	
104	setembro	QDI01I30	WS	7	4	4	11		setembro	QDI01I30	WS	0	0	0	0	0
105			WD	7	4	4	11			WD	0	0	0	0	0	
106	outubro	QDJ01J31	WS	0	0	0	0		outubro	QDJ01J31	WS	0	0	0	0	0
107			WD	0	0	0	0			WD	0	0	0	0	0	
108	novembro	QDK01K30	WS	0	0	0	0		novembro	QDK01K30	WS	0	0	0	0	0
109			WD	0	0	0	0			WD	0	0	0	0	0	
110	dezembro	QDL01L31	WS	0	0	0	0		dezembro	QDL01L31	WS	0	0	0	0	0
111			WD	0	0	0	0			WD	0	0	0	0	0	
112	T	T	T	122	88	106	112		T	T	T	0	0	2	2	
	Totais								Totais							
	Inválidos por Quarto															

Tabela 31 – Total anual e geral dos valores inválidos de WS e WD, por quarto, para a torre D.

	A	B	C	D	E	F	G
115							
116				Qrt 00	Qrt 15	Qrt 30	Qrt 45
117	Total 2013			74	86	70	96
118	Total 2014			52	46	62	48
119	Total 2015			198	232	212	222
120	Total 2016			218	260	200	226
121	Total 2017			98	120	116	76
122	Total 2018			24	32	30	40
123	Total 2019			122	98	106	112
124	Total 2020			0	0	2	2
125							
126	Total Geral			Qrt 00	Qrt 15	Qrt 30	Qrt 45
127	2013 a 2020			786	874	798	822
128							

Através da tabela acima, pode ser constatado que o quarto de hora 00 (zero) foi o que apresentou o menor número de valores inválidos para os dados meteorológicos de velocidade e direção do vento coletados pela torre D. O próximo passo para a construção da série temporal envolvendo as grandezas WS e WD é o preenchimento dos dados faltantes e inválidos adquiridos pela torre D no quarto de hora 00.

3.5.3 - PREENCHIMENTO DOS DADOS FALTANTES E INVÁLIDOS

É desejável que uma série temporal não apresente lacunas ou dados incorretos, então os dados faltantes e inválidos devem ser reparados. Para tal, alguns critérios foram estabelecidos de modo a atender a esse objetivo com a maior coerência possível. Cada critério considera algumas condições e o não atendimento dessas conduzirá a escolha do critério seguinte. Os critérios estabelecidos são:

- 1 – Média entre o valor imediatamente anterior e posterior ao dado faltante ou inválido. Os valores imediatamente anterior e posterior correspondem a uma diferença de 15 minutos antes e depois do valor faltante ou inválido, respectivamente.
- 2 – Utilização do valor imediatamente anterior, caso o imediatamente posterior não exista ou seja inválido.
- 3 – Utilização do valor imediatamente posterior, caso o imediatamente anterior não exista ou seja inválido.
- 4 – Na ausência de valores imediatamente anterior e posterior, a utilização do valor acima do imediatamente anterior, que corresponde a uma diferença de 30 minutos antes do dado faltante ou inválido.
- 5 – Na impossibilidade de aplicação do critério 4, utiliza-se o valor abaixo do imediatamente posterior, que corresponde a uma diferença de 30 minutos à frente do dado faltante ou inválido.
- 6 – Diante da impossibilidade de aplicação dos critérios acima, o dado a ser utilizado deverá ser aquele registrado pela torre C.
- 7 – Como último recurso, a utilização de dados coletados pela torre A no nível de 100m.

Através da aplicação dos critérios acima, foi possível a criação e substituição de todos os registros que se fizeram necessários. Há que registrar que os critérios 1, 2 e 3 atenderam a aproximadamente 98% das necessidades, as demais foram atendidas pelos critérios 4 e 5. O critério 6 foi utilizado em cinco situações e o critério 7 em apenas uma ocasião.

3.5.4 - ANÁLISE DE *OUTLIERS*

Finalmente, após a construção da série temporal, faz-se necessário distinguir quais valores de velocidade e direção do vento podem ser classificados como *outliers*. Esse procedimento foi realizado utilizando-se a técnica de *Boxplot*.

4 – APLICAÇÃO DOS MÉTODOS E MODELOS DE ANÁLISE DE SÉRIES TEMPORAIS NA PREVISÃO DE VELOCIDADE DO VENTO

4.1 – MODELO CLÁSSICO DE ANÁLISE DE SÉRIES TEMPORAIS

4.1.1 – DEFINIÇÃO DA SÉRIE TEMPORAL

A série horária WS_h, foi construída a partir dos dados recebidos pela torre D e tratados conforme descrito acima. A série é composta por valores de velocidade do vento tomados de hora em hora, tendo como referência os valores coletados no quarto 00 durante o período de oito anos (2013 a 2020) e totalizando 70080 registros. A Figura 45 representa a distribuição das velocidades do vento durante esse período.

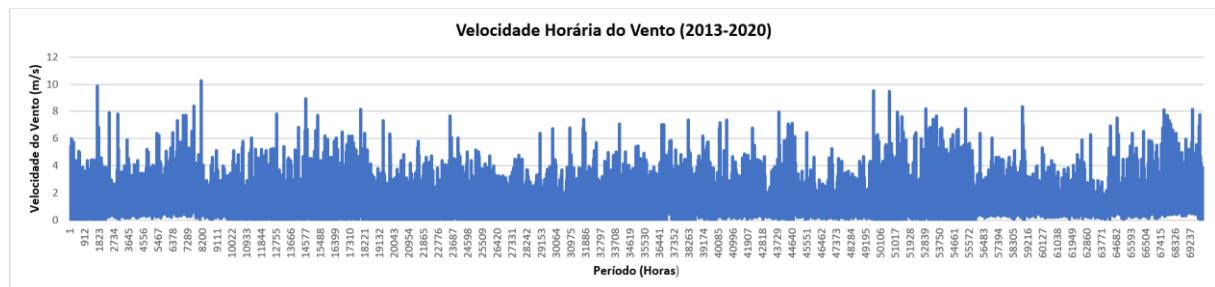


Figura 45 – Velocidade do vento tomada a cada hora entre os anos 2013 e 2020.

4.1.2 – OBTENÇÃO DA COMPONENTE DE TENDÊNCIA

Nos modelos de regressão utilizados para a obtenção da tendência, a variável independente X é o tempo, que corresponde ao período de aquisição (hora) de cada velocidade e numerado de 1 (janeiro de 2013) a 70080 (dezembro de 2020). Os modelos de regressão que descrevem a série WS_h são descritos pelas seguintes equações:

- Linear: $Y = 0,0000008 X + 1,3238$
- Logarítmica: $Y = -0,006 \ln(X) + 1,4137$
- Pol. de 2º grau: $Y = 0,0000000002 X^2 - 0,00001 X + 1,5018$
- Potência: $Y = 1,0199 X^{0,0017}$
- Exponencial: $Y = 0,9783 e^{0,000002 \cdot X}$

Após o cálculo das métricas (definidas em 2.16.1.1) para cada modelo, o escolhido será aquele que apresentar os valores de métricas mais próximo de zero, como mostrado pela cor verde da Tabela 32. O modelo potência é o que apresenta o maior número de métricas mais próximas de zero, então esse será o modelo escolhido para representar a tendência nos processos de recomposição e previsão.

Tabela 32 – Valores das métricas para cada modelo de regressão.

	EAM	EQM	EPM	EPAM
Reta	0,670213	0,835116	-375,462	400,646
Logarítmico	0,669564	0,835323	-376,274	401,3647
Pol. 2º grau	0,704246	0,84953	-408,702	429,4153
Potência	0,662912	0,933402	-265,406	307,7924
Exponencial	0,662878	0,926057	-268,251	309,9087

4.1.3 – OBTENÇÃO DA COMPONENTE SAZONAL

4.1.3.1 – MÉDIAS MÓVEIS CENTRADAS

Como a série temporal WS_h possui periodicidade inferior a um ano, então caso essa série apresente sazonalidade, poder-se-á estimar seus índices através da média móvel centrada. Para tal, define-se o tamanho da janela móvel com o mesmo valor de periodicidade da série, ou seja, um ano. Então, a janela móvel deverá abranger 8760 períodos, que correspondem ao número de horas em um ano. Devido à própria definição de média móvel centrada, haverá uma perda de atribuição de valores para os primeiros e últimos 4380 períodos da série, totalizando 8760 períodos sem valores de média móvel centrada atribuídos. Em outras palavras, haverá uma perda de um ano.

4.1.3.2 – ÍNDICES SAZONALIS ADITIVOS E MULTIPLICATIVOS

Uma vez calculado o valor da média móvel centrada atribuído a cada período de direito da série, passa-se ao cálculo dos índices sazonais aditivos e multiplicativos, utilizando-se as equações 1 e 2, respectivamente.

$$\text{Índice Sazonal Aditivo} = Y - \text{Média Móvel Centrada} \quad (1)$$

$$\text{Índice Sazonal Multiplicativo} = Y / \text{Média Móvel Centrada} \quad (2)$$

O próximo passo consiste na determinação de um representante dos índices sazonais aditivos e multiplicativos para cada hora, isto é, um valor de síntese para cada um dos 8760 períodos que definem um ano, conforme definidos nos fluxogramas das Figuras 30 e 31 acima. Uma vez determinado os valores de síntese para um ano, esses serão repetidos para os demais anos presentes na série para posterior uso nos processos de recomposição e previsão.

As Tabelas 33 e 34 apresentam os primeiros e últimos valores dos índices sazonais aditivos, respectivamente, calculados para os 8760 períodos e seus valores corrigidos. Lembrando que o somatório esperado para os representantes de índices sazonais aditivos deve resultar em 0 (zero).

Tabela 33 – Representantes dos índices (coluna BB) e seus valores corrigidos (coluna BC).

	AY	AZ	BA	BB	BC
1	Ano/Hora	Índice Sazonal Aditivo	Hora (1 a 8760)	Médias Simples Índices Aditivos	Índices Sazonais Aditivos Corrigidos
2	2013/1		1	0,141496005	0,145919054
3	2013/2		2	-0,141472094	-0,137049045
4	2013/3		3	0,491314694	0,495737743
5	2013/4		4	0,47779579	0,482218839
6	2013/5		5	0,179482306	0,183905355
7	2013/6		6	0,354325898	0,358748947
8	2013/7		7	-0,108875921	-0,104452872
9	2013/8		8	0,235236772	0,239659821
10	2013/9		9	-0,323574391	-0,319151342
11	2013/10		10	0,062875957	0,067299006
12	2013/11		11	-0,253120586	-0,248697537
13	2013/12		12	0,086583291	0,09100634
14	2013/13		13	0,456484421	0,46090747
15	2013/14		14	0,058147875	0,062570924
16	2013/15		15	0,130244305	0,134667354
17	2013/16		16	0,060501786	0,064924835
18	2013/17		17	0,147850089	0,152273138
19	2013/18		18	0,758656116	0,763079165
20	2013/19		19	-0,487200051	-0,482777002
21	2013/20		20	0,324728806	0,329151855

Tabela 34 – Somatório da coluna BB diferente de zero e somatório da BC (corrigido) igual a zero.

	AY	AZ	BA	BB	BC
8744	2013/8743	-0,535662646	8743	-0,104390791	-0,099967742
8745	2013/8744	-0,968714193	8744	-0,20751969	-0,203096641
8746	2013/8745	0,252201314	8745	-0,017043812	-0,012620763
8747	2013/8746	-0,49087081	8746	-0,473361165	-0,468938116
8748	2013/8747	-1,392424001	8747	-0,437708398	-0,433285349
8749	2013/8748	0,093384754	8748	0,050211023	0,054634072
8750	2013/8749	0,300959754	8749	0,033185977	0,037609026
8751	2013/8750	-1,067870919	8750	-0,225251332	-0,220828283
8752	2013/8751	-0,14895397	8751	0,141729385	0,146152434
8753	2013/8752	0,285018705	8752	0,122795934	0,127218983
8754	2013/8753	-0,135169485	8753	-0,068904686	-0,064481637
8755	2013/8754	2,614344156	8754	0,610987817	0,615410866
8756	2013/8755	0,303416582	8755	0,539430775	0,543853824
8757	2013/8756	0,884681542	8756	0,074254134	0,078677183
8758	2013/8757	0,444567483	8757	0,299411238	0,303834287
8759	2013/8758	1,462768373	8758	0,042008285	0,046431334
8760	2013/8759	1,077366393	8759	0,026156511	0,03057956
8761	2013/8760	-0,269496187	8760	-0,074472641	-0,070049592
8762	2014/1	1,621101614	Soma -38,74590903		0,000000000
8763	2014/2	-0,338228678	Diferença -0,004423049		

Cabe também observar as seguintes características:

- O número de períodos (horas) por ano é igual a 8760, coluna AY (Ano/Hora).
- Coluna AZ (Índice Sazonal Aditivo) dos primeiros 4380 períodos sem valor associado, sendo a primeira associação ocorrendo para o período 4381 e a última para o período 65700.
- O somatório realizado sobre a coluna BB (Médias Simples Índices Aditivos) apresentando valor diferente de zero e, após a primeira correção na coluna BC (Índices Sazonais Aditivos Corrigidos), resultando no valor esperado 0 (zero).

As Tabelas 35 e 36 mostram os primeiros e últimos valores dos índices sazonais multiplicativos, respectivamente, calculados para os 8760 períodos e seus valores corrigidos. Lembrando que o somatório esperado para os índices sazonais multiplicativos deve resultar em 8760, que corresponde ao valor da periodicidade da série e da janela móvel da média móvel centrada.

Tabela 35 – Representantes dos índices (Coluna BH) e a quarta correção (coluna BL).

BE	BF	BG	BH	BI	BJ	BK	BL
Ano/Hora	Índice Sazonal Mult.	Hora (1 a 8760)	Médias Internas Índ. Mult.	Índ. Sazonal Mult. Corrigido 1	Índ. Sazonal Mult. Corrigido 2	Índ. Sazonal Mult. Corrigido 3	Índ. Sazonal Mult. Corrigido 4
2013/1		1	1,004333259	1,098221739	1,107819257	1,107903864	1,107903871
2013/2		2	0,82215044	0,899007852	0,906864412	0,906933672	0,906933677
2013/3		3	1,225339093	1,339887948	1,351597422	1,351700647	1,351700655
2013/4		4	1,210631571	1,323805517	1,335374445	1,335476431	1,335476439
2013/5		5	1,031050755	1,127436877	1,137289711	1,137376568	1,137376575
2013/6		6	1,151909775	1,259594209	1,270601985	1,270699024	1,270699031
2013/7		7	0,850600877	0,930117932	0,938246367	0,938318024	0,938318029
2013/8		8	1,055397407	1,154059536	1,164145029	1,164233938	1,164233945
2013/9		9	0,69768441	0,762906316	0,769573464	0,769632239	0,769632243
2013/10		10	0,969363712	1,05998312	1,069246466	1,069328127	1,069328133
2013/11		11	0,744215368	0,813787145	0,820898949	0,820961643	0,820961648
2013/12		12	0,951083405	1,039993908	1,049082565	1,049162686	1,049162692
2013/13		13	1,200185753	1,312383188	1,323852295	1,323953401	1,323953409
2013/14		14	0,930758495	1,017768957	1,026663387	1,026741796	1,026741802
2013/15		15	0,996422582	1,089571545	1,099093468	1,099177409	1,099177415
2013/16		16	0,945474323	1,03386047	1,042895526	1,042975175	1,042975181
2013/17		17	1,000212893	1,093716188	1,103274332	1,103358592	1,103358598
2013/18		18	1,441077498	1,575794311	1,58965405	1,589686805	1,589686814
2013/19		19	0,599595244	0,655647441	0,661377239	0,66142775	0,661427754

Tabela 36 – Somatório de BH diferente de 8760 e somatório de BL (corrigido) igual a 8760.

BE	BF	BG	BH	BI	BJ	BK	BL
8742	2013/8741	0,787658366	8741	1,117881436	1,222384784	1,233067382	1,233161554
8743	2013/8742	1,498772263	8742	0,866337373	0,94732553	0,955604345	0,955677332
8744	2013/8743	0,629448619	8743	0,861204122	0,941712405	0,949942166	0,950014715
8745	2013/8744	0,329810933	8744	0,790248229	0,864123314	0,871675012	0,871741584
8746	2013/8745	1,17448564	8745	0,894634651	0,978268133	0,98681736	0,986892725
8747	2013/8746	0,660377119	8746	0,593485614	0,648966663	0,654638076	0,654668076
8748	2013/8747	0,036547462	8747	0,622427204	0,68601381	0,686561793	0,686614227
8749	2013/8748	1,064616502	8748	0,949595171	1,038366549	1,047440984	1,04752098
8750	2013/8749	1,208256434	8749	0,92517259	1,011660863	1,020501913	1,020579852
8751	2013/8750	0,260968996	8750	0,754688888	0,825239765	0,832451655	0,832515236
8752	2013/8751	0,896900114	8751	1,007722945	1,101928305	1,111558216	1,111643108
8753	2013/8752	1,197315608	8752	0,999845848	1,09331483	1,102869466	1,102953695
8754	2013/8753	0,906409789	8753	0,879030148	0,961204868	0,969604977	0,969679028
8755	2013/8754	2,810166923	8754	1,322119529	1,445715748	1,458350067	1,458461445
8756	2013/8755	1,210066509	8755	1,280871861	1,400612108	1,41285226	1,412960163
8757	2013/8756	1,612440455	8756	0,966596047	1,056956725	1,066193622	1,066275056
8758	2013/8757	1,307758723	8757	1,126515482	1,231825971	1,242591076	1,242685983
8759	2013/8758	2,012694783	8758	0,925039698	1,011551547	1,020355328	1,020433255
8760	2013/8759	1,745874637	8759	0,92960315	1,016505606	1,025388995	1,025467307
8761	2013/8760	0,8134324	8760	0,876523116	0,958463471	0,966839622	0,966913467
8762	2014/1	2,122259208	Soma 7941,085484		8683,445093	8759,330976	8759,999949
8763	2014/2	0,765855339	Diferença -0,93483392		-0,008739145	-7,63726E-05	-5,83278E-09
8764	2014/3	2,49709167	Fator 1,093483392		1,008739145	1,000076373	1,000000000
							8760

Semelhante ao modelo aditivo, observa-se as seguintes características:

- O número de períodos (horas) por ano é igual a 8760, coluna BE (Ano/Hora).
- Coluna BF (Índice Sazonal Mult.) dos primeiros 4380 períodos sem valor associado, sendo a primeira associação ocorrendo para o índice 4381 e a última para o período 65700.
- O somatório realizado sobre a coluna BH (Médias Internas Índ. Mult.) apresentando um valor diferente de 8760 e, somente após a quarta correção, a coluna BL (Índ. Sazonal Mult. Corrigido 4), resultando no valor esperado (8760).

4.1.4 – OBTENÇÃO DAS COMPONENTES CI

Os índices de CI (componentes ciclo e irregular) para os modelos aditivo e multiplicativo são obtidos através das equações 3 e 4, respectivamente, e aplicados para todos os períodos da série.

$$CI = Y - T - S \quad (3)$$

$$CI = Y / (T * S) \quad (4)$$

As Tabelas 37 e 38 listam os valores dos índices de CI e, na barra de funções, as metodologias de cálculo (equações 3 e 4) para os modelos aditivo e multiplicativo, respectivamente.

Tabela 37 – Valores calculados dos índices de CI para o modelo aditivo (coluna BR).

BR2		8760 índices Saz. Adt. Corr. (repetido p/cada ano)				BR
	BN	BO	BP	BQ	CI Aditivo (2013-2020)	
1	Ano/Hora	WS	T (potência)			
2	2013/1	1,9145	1,0199	0,145919054	0,748680946	
3	2013/2	0,27773	1,021102508	-0,137049045	-0,606323463	
4	2013/3	0,65602	1,021806587	0,495737743	-0,86152433	
5	2013/4	2,2498	1,022306433	0,482218839	0,745274728	
6	2013/5	2,0041	1,022694313	0,183905355	0,797500332	
7	2013/6	0,67779	1,023011342	0,358748947	-0,703970289	
8	2013/7	1,4011	1,023279464	-0,104452872	0,482273408	
9	2013/8	2,6229	1,023511778	0,239659821	1,359728401	
10	2013/9	1,7048	1,023716738	-0,319151342	1,000234604	
11	2013/10	2,7182	1,023900115	0,067299006	1,627000879	
12	2013/11	2,3504	1,024066028	-0,248697537	1,575031509	
13	2013/12	2,1356	1,024217519	0,09100634	1,020376141	
14	2013/13	1,7222	1,024356896	0,46090747	0,236935634	
15	2013/14	1,0948	1,024485956	0,062570924	0,00774312	
16	2013/15	0,4243	1,024606123	0,134667354	-0,734973478	
17	2013/16	1,0495	1,024718544	0,064924835	-0,040143379	
18	2013/17	0,19731	1,024824159	0,152273138	-0,979787298	
19	2013/18	0,68343	1,024923746	0,763079165	-1,104572911	
20	2013/19	1,7324	1,025017955	-0,482777002	1,190159047	

Conforme definido pela equação 3, o cálculo dos índices de CI no modelo aditivo (coluna BR) utiliza os valores de velocidade de vento da série (coluna BO), a tendência calculada através do modelo de regressão potência (coluna BP) e os 8760 representantes dos índices sazonais aditivos (coluna BQ), que se repetem para cada ano.

Tabela 38 – Valores calculados dos índices de CI para o modelo multiplicativo (coluna BY).

BY2		BW	BX	BY
1	Ano/Hora	8760 Índ. Saz. Mult. Corr. repetido p/cada ano		CI Multiplicativo (2013-2020)
2	2013/1	1,107903871	1,694321022	
3	2013/2	0,906933677	0,299901006	
4	2013/3	1,351700655	0,474971835	
5	2013/4	1,335476439	1,647883815	
6	2013/5	1,137376575	1,722936485	
7	2013/6	1,270699031	0,521401179	
8	2013/7	0,938318029	1,459233574	
9	2013/8	1,164233945	2,201144889	
10	2013/9	0,769632243	2,163766432	
11	2013/10	1,069328133	2,482634712	
12	2013/11	0,820961648	2,795702469	
13	2013/12	1,049162692	1,987398114	
14	2013/13	1,323953409	1,269870947	
15	2013/14	1,026741802	1,040800604	
16	2013/15	1,099177415	0,376745686	
17	2013/16	1,042975181	0,981982783	
18	2013/17	1,103358598	0,174495028	
19	2013/18	1,589686814	0,419460347	
20	2013/19	0,661427754	2,555255243	

Conforme definido pela equação 4, o cálculo dos índices de CI no modelo multiplicativo (coluna BY) utiliza os valores da velocidade do vento da série (coluna BO mostrada na tabela anterior), a tendência calculada através do modelo de regressão potência (coluna BP também apresentada na tabela anterior) e os 8760 representantes de índices sazonais multiplicativos (coluna BX), que se repetem para cada ano.

Os valores dos índices de CI para os modelos aditivos e multiplicativos foram plotados nas Figuras 46 e 47, respectivamente.

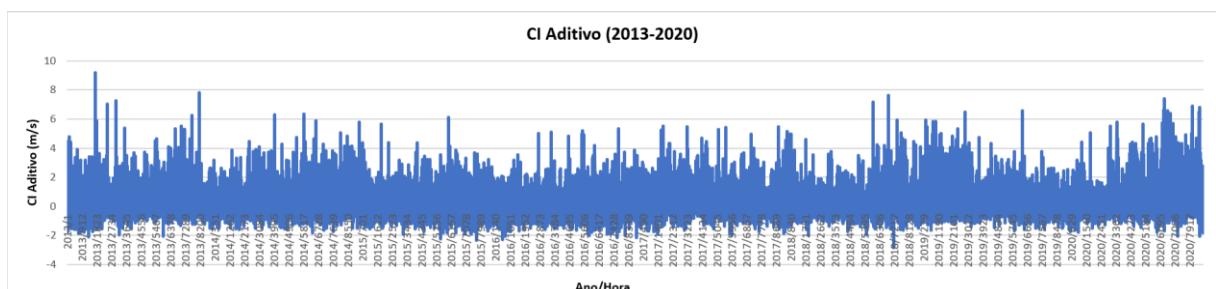


Figura 46 – Índices CI para o modelo aditivo.

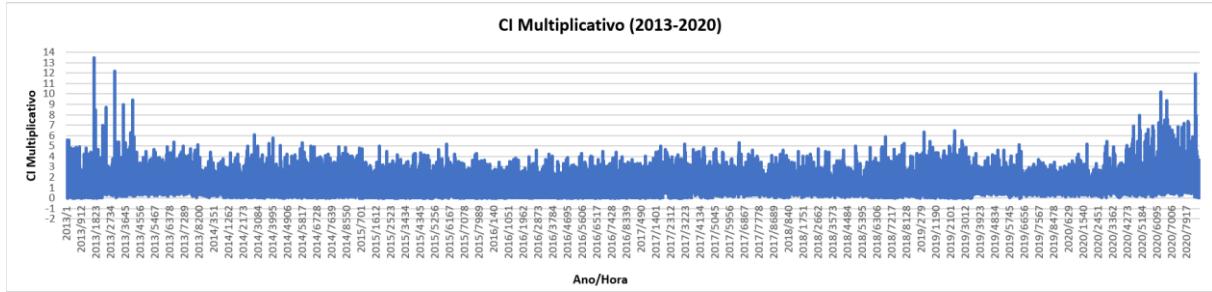


Figura 47 – Índices CI para o modelo multiplicativo.

Apesar da aglutinação dos valores nos gráficos acima, fica bem identificado o início e o fim de cada ano. Contudo, nenhum dos dois gráficos (mesmo quando devidamente expandidos) apresentam uma alternância sistemática de valores de alta e baixa por um período superior a um ano. No modelo aditivo, essa alternância deveria ocorrer em torno do valor 0 (zero) e, no multiplicativo, em torno do valor 1 (um).

A inexistência de alternância sistemática de intervalos de alta e baixa por um período superior a um ano significa que a série ou não apresenta influência da componente cíclica ou o número de anos analisados (2013-2020) não foi suficiente para evidenciá-la. Esta segunda opção é a mais plausível, visto que sabidamente admite-se a existência de ciclos de longo prazo envolvendo fenômenos meteorológicos. Esse fato compromete a recomposição da série e, por conseguinte, a previsão conforme definidas em 2.16.4.1, 2.16.4.2 e 2.16.5.

4.2 – DEFININDO A SÉRIE TEMPORAL PARA OS PRÓXIMOS MÉTODOS E MODELOS

Os métodos e modelos tratados a seguir utilizam a série temporal chamada WS_m, que armazena velocidades de vento coletadas a cada 15 minutos pela torre C em maio de 2016. A escolha dessa planilha se deve ao fato de seus valores não apresentarem originariamente quaisquer ocorrências de invalidade e/ou ausência (vide Tabela 21). A transformação logarítmica realizada sobre WS_m e a especificação dos conjuntos de treinamento e teste apresentados a seguir também serão os mesmos utilizados nos métodos e modelos seguintes, de modo a manter coerência durante a comparação de resultados.

A cada hora estão associados quatro valores de velocidade de vento registrados nos minutos 00, 15, 30 e 45. Sendo assim, a série WS_m conta com 2976 valores de velocidade de

vento correspondente aos 31 dias do mês de maio. A Figura 48 mostra a distribuição desses valores.

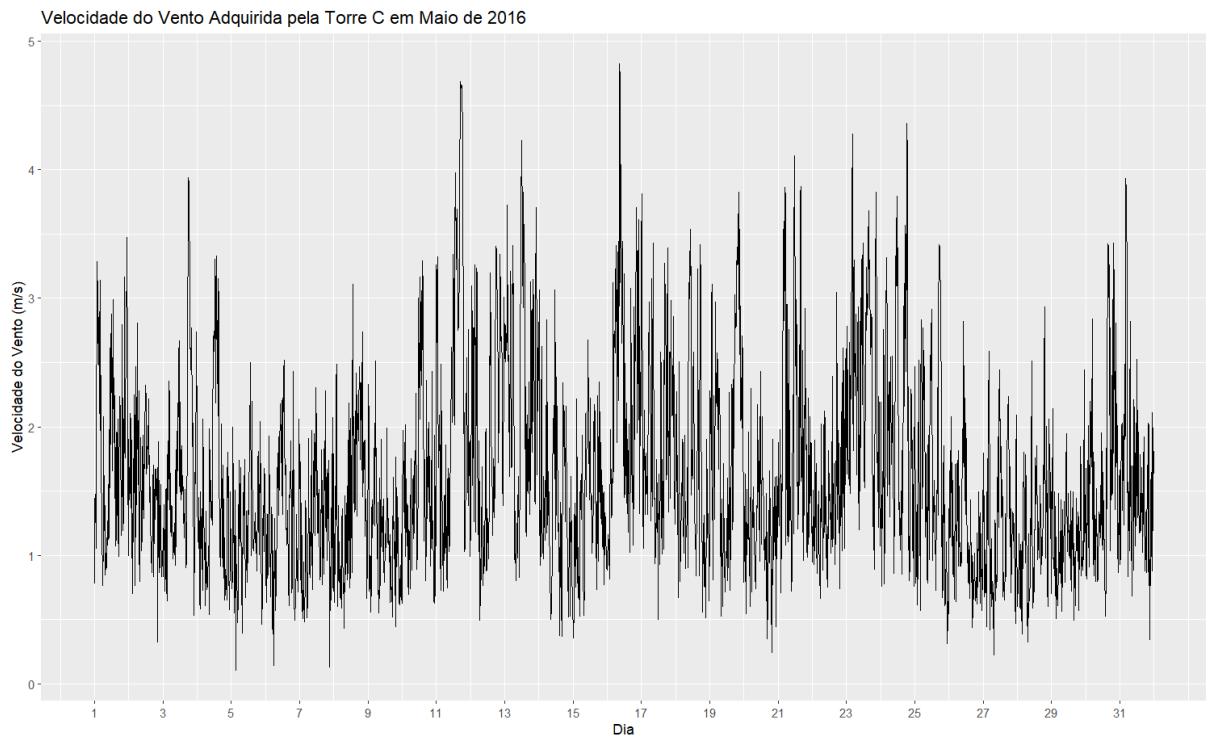


Figura 48 – Velocidade do vento registrada pela torre C em maio de 2016.

Para melhor visualização do comportamento diário da velocidade do vento, foram selecionados os 5 primeiros dias, conforme mostrado na Figura 49.

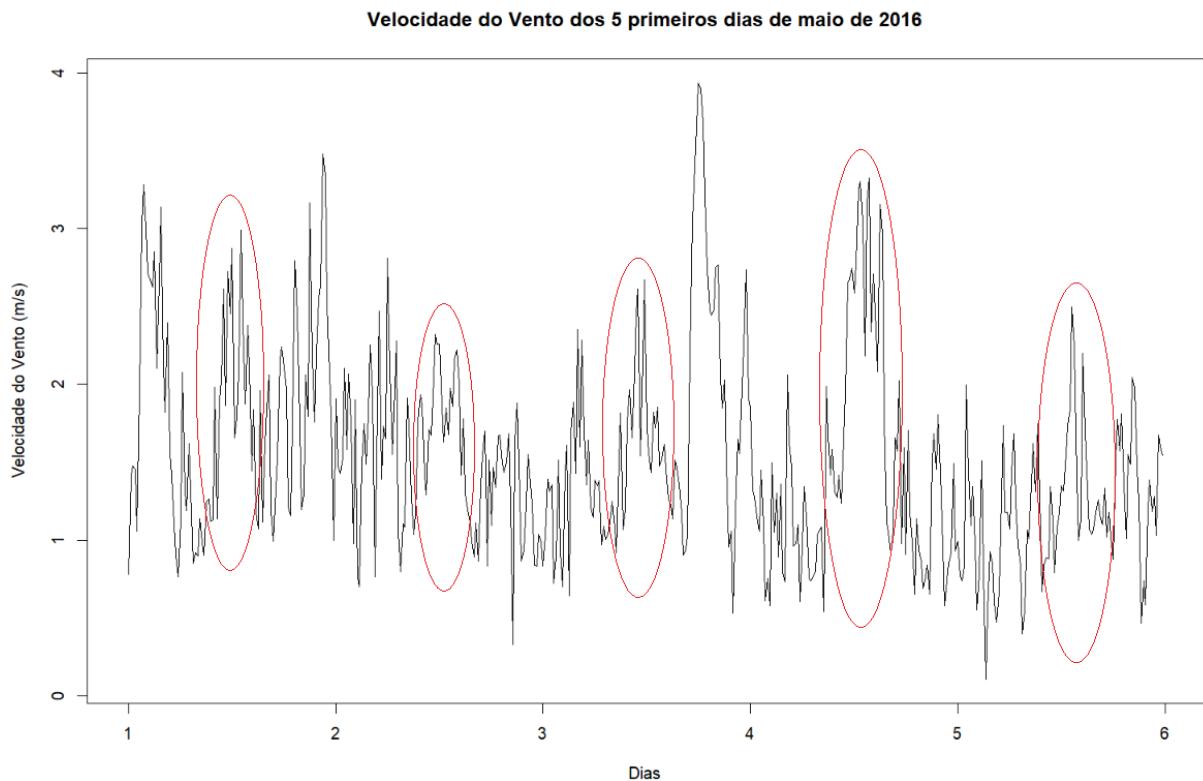


Figura 49 – Velocidade do vento dos cinco primeiros dias de maio.

Observa-se um razoável padrão sazonal (circundado em vermelho) ocorrendo aproximadamente na metade de cada dia, indicando assim um padrão de frequência aproximado de 48, que corresponde a 12 horas ($48=12 \times 4$, onde 4 é número de quartos em uma hora).

4.2.1 – APLICANDO A TRANSFORMAÇÃO LOGARÍTMICA SOBRE A SÉRIE WS_m

Com o intuito de estabilizar a variância existente na série WS_m, conforme mostrada na Figura 45, foi aplicada uma transformação logarítmica (log neperiano), resultando na distribuição apresentada pela Figura 50 abaixo.

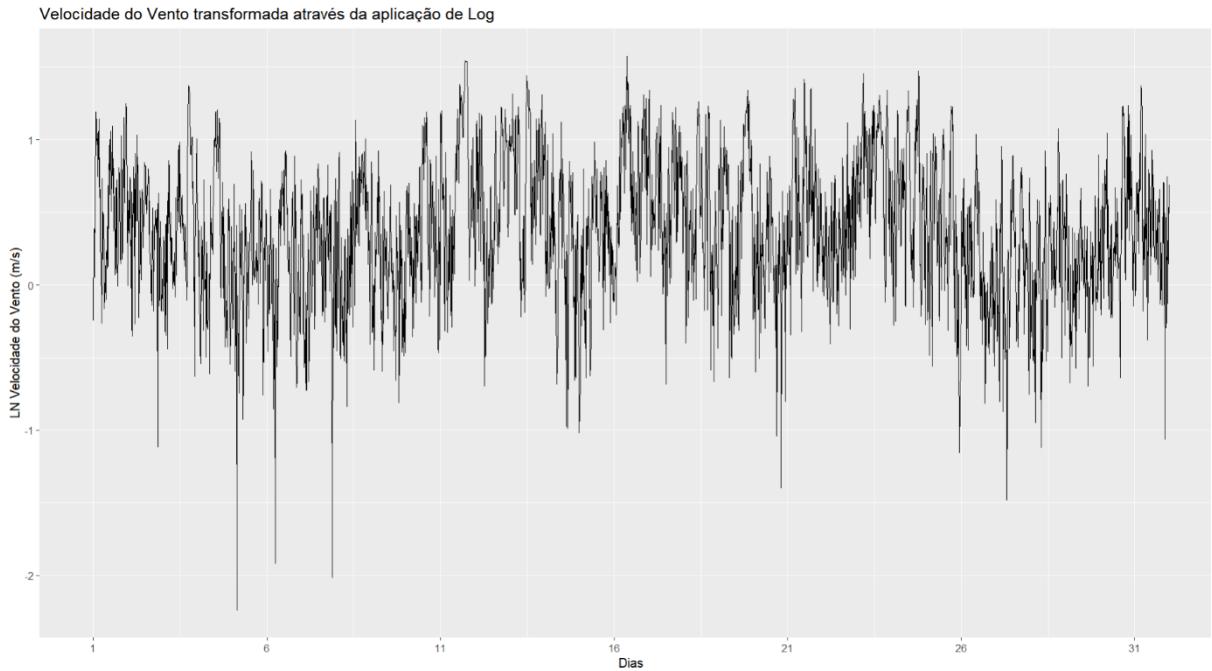


Figura 50 – Série WS_m transformada através da aplicação do LN.

Nota-se que a série sofreu uma leve melhora em relação à estabilidade de sua variância, cabendo também observar a alteração na escala das ordenadas devido à transformação logarítmica.

4.2.2 – DEFINIÇÃO DO CONJUNTO DE TREINAMENTO E TESTE SOBRE A SÉRIE WS_m APÓS SUA TRANSFORMAÇÃO LOGARÍTMICA

Na série temporal WS_m, cada dia é composto por 96 observações de velocidade do vento, obtidas a cada 15 minutos ($96 = 24 * 4$). Com o objetivo de comparar os próximos modelos estudados com o modelo de previsão adotado pelo SCA num ponto crítico, a escolha da extensão do conjunto de treinamento procurou satisfazer dois critérios. O primeiro diz respeito à sua dimensão percentual, devendo estar entre os limites $\geq 60\%$ e $\leq 70\%$ do total de valores registrados na série WS_m (2976 observações). O segundo critério buscou encontrar um ponto crítico, isto é, onde o último valor do conjunto de treinamento fosse significativamente “destoante” do primeiro valor do conjunto de teste. Atendendo aos critérios acima, o conjunto de treinamento selecionado conta com 1942 observações, correspondendo a 65.25% ou 20 dias mais 22 observações ($1942 = 20*96 + 22$) e o conjunto de teste correspondendo a 34.75% do total, contando com 1034 observações ou 10 dias mais 74 observações ($1034 = 10*96 + 74$). A Figura 51 apresenta toda a distribuição da série velocidade

do vento WS_m após sua transformação logarítmica, distinguindo o conjunto de treinamento em vermelho e o conjunto de teste em verde.

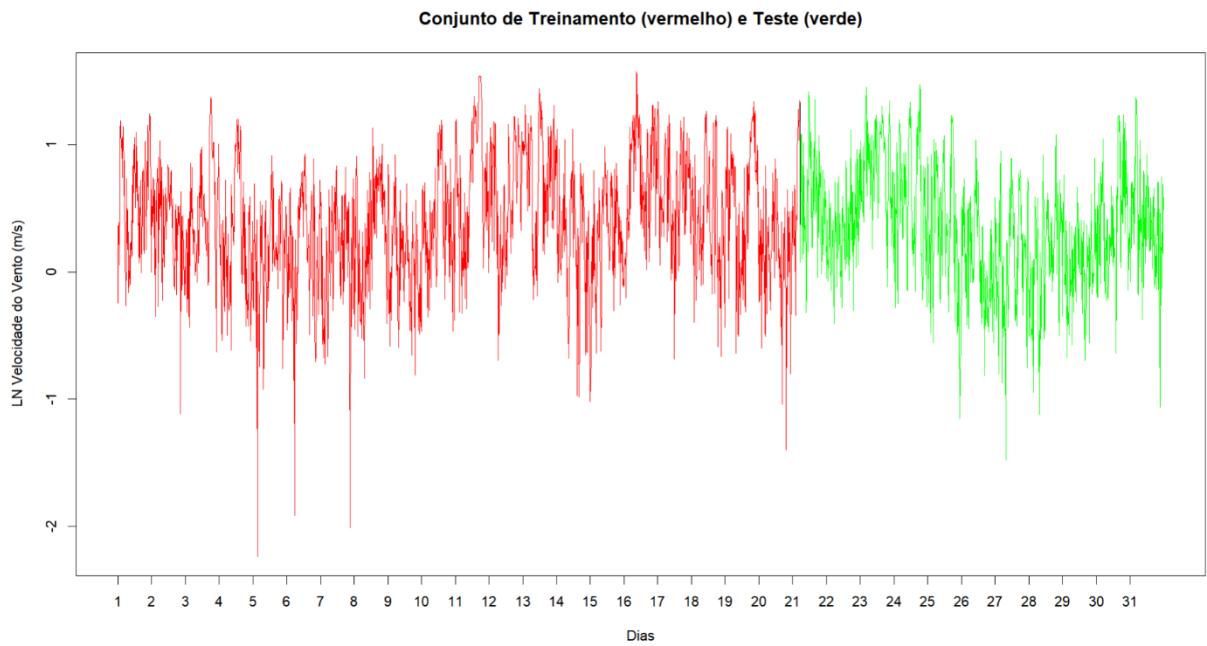


Figura 51 –Conjunto de Treinamento (vermelho) e Teste (verde) de WS_m após sua transformação LN.

Para melhor visibilidade do ponto crítico escolhido, a Figura 52 apresenta as 5 últimas horas do conjunto de treinamento (vermelho) e as 5 primeiras horas do conjunto de teste (verde), destacando o último valor do conjunto de treinamento (3.783 m/s) e o primeiro valor do conjunto de teste (1.0787). Ainda com intuito de gerar uma percepção real, a Figura 52 foi confeccionada a partir da série WS_m antes da transformação logarítmica.

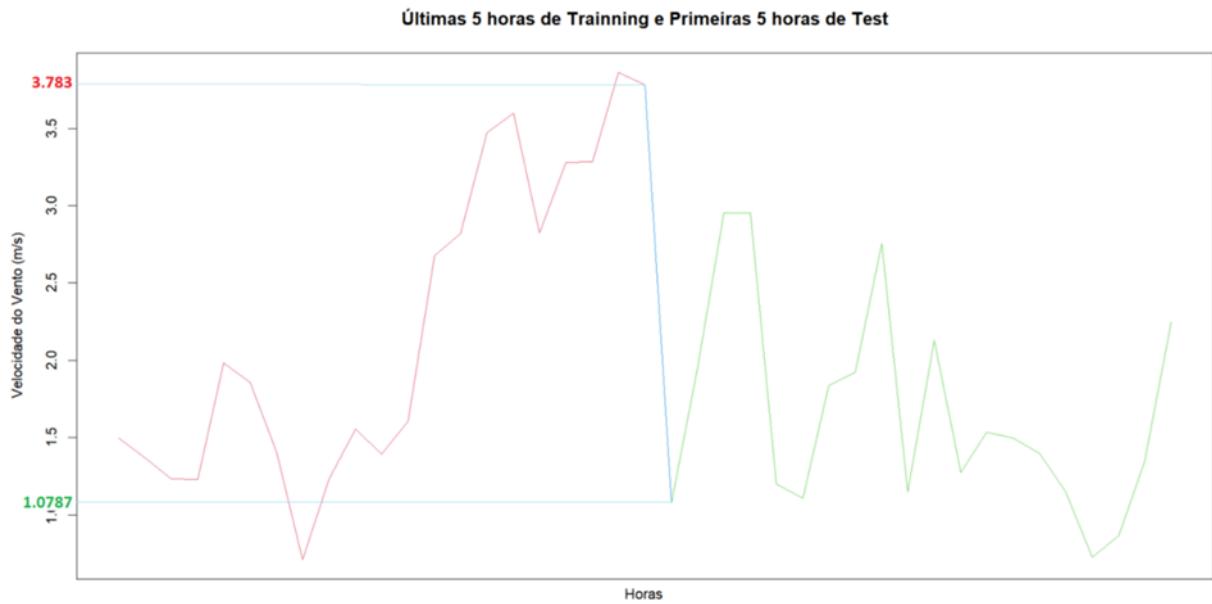


Figura 52 – 5 últimas horas do conjunto de treinamento e 5 primeiras horas do teste.

4.3 – APLICAÇÃO DO MÉTODO DE HOLT-WINTERS

A função *HoltWinters()* do pacote estatístico R, foi utilizada para obtenção deste ajuste sobre o conjunto de treinamento. A Figura 53 apresenta em preto a distribuição do conjunto de treinamento e em vermelho o ajuste Holt-Winters obtido sobre este conjunto.

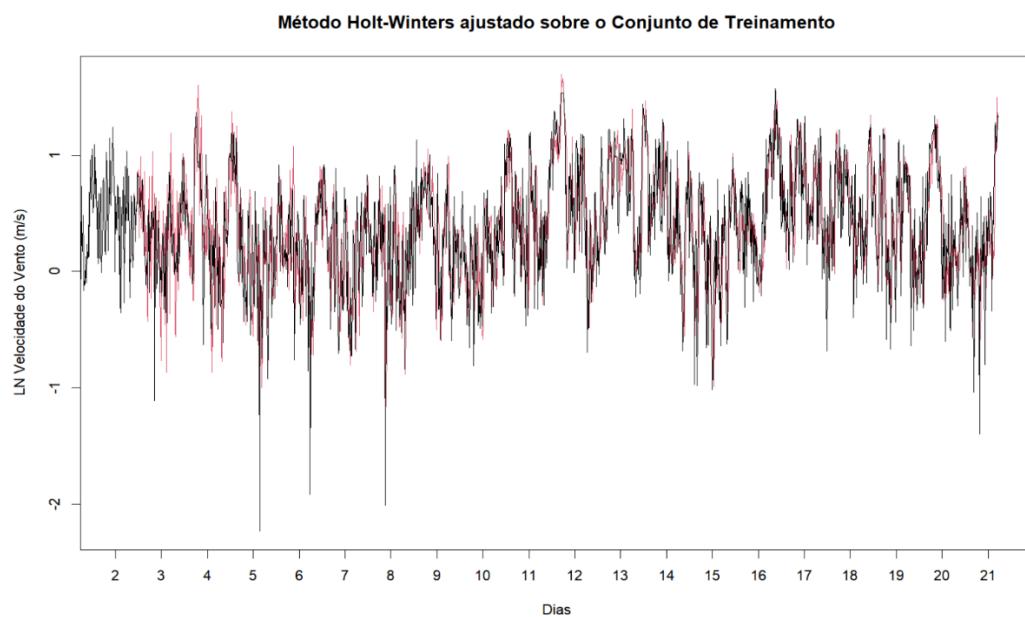


Figura 53 – Ajuste Holt-Winters (vermelho) sobre o conjunto de treinamento (preto).

Analisando a figura acima, pode-se considerar que o ajuste do modelo Holt-Winters sobre o conjunto de treinamento ocorreu de maneira bastante satisfatória. O próximo passo consiste em analisar os resíduos produzidos pela aplicação do método Holt-Winters. O desejável é que os resíduos apresentem um comportamento semelhante a um ruído branco.

- Análise da distribuição dos resíduos.

A Figura 54 apresenta a distribuição dos resíduos produzidos pela aplicação do método Holt-Winters sobre o conjunto de treinamento. Pode-se constatar que a variação de seus valores gira em torno de 0 (zero) como esperado e, realmente, o valor da média calculada é 0.003285.

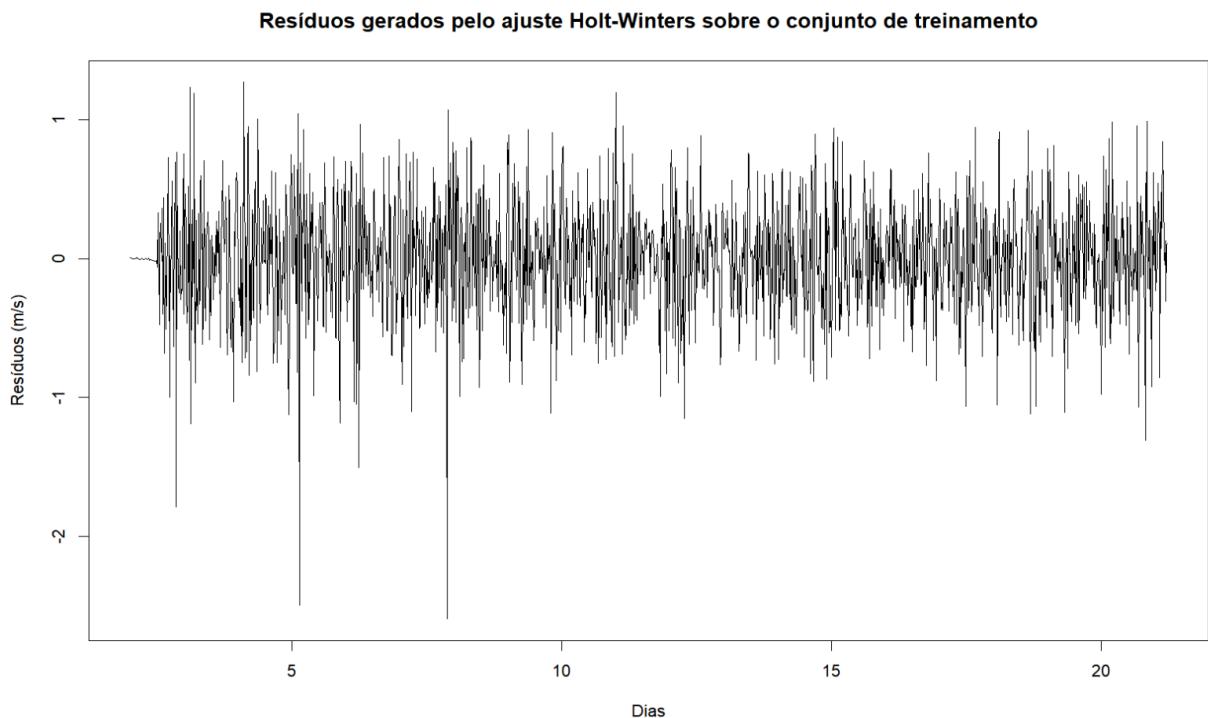


Figura 54 – Gráfico dos resíduos gerados pelo método Holt-Winters.

- Análise de normalidade dos resíduos do modelo.

A Figura 55 esboça o gráfico *QQ-plot* gerado pelos resíduos e a linha vermelha indica uma distribuição normal ou gaussiana. Como constatado no gráfico abaixo, o número de pontos sobre a reta parece indicar a não normalidade dos resíduos, contudo, isso ainda

necessita ser comprovado através do seu histograma e do teste de hipóteses de Shapiro-Wilker.

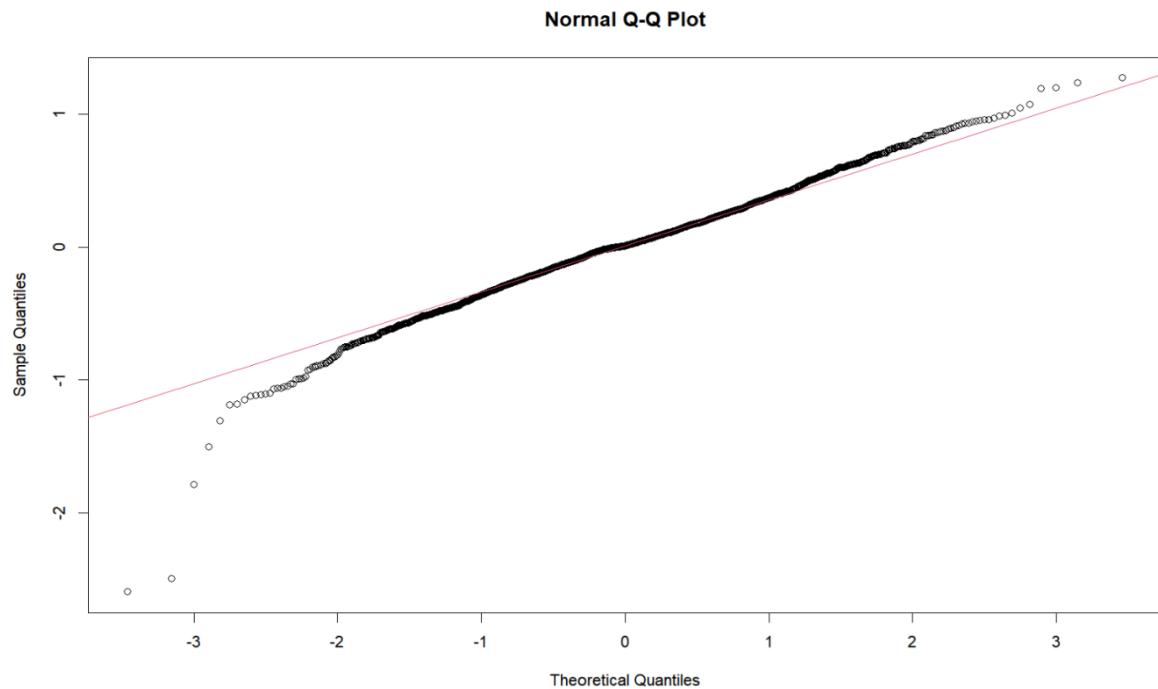


Figura 55 – Gráfico *QQ-plot* dos resíduos gerados pelo modelo Holt-Winters.

A Figura 56 mostra o histograma e a curva densidade de probabilidade dos resíduos gerados pelo método Holt-Winters.

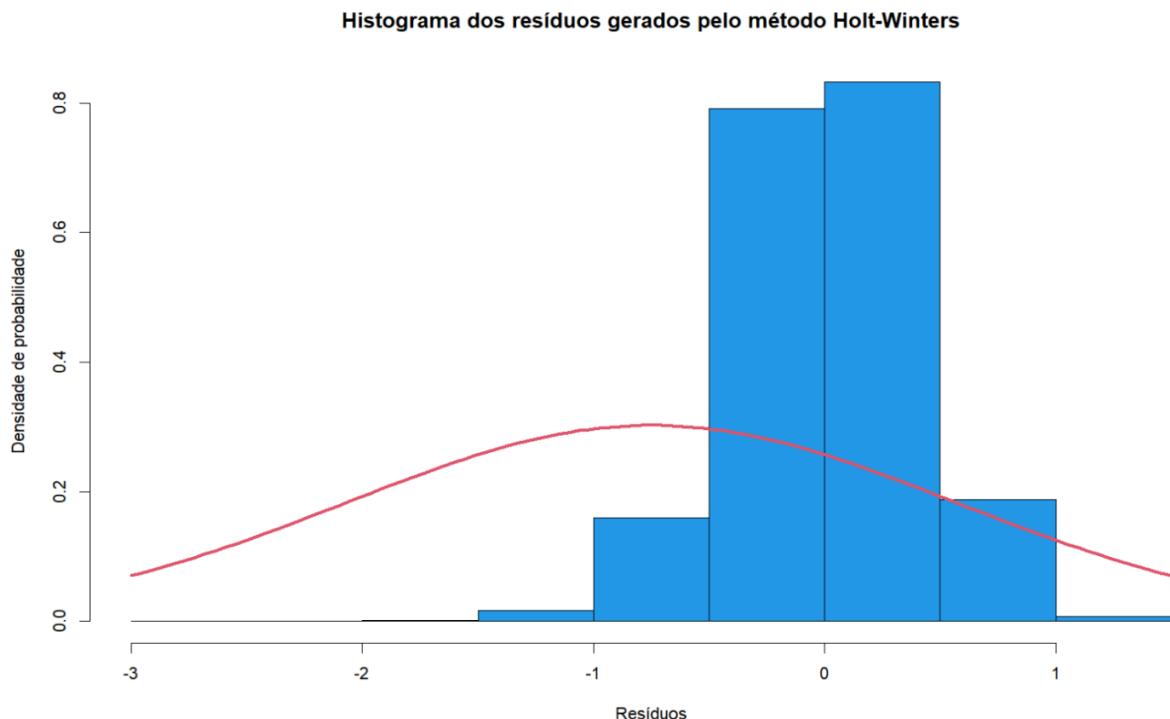


Figura 56 – Histograma e curva densidade de probabilidade dos resíduos gerados pelo modelo Holt-Winters.

A Tabela 39 apresenta o valor do teste de hipóteses de Shapiro-Wilker. Como o valor do teste é muito menor que o nível de significância, o teste indica não normalidade.

Tabela 39 – Teste de normalidade Shapiro-Wilker para os resíduos gerados pelo modelo Holt-Winters.

Valor do teste Shapiro-Wilker	Nível de Significância
3.251e-14	0.05

As três análises acima comprovam que os resíduos gerados pela aplicação do modelo Holt-Winters sobre o conjunto de treinamento não seguem uma distribuição normal, como esperado.

- Análise do correlograma FAC dos resíduos.

A literatura sugere a escolha de um *lag máximo* igual a $10 * \log_{10}(N)$, que corresponde a aproximadamente 33 *lags* ($10 * \log_{10}1942 = 32,882$). Neste trabalho, para a exibição dos correlograma FAC dos resíduos, optou-se por um *lag* máximo igual 192, de modo a cobrir

quatro períodos sazonais ($4*48$). A Figura 57 representa o correlograma FAC dos resíduos obtidos pelo método Holt-Winters.

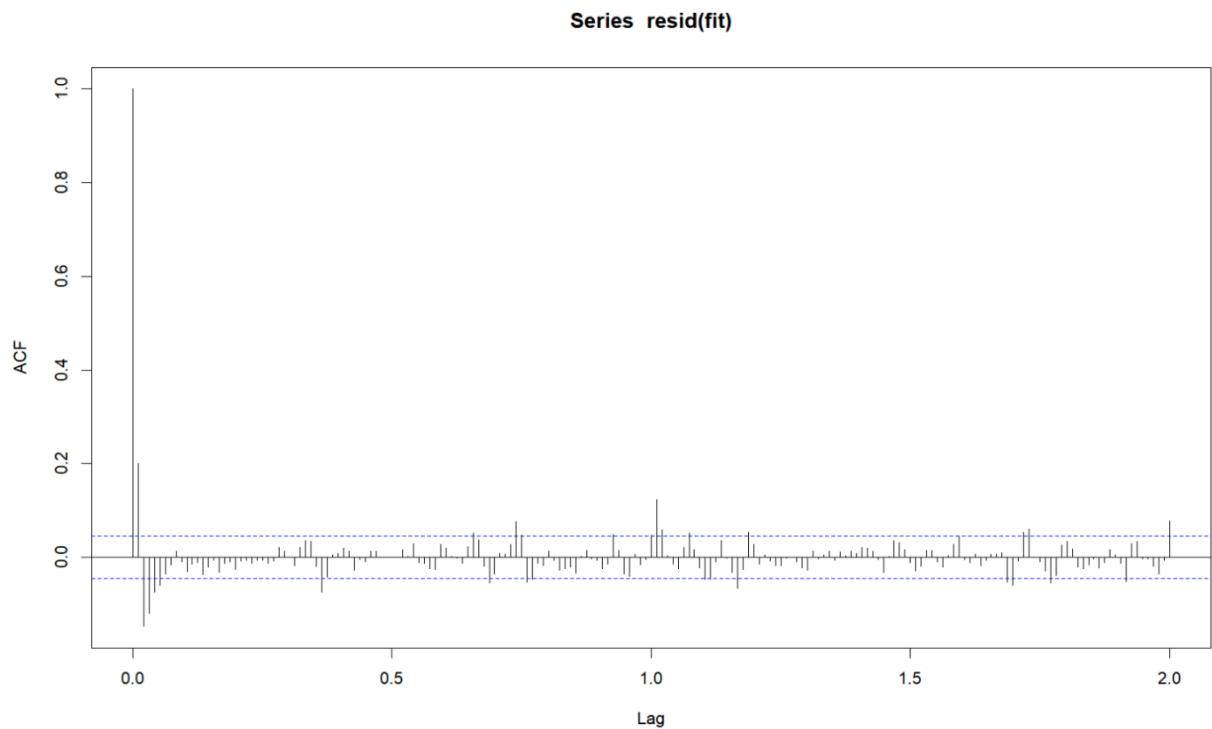


Figura 57 – Correlograma FAC dos resíduos gerados pelo modelo Holt-Winters.

Como pode-se observar na figura acima os valores dos *lags* (abscissa) estão representados em escala, onde 0.5 corresponde ao *lag* 48. Em teoria, o correlograma FAC dos resíduos deveria apresentar todos os *lags* dentro do intervalo de confiança (95%), indicando assim uma independência. Contudo, segundo Hyndman *at al.* (2018), um valor aproximado de até 5% de *lags* fora do intervalo de confiança é aceitável, entretanto o que se observa é um valor bem acima desse limite. Conclui-se, então, que o método não foi capaz de capturar toda a autocorrelação da série.

A Figura 58 esboça a parte final do conjunto de treinamento (vermelho), todo o conjunto de teste (verde) e o ajuste Holt-Winters (azul) sobre este último.

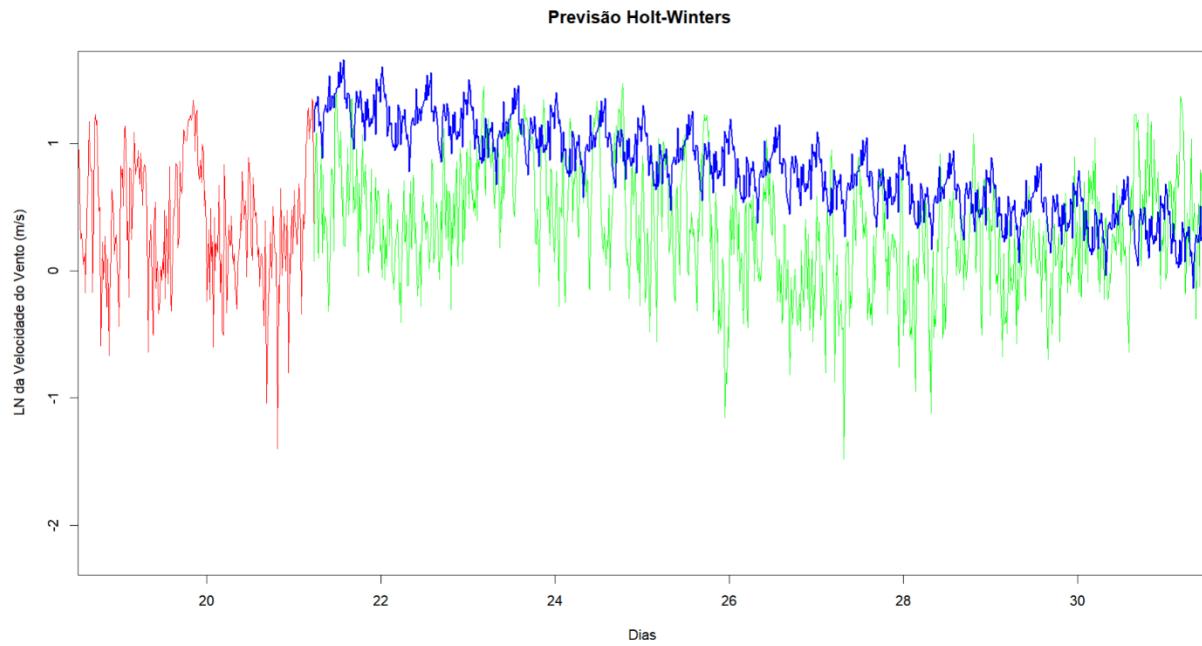


Figura 58 – Parte do conjunto de treinamento (vermelho), todo o conjunto de teste (verde) e a previsão Holt-Winters (azul) sobre o conjunto de teste.

Conforme pode ser constatado do gráfico acima, a previsão para os primeiros valores se mostra adequada assim como a captura do efeito sazonal, mas a partir desse ponto o modelo se descola acentuadamente do conjunto de teste. Isso pode ser explicado pelo fato de o período de previsão ser extremamente longo (1034 valores previstos).

O valor da métrica de acurácia MAPE da previsão Holt-Winters para os oito primeiros valores do conjunto de teste, é apresentado na Tabela 40 e a Tabela 41 mostra o MAPE da previsão Holt-Winters para todo o conjunto de teste (1034 observações).

Tabela 40 – Medidas de acurácia da previsão Holt-Winters para os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão Holt-Winters	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,091907983	-0,01320798	0,01321	0,00017445	-1,22443523	1,2244352	0,71758319	0,871605	24,3762683	31,429485
1,9438	1,266908871	0,676891129	0,67689	0,4581816	34,8230851	34,823085				
2,9541	1,321728555	1,632371445	1,63237	2,66463654	55,2578262	55,257826				
2,9527	1,294049538	1,658650462	1,65865	2,75112135	56,1740259	56,174026				
1,1976	1,371433989	-0,17383399	0,17383	0,03021826	-14,5151962	14,515196				
1,1067	1,244741294	-0,13804129	0,13804	0,0190554	-12,4732352	12,473235				
1,8351	1,139366134	0,695733866	0,69573	0,48404561	37,912586	37,912586				
1,9253	1,173364648	0,751935352	0,75194	0,56540677	39,0554901	39,05549				

Tabela 41 – Medidas de acurácia da previsão Holt-Winters para todo o conjunto de teste.

AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
Valores Reais	Previsão Holt-Winters	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
1,0787	1,091907983	-0,01320798	0,01320798	0,00017445	-1,22443523	1,22443523	0,83029674	1,17024961	41,300843	47,79052
1,9438	1,266908871	0,676891129	0,67689113	0,4581816	34,82308513	34,8230851				
2,9541	1,321728555	1,632371445	1,63237145	2,66463654	55,25782625	55,2578262				
2,9527	1,294049538	1,658650462	1,65865046	2,75112135	56,17402587	56,1740259				
1,1976	1,371433989	-0,17383399	0,17383399	0,03021826	-14,5151962	14,5151962				
1,1067	1,244741294	-0,13804129	0,13804129	0,0190554	-12,4732352	12,4732352				
1,8351	1,139366134	0,695733866	0,695733867	0,48404561	37,91258599	37,912586				
1,9253	1,173364648	0,751935352	0,75193535	0,56540677	39,05549015	39,0554901				
2,7568	1,079870391	1,676929609	1,67692961	2,81209291	60,82884538	60,8288454				
1,1488	0,882384625	0,266415375	0,26641538	0,07097715	23,19075343	23,1907534				
2,1292	1,091525198	1,037674802	1,0376748	1,07676899	48,73543122	48,7354312				
1,2756	1,239985913	0,035614087	0,03561409	0,00126836	2,791947893	2,79194789				
1,5334	1,275520607	0,257879393	0,25787939	0,06650178	16,81749007	16,8174901				
1,4949	1,303742569	0,191157431	0,19115743	0,03654116	12,78730557	12,7873056				
1,3953	1,255889512	0,139410488	0,13941049	0,01943528	9,991434664	9,99143466				
1,1473	1,242819887	-0,09551989	0,09551989	0,00912405	-8,32562424	8,32562424				
0,7244	1,365011699	-0,6406117	0,6406117	0,41038335	-88,4334207	88,4334207				
0,8622	1,532095276	-0,66989528	0,66989528	0,44875968	-77,6960422	77,6960422				
1,3472	1,259563203	0,087636797	0,0876368	0,00768021	6,505106668	6,50510667				
2,2483	1,283494938	0,96480562	0,96480562	0,93084988	42,91267269	42,9126727				
2,1999	1,375065262	0,824834738	0,82483474	0,68035235	37,49419238	37,4941924				
1,9508	1,359450307	0,59134963	0,59134963	0,34969439	30,31318589	30,3131859				
1,166	1,278733352	-0,11273335	0,11273335	0,01270881	-9,66838349	9,66838349				
3,3066	1,391229664	1,915370336	1,91537034	3,66864352	57,92567398	57,925674				
4,1093	1,3936675	2,7156325	2,7156325	7,37465987	66,0850388	66,0850388				
3,9172	1,429592947	2,487607053	2,48760705	6,18818885	63,50472411	63,5047241				
2,7303	1,443639924	1,286660076	1,28666008	1,65549415	47,12522711	47,1252271				
2,6955	1,563598067	1,131901933	1,13190193	1,28120199	41,99228094	41,9922809				
2,2498	1,439869126	0,809930874	0,80993087	0,65598802	36,00012776	36,00012778				
2,9872	1,639178181	1,348021819	1,34802182	1,81716282	45,1266008	45,1266008				
2,6069	1,475035152	1,131816488	1,13181649	1,28100854	43,41618349	43,4161835				
1,5367	1,539382643	-0,00268264	0,00268264	7,1966E-06	-0,17457169	0,17457169				
1,9246	1,600292199	0,324307801	0,3243078	0,10517555	16,85065993	16,8506599				
1,2238	1,660571332	-0,43677133	0,43677133	0,19076962	-35,689764	35,689764				
1,2072	1,3505249	-0,1433249	0,1433249	0,02054203	-11,8725066	11,8725066				
1,4937	1,260001001	0,062708000	0,06270800	0,00405271	4,474010772	4,47401077				

4.4 – APLICAÇÃO DO MODELO AUTOREGRESSIVO SARIMA

4.4.1 – AVALIAÇÃO DA ESTACIONARIEDADE DA SÉRIE

A utilização de modelos autorregressivos exige que a série seja do tipo estacionária ou que possa ser transformada em tal, visto que modelos da classe ARIMA(p, d, q) podem realizar processos de diferenciação com esse objetivo. Então, a primeira providência antes da aplicação de um modelo autorregressivo é examinar a estacionariedade da série. Nesse trabalho foram utilizados os testes de estacionariedade de Dickey-Fuller Aumentado (DFA) e KPSS, executados no ambiente computacional R (pacote *urca*), através da utilização das funções *ur.df* e *ur.kpss*, respectivamente. O critério de informação adotado nos testes foi o AIC (Akaike) ao invés do BIC (bayesiano), pois este é mais parcimonioso que aquele.

A série a ter a sua estacionariedade avaliada é a WS_m (velocidade do vento enviada pela torre C no mês de maio de 2016 com periodicidade de 15 minutos). Após a aplicação do teste DFA, seu resultado é confirmado através do teste KPSS.

A princípio, o teste DFA foi executado utilizando-se os valores default da função *ur.df*, porém, a análise de resíduos do teste através das FAC e FACP apresentaram as configurações mostradas na Figura 59.

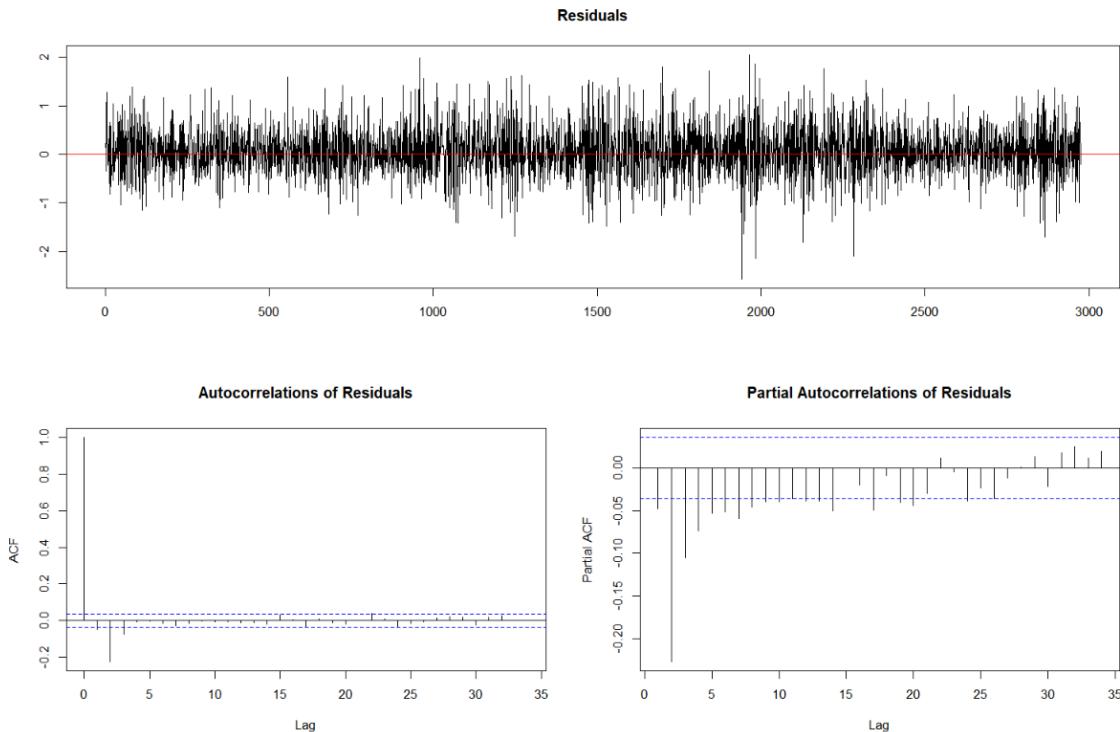


Figura 59 – Análise dos resíduos após aplicação de ADF com valores default.

Através dos gráficos de FAC e FACP dos resíduos do teste DFA, observa-se a existência de *lags* ultrapassando o intervalo de confiança (95%), indicando que esses apresentam correlação e não configuraram um ruído branco. Diante disso, faz-se necessário aumentar o número de *lags* utilizados na função *ur.df* até que não haja mais quaisquer autocorrelações. Esse objetivo foi alcançado a partir do *lag* 38 (trinta e oito), conforme pode ser verificado nos gráficos da Figura 60.

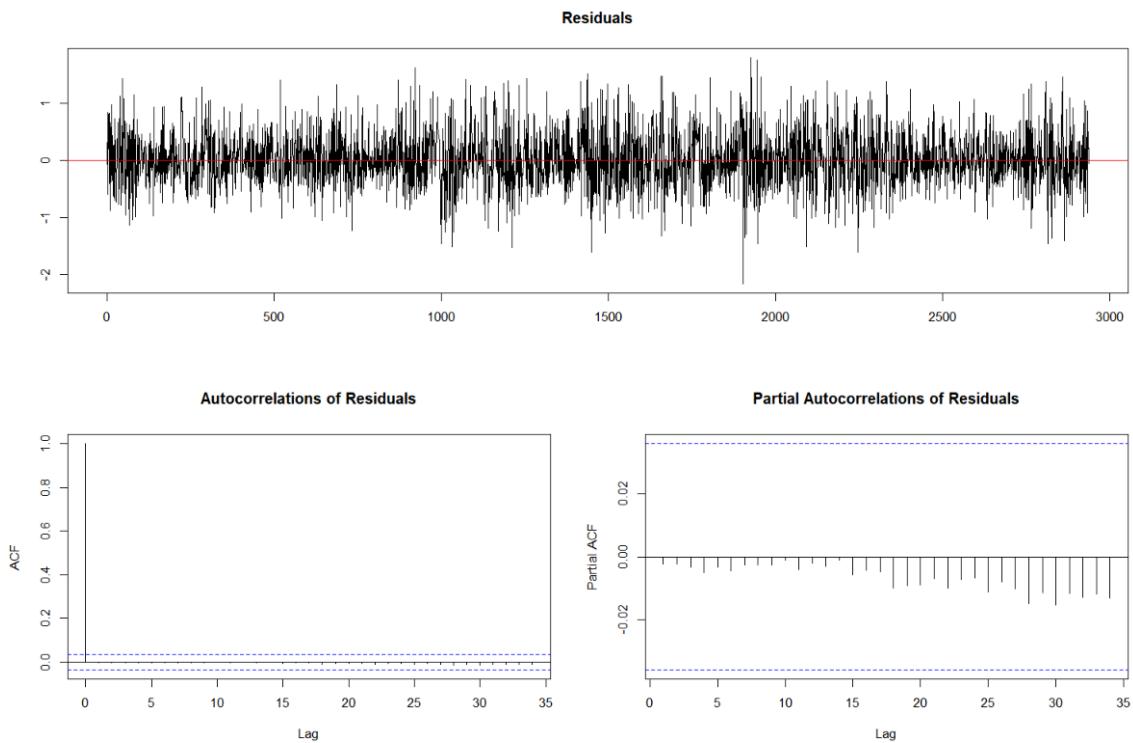


Figura 60 – Análise dos resíduos após aplicação do ADF com $lag = 38$.

Assim, após a inclusão do parâmetro $lags=38$ na sintaxe da função `ur.df`, o teste estatístico ADF foi executado e apresentou os resultados mostrados na Tabela 42.

Tabela 42 – Valores do teste de raiz unitária ADF para a série WS_m não diferenciada.

Valor do Teste ADF	Valores Críticos		
	1%	5%	10%
-1.5114	-2.58	-1.95	-1.62

Interpretação dos resultados:

Como o valor do teste (-1.5114) é maior que todos os valores críticos (1%, 5% e 10%), não há motivos para se rejeitar a hipótese nula (H_0) do teste ADF, que indica a existência de pelo menos uma raiz unitária e, por conseguinte, o estado de não estacionariedade da série.

Diante dessa interpretação, foi realizada uma diferenciação de 1 (um) sobre a série e a repetição do teste, emitindo os resultados mostrados na Tabela 43.

Tabela 43 – Valores do teste de raiz unitária ADF para a série WS_m após uma diferenciação I(1).

Valor do Teste ADF	Valores Críticos		
	1%	5%	10%
-13.3870	-2.58	-1.95	-1.62

Interpretação dos resultados:

Como o valor do teste (-13.387) é menor que todos os valores críticos (1%, 5% e 10%), pode-se rejeitar a hipótese nula (H_0) e aceitar-se a hipótese alternativa (H_a), indicando assim que a série diferenciada é estacionária, isto é, não possui raiz unitária e é dita ser integrada de ordem 1 (um) ou I(1).

A fim de se confirmar os resultados do teste ADF, foi utilizado o teste de raiz unitária KPSS através da aplicação da função *ur.kpss*, sobre a série WS_m não diferenciada, resultando nos valores mostrados na Tabela 44.

Tabela 44 – Valores do teste de raiz unitária KPSS para a série WS_m não diferenciada.

Valor do Teste KPSS	Valores Críticos			
	10%	5%	2%	1%
4.2741	0.347	0.463	0.574	0.739

Interpretação dos resultados:

A estatística do teste apresenta valor maior que os valores críticos de 10%, 5%, 2.5% e 1%, indicando que a hipótese nula é rejeitada, ou seja, a série é não estacionária. Em outras palavras, a série apresenta pelo menos uma raiz unitária, que leva à rejeição da hipótese nula (H_0) do KPSS. Cabe lembrar que as hipóteses do teste KPSS são contrárias às hipóteses do teste ADF. Após a primeira diferenciação e nova aplicação do KPSS, obtém-se o resultado apresentado pela Tabela 45.

Tabela 45 – Valores do teste de raiz unitária KPSS para a série WS_m após uma diferenciação I(1).

Valor do Teste KPSS	Valores Críticos			
	10%	5%	2%	1%
0.001	0.347	0.463	0.574	0.739

Interpretação dos resultados:

Diante desses valores, a hipótese nula do KPSS não pode ser rejeitada, confirmando as conclusões obtidas pelo teste ADF. Então, com apenas uma diferenciação a série tornou-se estacionária e isso garante a possibilidade de bons resultados preditivos através da aplicação de modelos autorregressivos. Essa possibilidade de estacionariedade, com diferenciação I(1), permite a aplicação do modelo ARIMA, no caso a variante SARIMA.

4.4.2 – DETERMINAÇÃO DAS ORDENS DO MODELO SARIMA

Neste momento cabe esclarecer que no software estatístico R a função chamada *Arima* acompanhada das ordens (p,d,q) , (P,D,Q) e um valor numérico $[m]$, se refere ao modelo $SARIMA(p,d,q)(P,D,Q)[m]$ com m períodos sazonais.

O próximo passo consiste na determinação dos valores mais adequados para as ordens (p,d,q) e (P,D,Q) do modelo SARIMA. O fato de a série apresentar a possibilidade de se tornar estacionária após uma diferenciação I(1) já indica o valor do parâmetro $d = 1$. Após a aplicação do modelo SARIMA, é aconselhável avaliar a estabilidade de suas raízes unitárias, de modo a garantir uma estabilidade estatística que possa assegurar a capacidade preditora do modelo. Por exemplo, o modelo $SARIMA(4,1,2)(0,01)[96]$ apresenta duas raízes unitárias externas (vermelho) ao círculo, Figura 61, não atendendo a esse critério de estabilidade estatística.

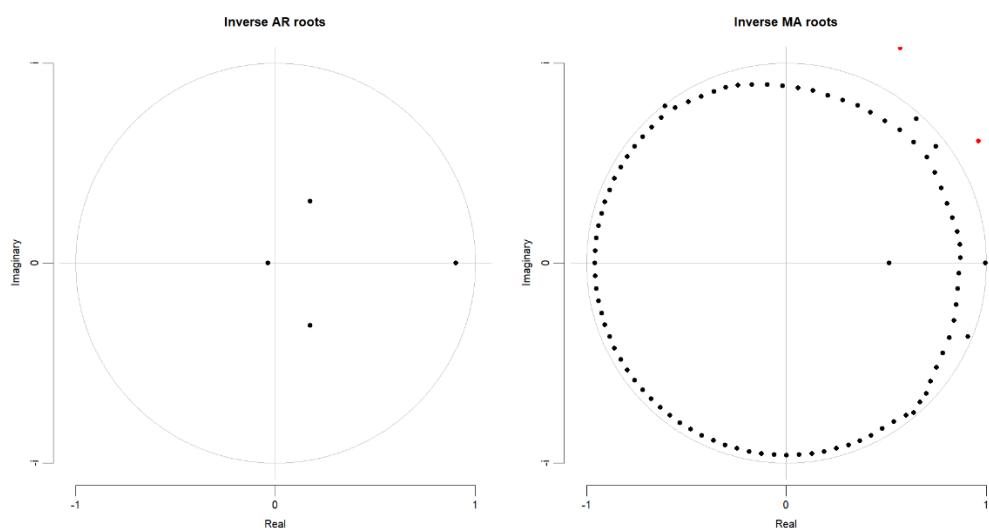


Figura 61 – Existência de 2 raízes (vermelho) fora do círculo unitário.

1 - Análise da série utilizada

Conforme descrito na seção 4.2, foi realizada uma transformação logarítmica sobre a série WS_m, com o objetivo de estabilizar a sua variância. Esse procedimento resultou na distribuição mostrada pela Figura 62.

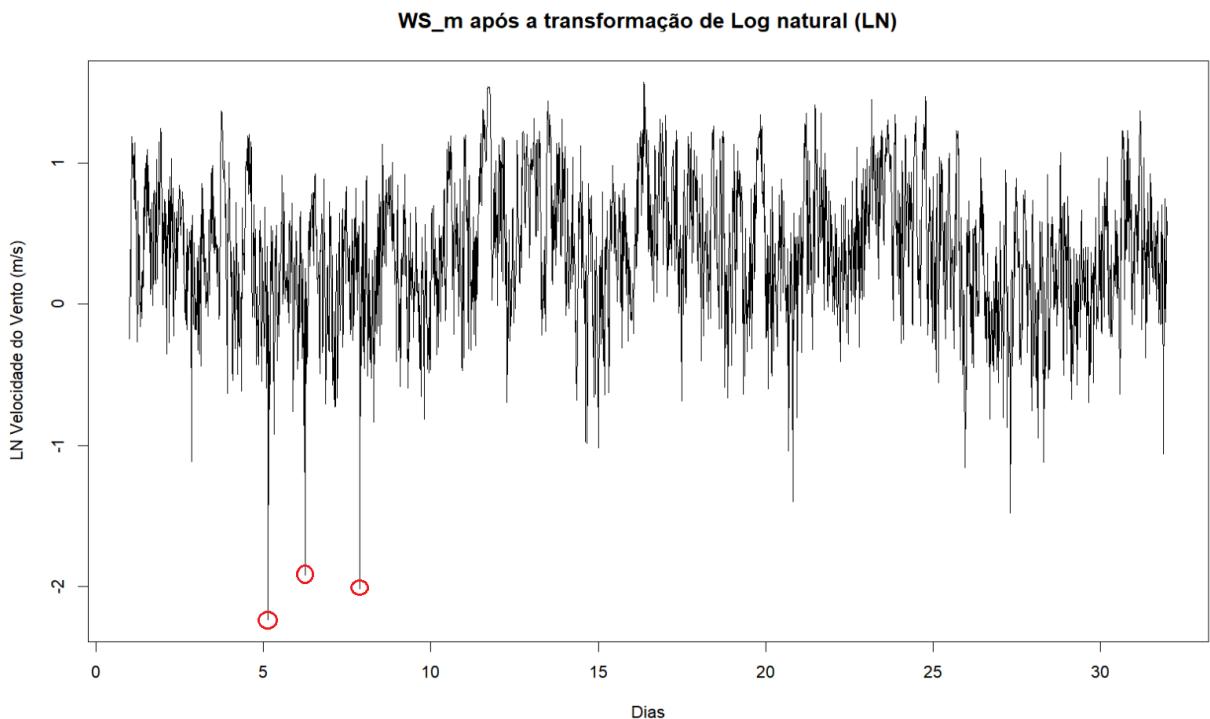


Figura 62 – Série WS_m após uma transformação logarítmica.

Após a transformação logarítmica, observa-se a presença de picos indesejados, circundados em vermelho. Então, foi aplicada uma nova transformação, desta feita através da função *tsclean* disponibilizada pelo *software R*, que identifica possíveis *outliers* e os substitui a partir de uma interpolação linear. A Figura 63 exibe o resultado obtido após a aplicação da função *tsclean* sobre a série WS_m previamente transformada através de logaritmo.

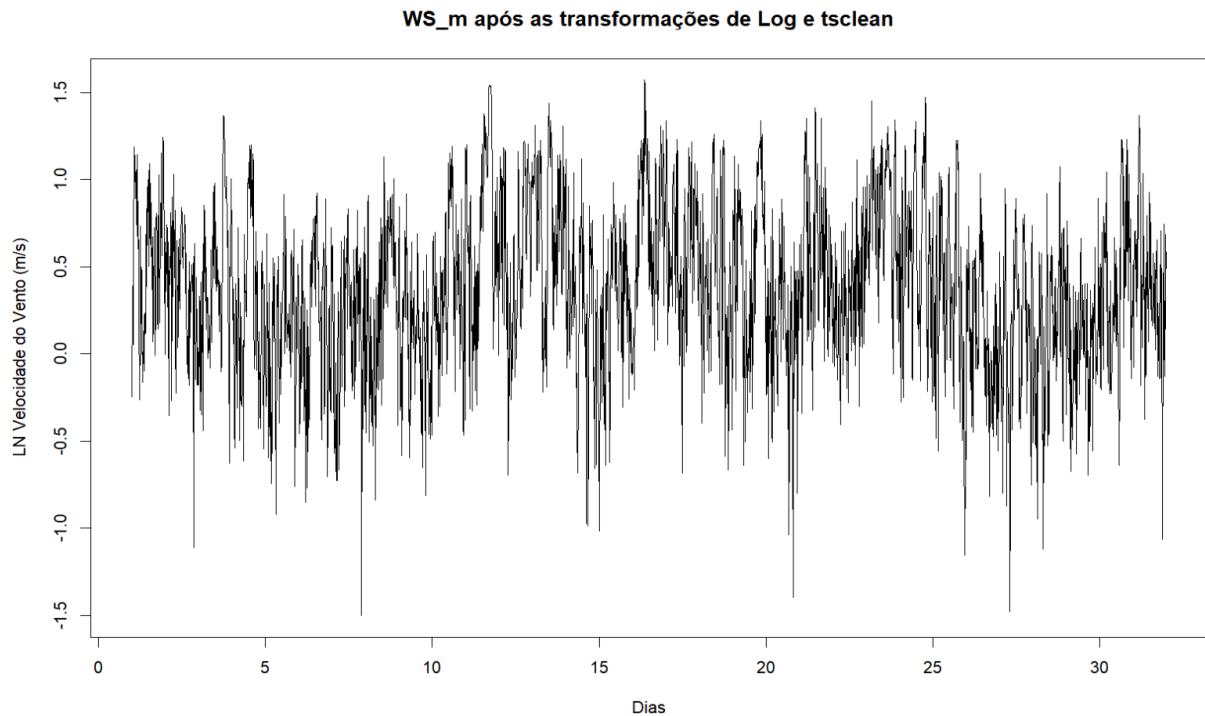


Figura 63 – WS_m após as transformações de Log e *tsclean*.

2 – Determinação das ordens do modelo SARIMA

O ajuste do modelo SARIMA acontecerá sobre o conjunto de treinamento, tendo como alvo da previsão os oito primeiros valores a partir da última observação desse conjunto ajustado, vide seção 4.2. Cabe observar que a diferenciação de primeira ordem realizada na seção 4.4.1, foi somente para a verificação da possibilidade de tornar a série WS_m estacionária.

A fim de orientar a escolha das ordens do modelo SARIMA, será realizada uma diferenciação de segunda ordem seguida da análise de sua normalidade e de seus correlogramas. A primeira diferenciação tem o intuito de remover a sazonalidade e a segunda de remover a tendência determinística. Assim, a primeira diferenciação será de 48 e a segunda diferenciação de 1. Os próximos quatro gráficos abaixo oferecem uma orientação para o estabelecimento dos parâmetros do modelo SARIMA a ser aplicado sobre o conjunto de treinamento.

O primeiro gráfico, Figura 64, apresenta a série WS_m após as transformações logarítmica e *tsclean* e após a diferenciação de segunda ordem.

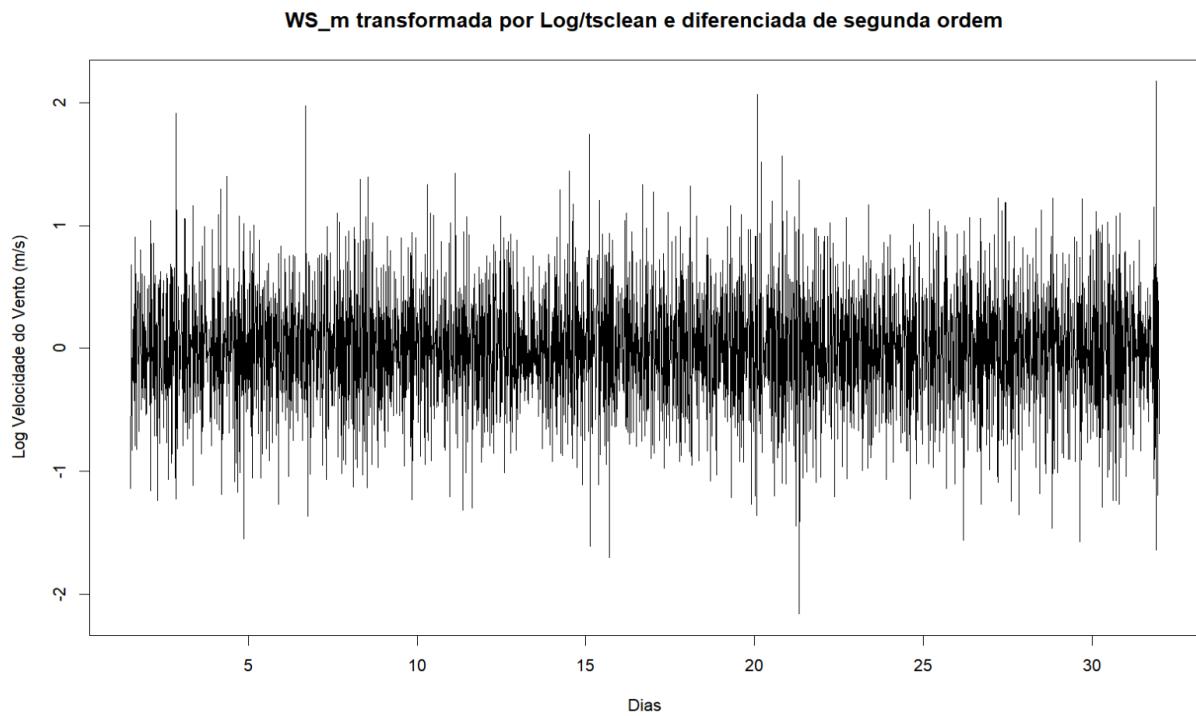


Figura 64 – WS_m após as transformações de Log e *tsclean* e diferenciação de segunda ordem.

O gráfico da Figura 65 se refere à normalidade da série resultante das transformações acima citadas. O gráfico *Q-Q plot* apresenta uma indicação de possível normalidade da série, visto que, excetuando-se os pontos iniciais e finais, os demais se encontram sobre a reta.

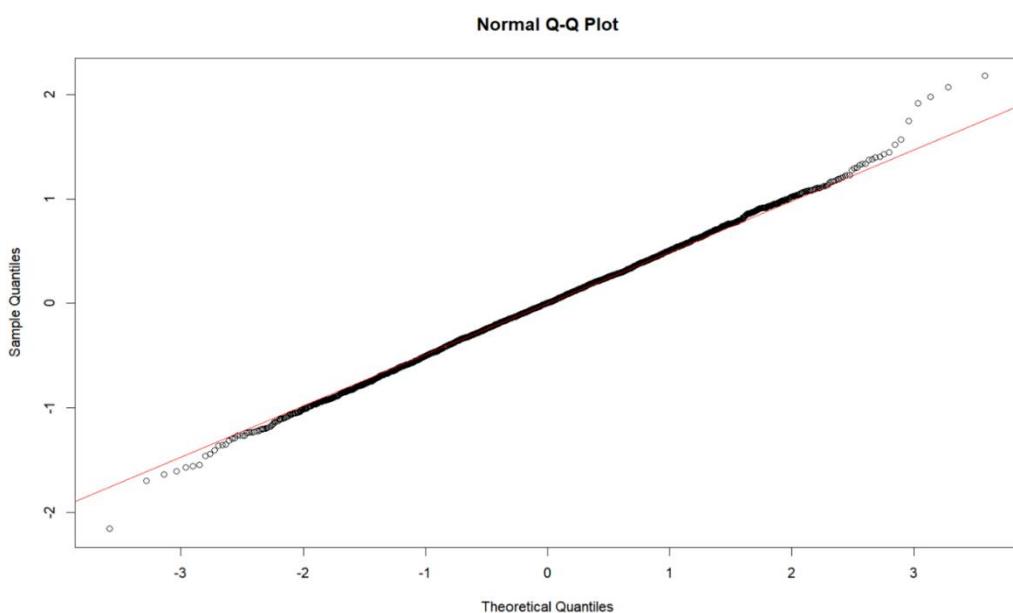


Figura 65 – Gráfico *QQ-plot* de WS_m após as transformações de Log, *tsclean* e diferenciações de segunda ordem.

Conforme comentado anteriormente, além do gráfico *QQ-plot*, a certificação de normalidade deve contar com o apoio de um histograma e de um teste de normalidade. A Figura 66 apresenta o histograma e a curva densidade de probabilidade correspondente.

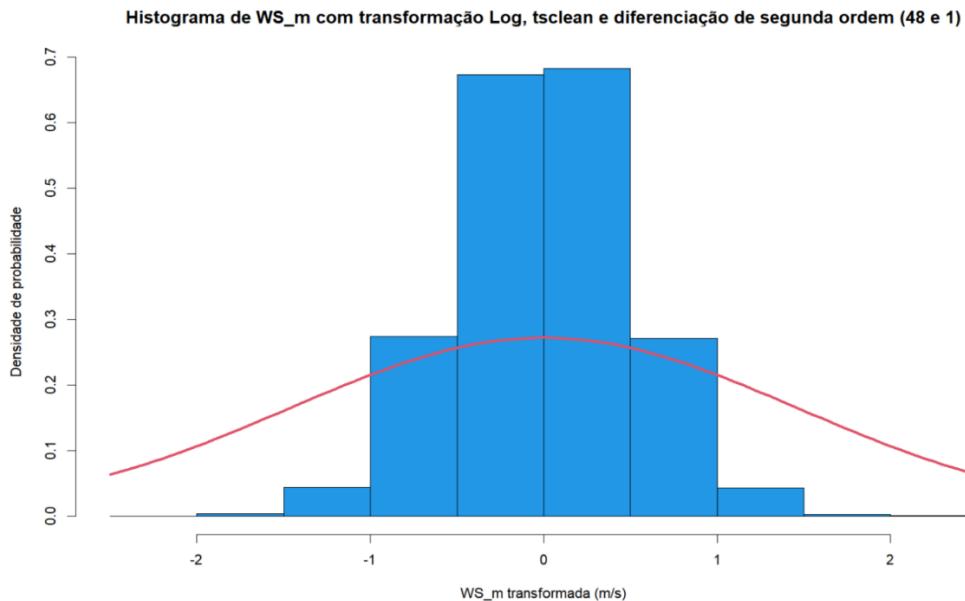


Figura 66 – Histograma e curva densidade de probabilidade de WS_m após as transformações e diferenciação de segunda ordem.

Finalmente, para comprovação da análise de normalidade, foi executado o teste de hipóteses de normalidade de Shapiro-Wilker, cujos resultados são apresentados na Tabela 46. Como o valor do teste (0.0547) é maior que o nível de significância (0.05), então não há motivos para se rejeitar a hipótese nula do teste, indicando assim que a série pode ser considerada normal.

Tabela 46 – Valor do teste de hipóteses de normalidade Shapiro-Wilker para WS_m após as transformações e diferenciação de segunda ordem.

Valor do teste Shapiro-Wilker	Nível de Significância
0.0547	0.05

Os dois gráficos da Figura 67 se referem aos correlogramas da série após as transformações (log e *tsclean*) e diferenciação de segunda ordem (diferenciação de 48 e a seguir diferenciação de 1). Para a exibição desses correlogramas foi estabelecido um *lag* igual a 192,

de modo a cobrir quatro períodos sazonais ($4*48$). A opção por um valor de *lag* maior que o máximo recomendado pela literatura foi para melhor visualização das partes sazonal e não sazonal dos correlogramas. Como também pode ser observado, os *lags* são apresentados em escala, onde o valor 0.5 da abscissa corresponde ao *lag* 48, o valor 1 corresponde ao *lag* 96 e assim por diante. Esses valores 48, 96, 144 etc. correspondem aos períodos sazonais. A partir deste ponto, as referências serão os números de *lags* propriamente ditos e não o valor em escala mostrado no eixo das abscissas.

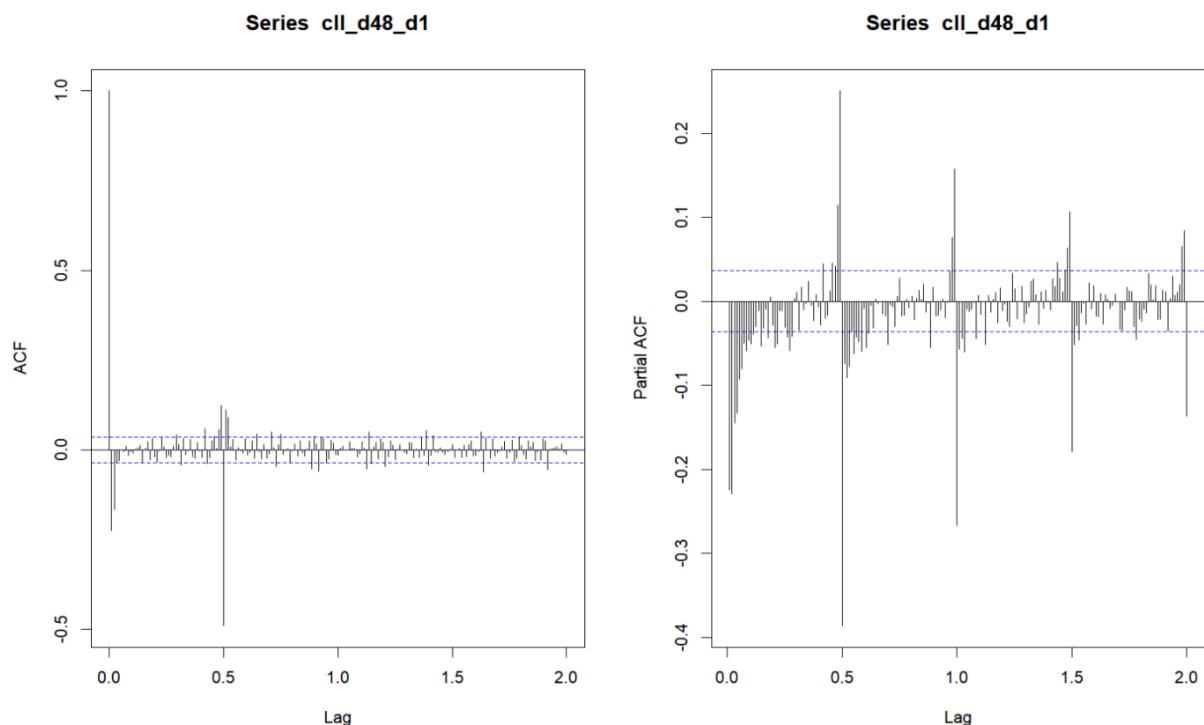


Figura 67 – FAC e FACP de WS_m após diferenciação de segunda ordem.

Em ambos os gráficos, FAC e FACP, o início (*lag* 1 ao *lag* 47) se refere à parte não sazonal e os *lags* seguintes se referem à parte relativa ao efeito sazonal.

- Análise da parte *Não Sazonal* dos correlogramas

No trecho inicial da FAC, observa-se um corte claro no *lag* 1 e no trecho inicial da FACP observa-se um decaimento exponencial. Essas observações são indicativas de um modelo MA de ordem 1 ($q=1$). Por outro lado, na FACP verifica-se um corte no *lag* 2, que apresenta um valor de autocorrelação maior que o *lag* anterior, fornecendo indicação de um modelo AR de ordem 2 ($p=2$). Diante de tais constatações, pode-se supor a indicação de um modelo ARMA(2,1).

- Análise da parte *Sazonal* dos correlogramas

Na FAC, pode-se considerar um corte claro no primeiro *lag* nessa região, que sugere ordem 1 para o modelo MA ($Q=1$). Na FACP não se observa cortes claros nesta região, que sugere ordem 0 para o modelo AR ($P=0$).

As avaliações dos trechos *não sazonal* e *sazonal* acima sugerem um modelo SARIMA de ordem $(2,1,1)(0,0,1)$ [48]. Lembrando que o primeiro algarismo 1 (posição d da parte não sazonal) indica que a série deve sofrer uma diferenciação de 1 para torná-la estacionária, como visto na seção 4.4.1. O modelo SARIMA realiza tal operação automaticamente ao se deparar com o algarismo 1 colocado nessa posição (d).

Conforme mencionado anteriormente, o resultado obtido através dessa análise aponta para um ponto de partida, pois normalmente não é muito evidente avaliar-se os processos que estão atuando nas escalas de tendência e sazonalidade somente observando a FAC e FACP. Então, deve-se analisar outras combinações para as ordens de AR e MA, tomando-se por referência a configuração que apresentar o menor valor de AICc. Para cada configuração testada também deve-se verificar se suas raízes se encontram no interior do círculo unitário. Assim, após diversos testes de configurações, chegou-se ao melhor modelo SARIMA com a seguinte estrutura: *SARIMA*(3,1,2)(0,0,1)[48]. Cabe lembrar que o ajuste SARIMA é realizado sobre o conjunto de treinamento.

A Figura 68 apresenta a distribuição das raízes no círculo unitário para esse modelo escolhido, onde todas as raízes se encontram no interior do círculo, como desejado.

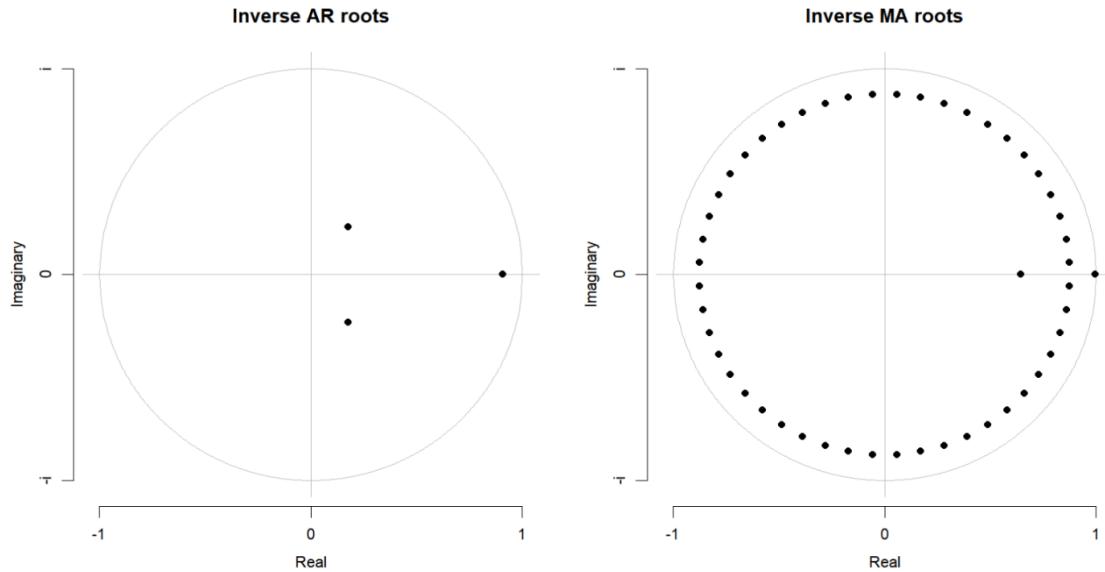


Figura 68 – Todas as raízes unitárias estão no interior do círculo.

Uma vez definidas as ordens do modelo SARIMA, passa-se então à análise dos resíduos produzidos pelo modelo $SARIMA(3,1,2)(0,0,1)[48]$.

- Distribuição dos resíduos.

A Figura 69 apresenta o gráfico relativo à distribuição dos resíduos. Visualmente, a média parece ser zero, e isso foi confirmado através de seu cálculo resultando em 0.000920.

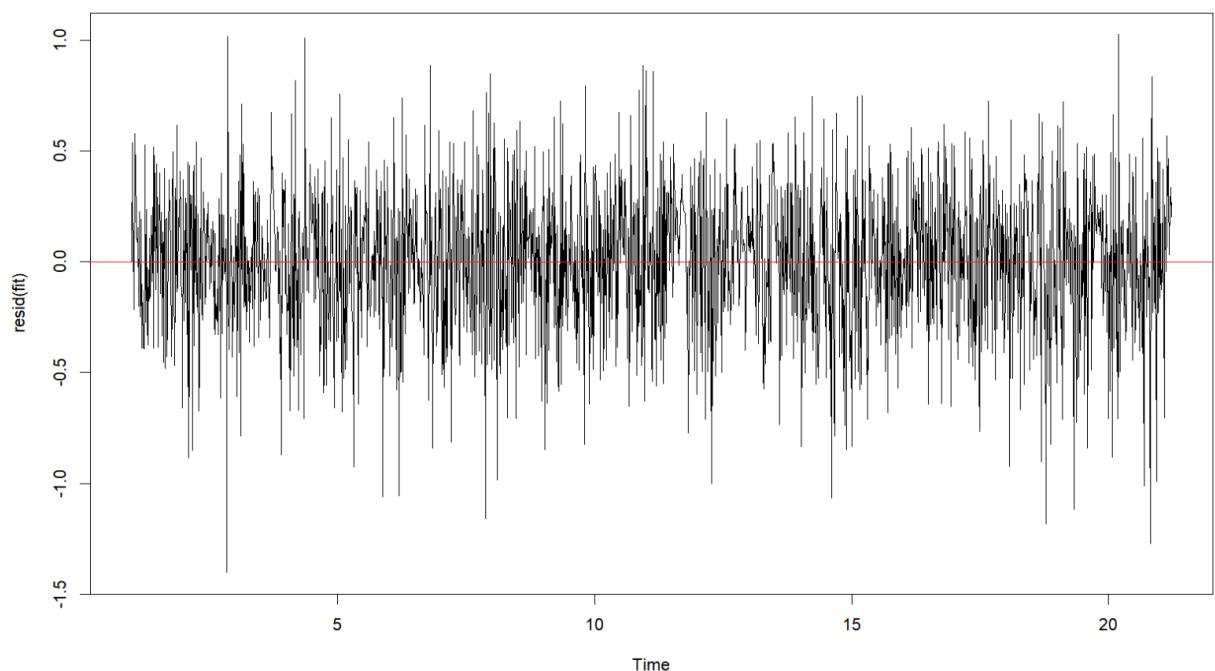


Figura 69 – Gráfico dos resíduos gerados pelo modelo SARIMA.

- Análise de normalidade dos resíduos do modelo.

A normalidade dos resíduos gerados pelo modelo $SARIMA(3,1,2)(0,01)[48]$ é analisada através do gráfico *QQ-plot* (Figura 70), histograma (Figura 71) e resultado do teste de hipóteses de Shapiro-Wilker (Tabela 47).

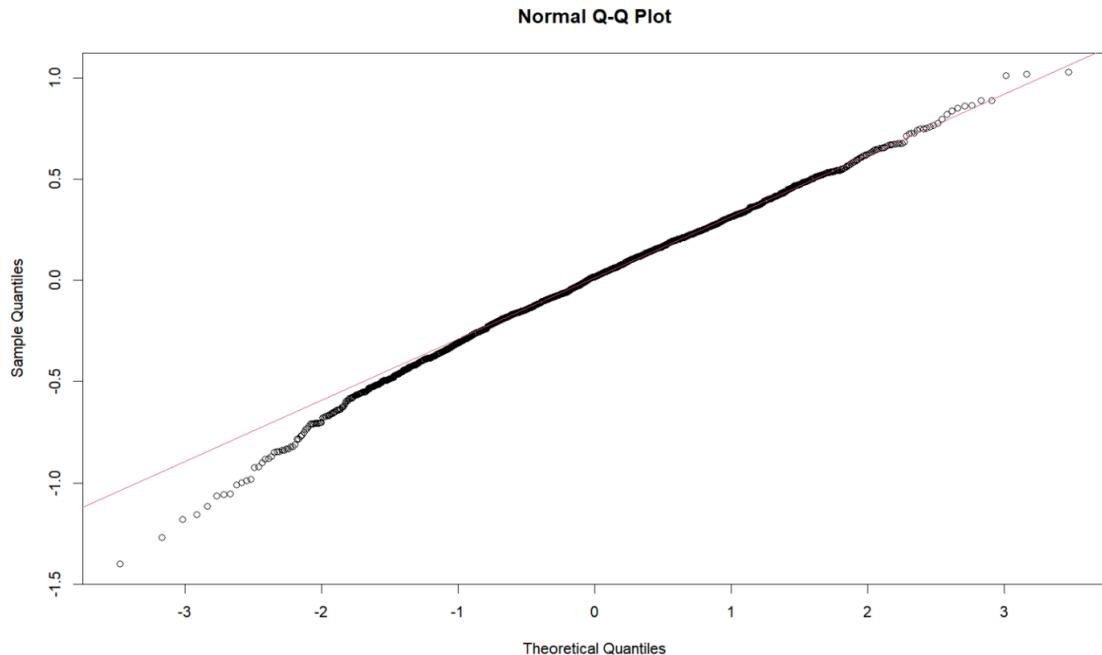


Figura 70 – Gráfico *QQ-plot* dos resíduos gerados pelo modelo SARIMA.

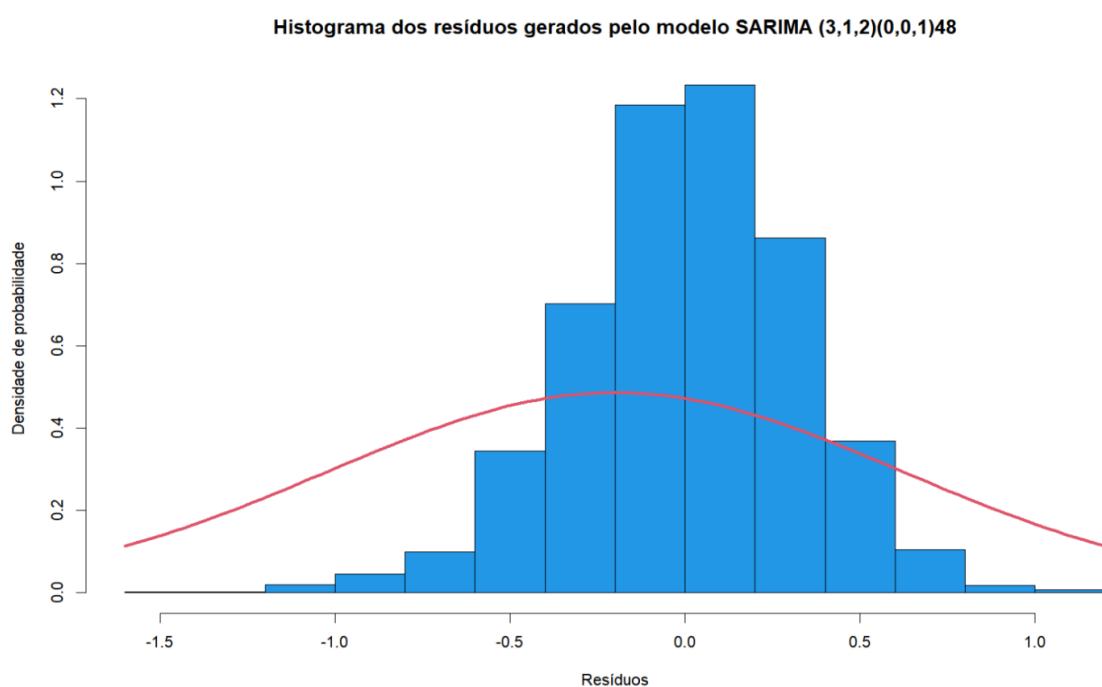


Figura 71 – Histograma e curva densidade de probabilidade dos resíduos gerados pelo modelo

SARIMA.

Tabela 47 – Teste de normalidade Shapiro-Wilker para os resíduos gerados pelo modelo SARIMA.

Valor do teste Shapiro-Wilker	Nível de Significância
8.008e-07	0.05

Através do gráfico *QQ-plot* dos resíduos, observa-se muitos pontos não pertencentes à reta, apesar do seu histograma sugerir uma forma normal. A certeza da não normalidade é definida pelo teste de hipótese de normalidade do teste de Shapiro-Winker acima, cujo valor é muito menor que 0.05. Assim, estas análises indicam a não normalidade dos resíduos, contrariando o desejável.

- Correlograma FAC dos resíduos do modelo, Figura 72.

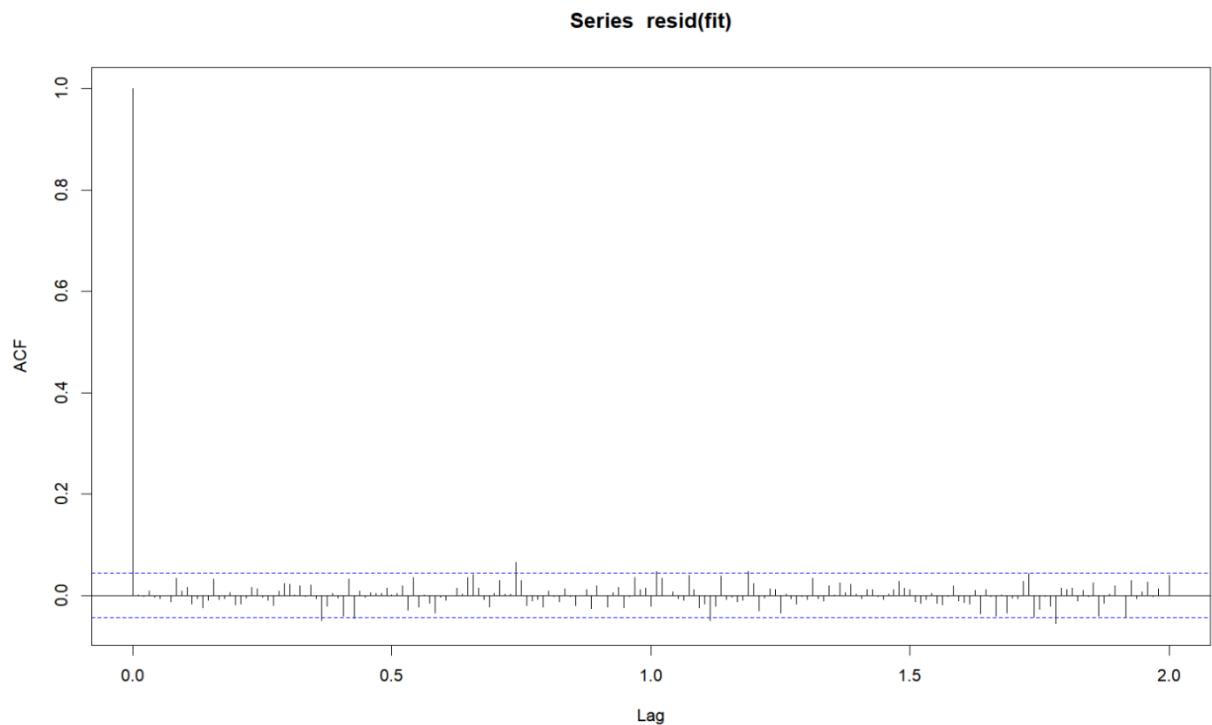


Figura 72 – Correlograma FAC dos resíduos gerados pelo modelo SARIMA.

Observa-se no correograma FAC dos resíduos gerados pelo modelo SARIMA acima, que 7 *lags* ultrapassam o intervalo de confiança e isso indica existência de correlação, fato não desejado. Contudo, conforme anteriormente citado, Hyndman *et al.* (2018) afirmam que quando

de menos de 5% dos *lags* estão fora do intervalo de confiança esse resultado pode ser considerado aceitável.

A FAC apresentou esses *lags* com autocorrelação muito provavelmente devido à dificuldade que os modelos autorregressivos enfrentam ao lidar com altas frequências de aquisição. De modo geral, esses modelos trabalham bem com frequência máxima de 24 e aqui trabalhou-se com quatro vezes esse valor.

- Aplicando o modelo SARIMA sobre o conjunto de treinamento

A Figura 73 mostra o conjunto de treinamento (preto) e o ajuste do modelo $SARIMA(3,1,2)(0,0,1)[48]$ realizado sobre esse conjunto (vermelho).

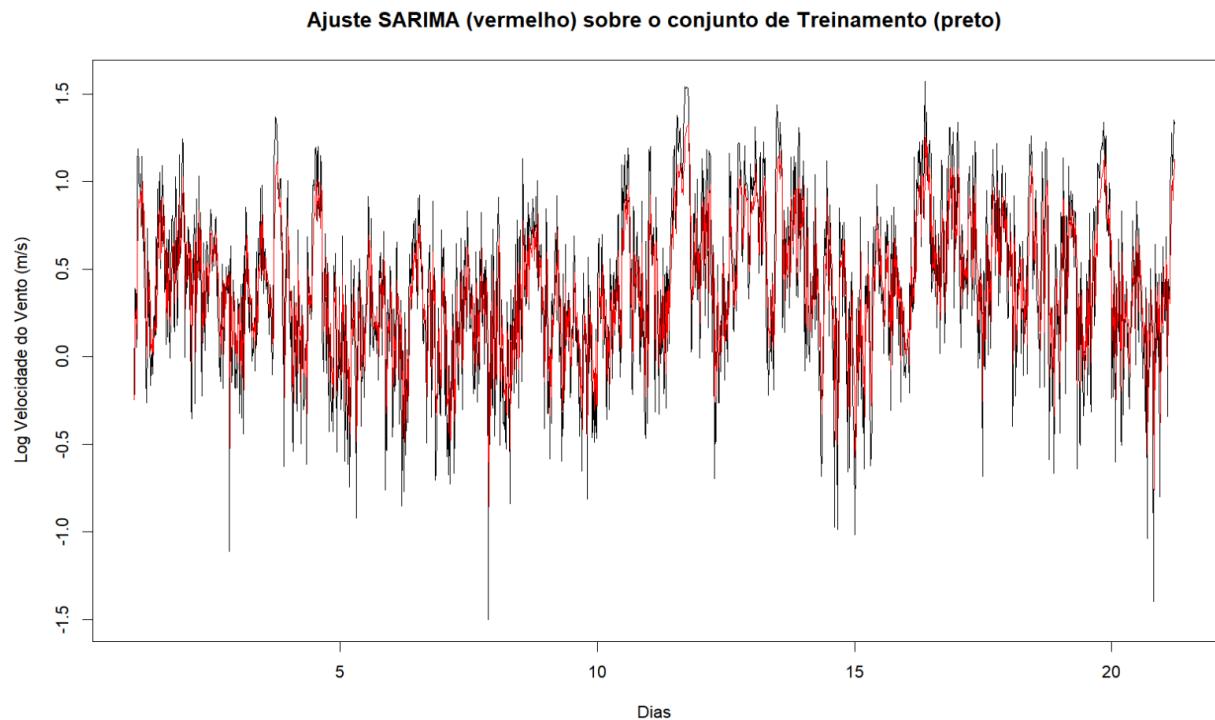


Figura 73 – Ajuste SARIMA (vermelho) sobre o conjunto de treinamento (preto).

- Aplicando o modelo SARIMA sobre o conjunto de teste

A Figura 74 mostra o conjunto de teste (preto) e o ajuste do modelo $SARIMA(3,1,2)(0,0,1)[48]$ realizado sobre esse conjunto (azul).

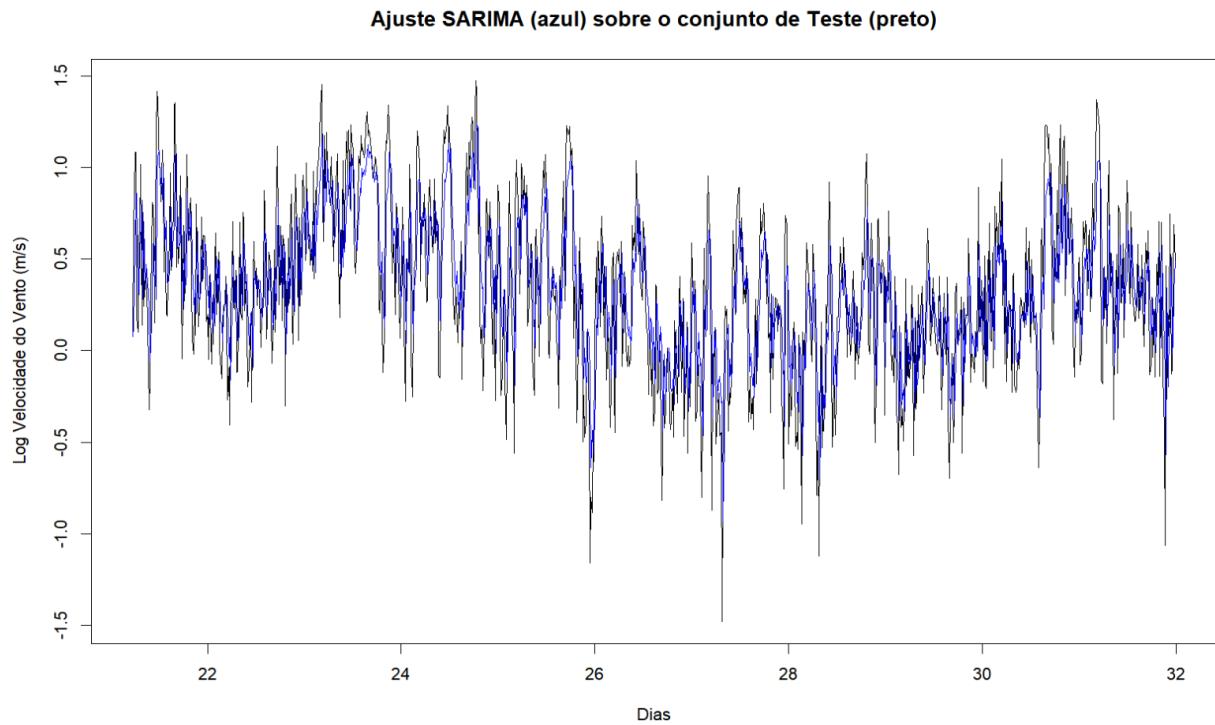


Figura 74 – Ajuste SARIMA (azul) sobre o conjunto de teste (preto).

Considerando as duas figuras acima, aparentemente o ajuste fornecido pelo modelo *SARIMA* $(3,2,1)(0,0,1)[48]$ parece satisfatório. Os picos acompanham bem, apesar de alguma variação de amplitude, mas no geral pode-se considerar aceitável. Sendo assim, passa-se agora ao processo de previsão, que no R pode ser realizado através do comando *predict* sobre o modelo ajustado. A previsão será realizada para 1034 observações à frente, a partir da última observação do conjunto de treinamento ajustado, embora o objetivo aqui seja prever somente os próximos 8 valores.

O gráfico da Figura 75 esboça na cor preta todo o conjunto de treinamento. A previsão para dos 1034 valores à frente é representada pela linha em azul escuro. O vão colorido exibe dois intervalo de confiança de previsão, o trecho mais escuro indicando 80% e o mais claro 95%.

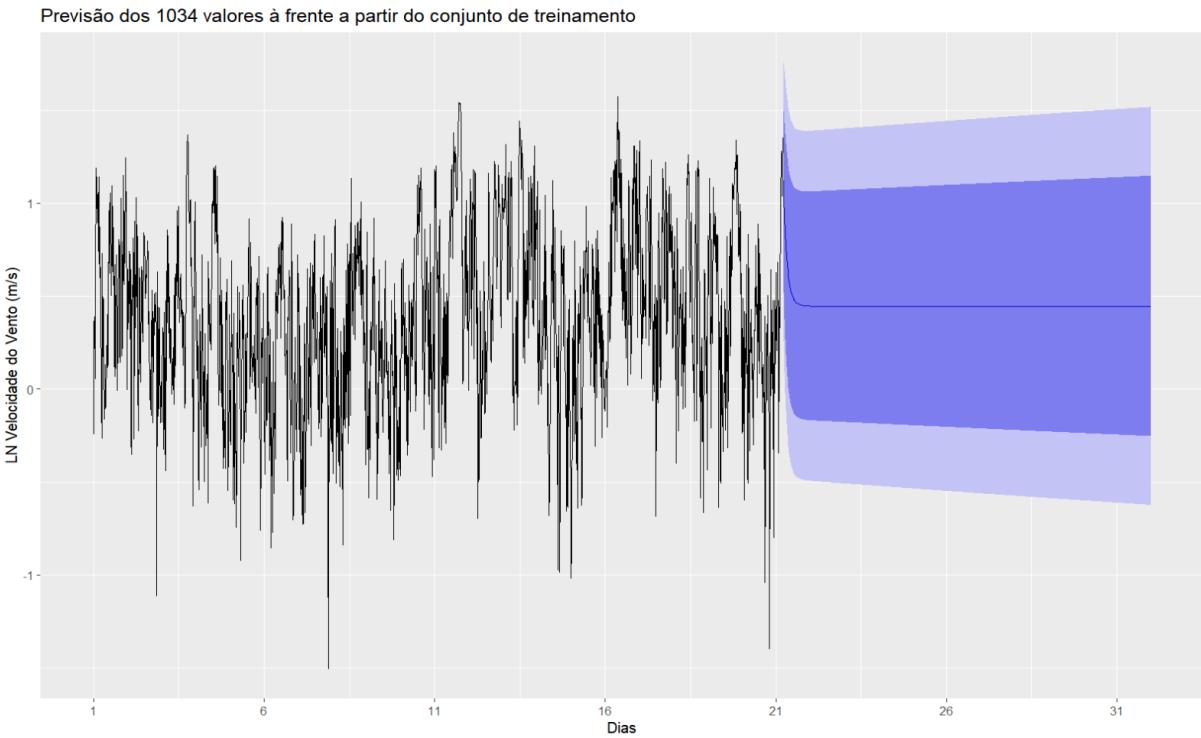


Figura 75 - Previsão SARIMA para todo o conjunto de teste.

- Acurácia da previsão para os oito primeiros valores do conjunto de teste - Tabela 48.

Tabela 48 – Medidas de acurácia da previsão SARIMA para os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão SARIMA	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,122090738	-0,04339074	0,04339	0,00188276	-4,02250279	4,0225028	0,98272178	1,46734065	44,0489668	45,0545925
1,9438	1,008348994	0,935451006	0,93545	0,87506859	48,1248588	48,124859				
2,9541	0,949455295	2,004644705	2,00464	4,01860039	67,8597442	67,859744				
2,9527	0,905058969	2,047641031	2,04764	4,19283379	69,3480893	69,348089				
1,1976	0,863215235	0,334384765	0,33438	0,11181317	27,9212396	27,92124				
1,1067	0,824671436	0,282028564	0,28203	0,07954011	25,4837412	25,483741				
1,8351	0,78951164	1,04558836	1,04559	1,09325502	56,9771871	56,977187				
1,9253	0,756654896	1,168645104	1,16865	1,36573138	60,6993769	60,699377				

- Acurácia da previsão para os todos os valores (1034) do conjunto de teste – Tabela 49

Tabela 49 – Medidas de acurácia da previsão SARIMA para todo o conjunto de teste.

AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
Valores Reais	Previsão SARIMA	Erro	Erro	Erro ²	Erro%	Erro%	MAE	MSE	MPE	MAPE
1,0787	1,122090738	-0,04339074	0,04339074	0,00188276	-4,02250279	4,02250279	1,10989527	1,76947199	64,097748	64,560051
1,9438	1,008348994	0,935451006	0,93545101	0,87506859	48,12485885	48,1248588				
2,9541	0,949455295	2,004644705	2,0046447	4,01860039	67,85974425	67,8597442				
2,9527	0,905058969	2,047641031	2,04764103	4,19283379	69,34808926	69,3480893				
1,1976	0,863215235	0,334384765	0,33438476	0,11181317	27,92123956	27,9212396				
1,1067	0,824671436	0,282028564	0,28202856	0,07954011	25,48374122	25,4837412				
1,8351	0,78951164	1,04558836	1,04558836	1,09325502	56,97718705	56,9771871				
1,9253	0,756654896	1,168645104	1,1686451	1,36573138	60,69937692	60,6993769				
2,7568	0,725881152	2,030918848	2,03091885	4,12463137	73,66943005	73,66943				
1,1488	0,701987039	0,446812961	0,44681296	0,19964182	38,89388591	38,8938859				
2,1292	0,677929597	1,451270403	1,4512704	2,10618578	68,16036086	68,1603609				
1,2756	0,657972648	0,617627352	0,61762735	0,38146355	48,41857571	48,4185757				
1,5334	0,6377803	0,8956197	0,8956197	0,80213465	58,40744101	58,407441				
1,4949	0,619379531	0,875520469	0,87552047	0,76653609	58,56715964	58,5671596				
1,3953	0,603688822	0,791611178	0,79161118	0,62664826	56,73412013	56,7341201				
1,1473	0,589473385	0,557826615	0,55782662	0,31117053	48,62081542	48,6208154				
0,7244	0,575779174	0,148620826	0,14862082	0,02208815	20,51640336	20,5164034				
0,8622	0,563394978	0,298805022	0,29880502	0,08928444	34,65611478	34,6561148				
1,3472	0,553342519	0,793857481	0,79385748	0,6302097	58,9264757	58,9264757				
2,2483	0,542852444	1,705447556	1,70544756	2,90855137	75,85498181	75,8549818				
2,1999	0,532131626	1,667768374	1,66776837	2,78145135	75,81109933	75,8110993				
1,9508	0,525682218	1,425117782	1,42511778	2,03090669	73,05299272	73,0529927				
1,166	0,517999274	0,648000726	0,64800073	0,41990494	55,57467631	55,5746763				
3,3066	0,512201346	2,794398654	2,79439865	7,80866384	84,50972763	84,5097276				
4,1093	0,505944682	3,603355318	3,60335532	12,9841696	87,68781346	87,6878135				
3,9172	0,49996992	3,41723008	3,41723008	11,6774614	87,23654856	87,2365486				
2,7303	0,495860443	2,234439557	2,23443956	4,99272014	81,83860958	81,8386096				
2,6955	0,490905902	2,204594098	2,2045941	4,86023514	81,78794652	81,7879465				
2,2498	0,486627322	1,763172678	1,76317268	3,10877789	78,37019638	78,3701964				
2,9872	0,482692654	2,504507346	2,50450735	6,27255705	83,84130109	83,8413011				
2,6069	0,479295472	2,127604528	2,12760453	4,52670103	81,61435145	81,6143514				
1,5367	0,477085233	1,059614767	1,05961477	1,12278345	68,95391206	68,9539121				
1,9246	0,474174256	1,450425744	1,45042574	2,10373484	75,36245165	75,3624516				
1,2238	0,471136258	0,752663742	0,75266374	0,56650271	61,5021852	61,5021852				
1,2072	0,467597514	0,739602486	0,73960249	0,54701184	61,2659448	61,2659448				
1,4227	0,466519575	0,057198405	0,05719840	0,01620585	67,22224107	67,222242				

4.5 – APLICAÇÃO DA DECOMPOSIÇÃO STL/SES/ARMA SOBRE A SÉRIE WS_m

A base do método STL é uma transformação não linear chamada Loess, que é um modelo de decomposição mais robusto, matematicamente falando. A decomposição STL é aplicada sobre o conjunto de treinamento gerando as componentes que o descrevem. O próximo passo consiste em determinar os modelos ou métodos que melhor se ajustem a cada componente STL gerada, de modo a viabilizar uma previsão para os próximos 1034 valores à frente. Lembrando que 1034 é o número de observações do conjunto de teste, para o qual se deseja uma realizar a previsão. Uma vez obtidos os modelos e feita a previsão sobre eles, as séries resultantes são somadas (equação 1), resultando na série temporal X_t que descreve a previsão. O motivo da adição das componentes é porque as séries tratadas nesse trabalho são do tipo aditivas.

$$X_t \text{ (série temporal)} = T_t \text{ (tendência)} + S_t \text{ (efeito sazonal)} + R_t \text{ (restante)} \quad (1)$$

Abaixo são descritos os procedimentos para obtenção de cada parcela da equação 1.

1 – Decomposição STL aplicada sobre o conjunto de treinamento.

A decomposição STL aplicada sobre o conjunto de treinamento gera três componentes (três séries temporais), a saber, a componente relativa à tendência (*trend*), sazonalidade (*seasonal*) e restante (*remainder*). Além dessas, o *software R* também representa graficamente a série alvo da decomposição (*data*), que neste caso se refere ao conjunto de treinamento, conforme mostrado na Figura 76.

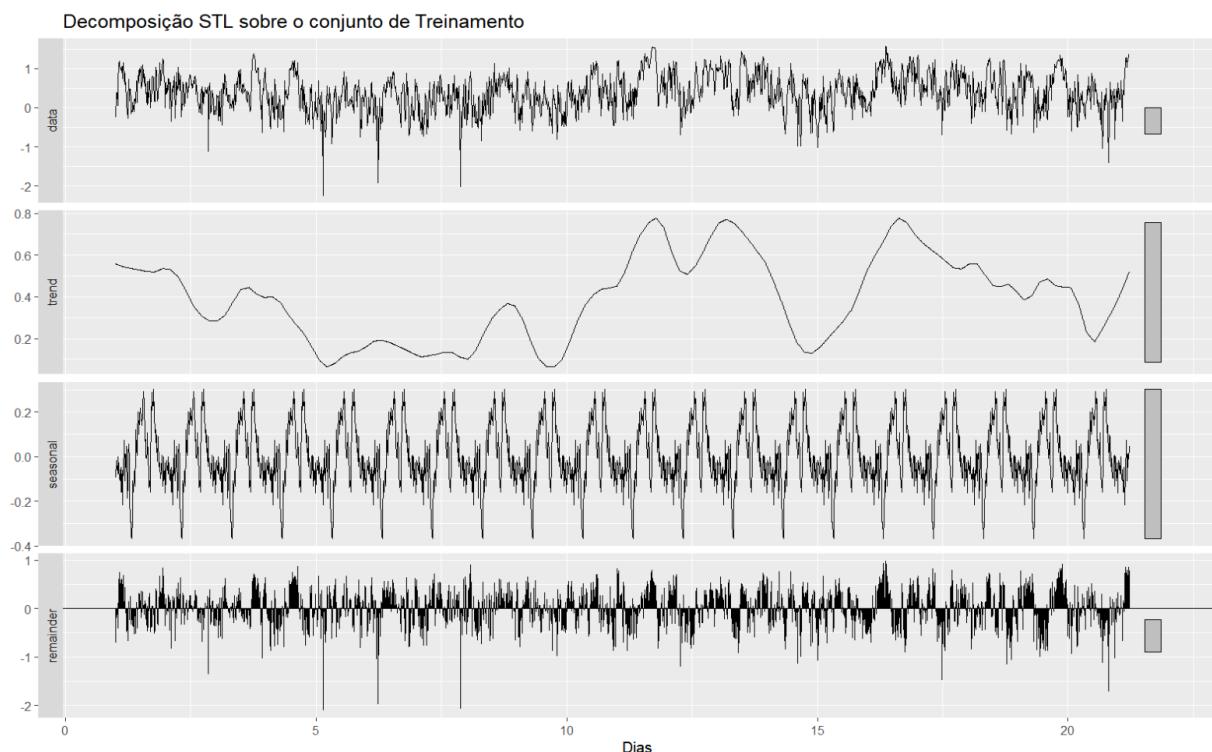


Figura 76 – Decomposição STL sobre o conjunto de treinamento.

Naturalmente, as séries resultantes da decomposição STL sobre o conjunto de treinamento possuem as mesmas características desse conjunto, ou seja, são compostas por 1942 observações e uma frequência de aquisição igual a 96.

2 – Estimativa do método para a componente tendência (*trend*).

A estimativa do método que melhor descreve a componente tendência (*trend*), obtida da decomposição STL sobre o conjunto de treinamento, pode englobar diversas abordagens como o modelo ingênuo, média móvel, regressões, métodos de suavização exponencial etc. A metodologia que se mostrou mais adequada para representar a componente tendência (*trend*) foi a suavização exponencial simples (SES) com valor de alfa = 0.17, gerando um valor pontual

igual a 0.4856 a ser replicado 1034 vezes e assim gerando a parcela T_t (tendência) da equação 1. Lembrando que tanto o modelo ingênuo, quanto o SES resultam somente um valor.

O gráfico abaixo apresenta em preto a componente *trend* gerada pela decomposição STL (preto) e o valor obtido pelo método SES (vermelho) replicado para T_t (tendência).

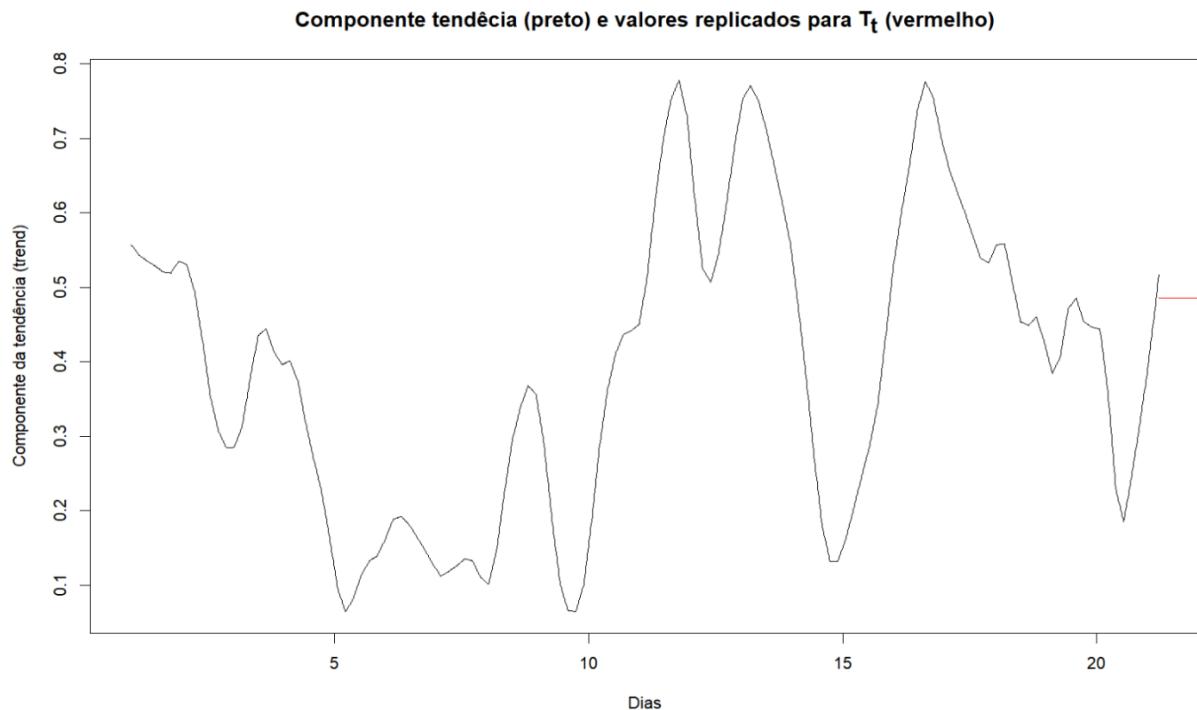


Figura 77 – Componente *trend* (preto) e valores replicados para T_t (tendência) (vermelho).

3 – Estimativa do método para a componente sazonalidade (*seasonal*).

Como a série WS_m apresenta somente um tipo de sazonalidade, que se repete periodicamente a cada dia (vide seção 4.2), então foram selecionados os últimos 1034 valores da componente sazonal (*seasonal*) para serem replicados e assim gerar a componente sazonal S_t (efeito sazonal).

A Figura 78 apresenta os últimos 1034 valores da componente sazonal (*seasonal*) do conjunto de treinamento (preto) e a replicação desses para a componente S_t (efeito sazonal) (vermelho).

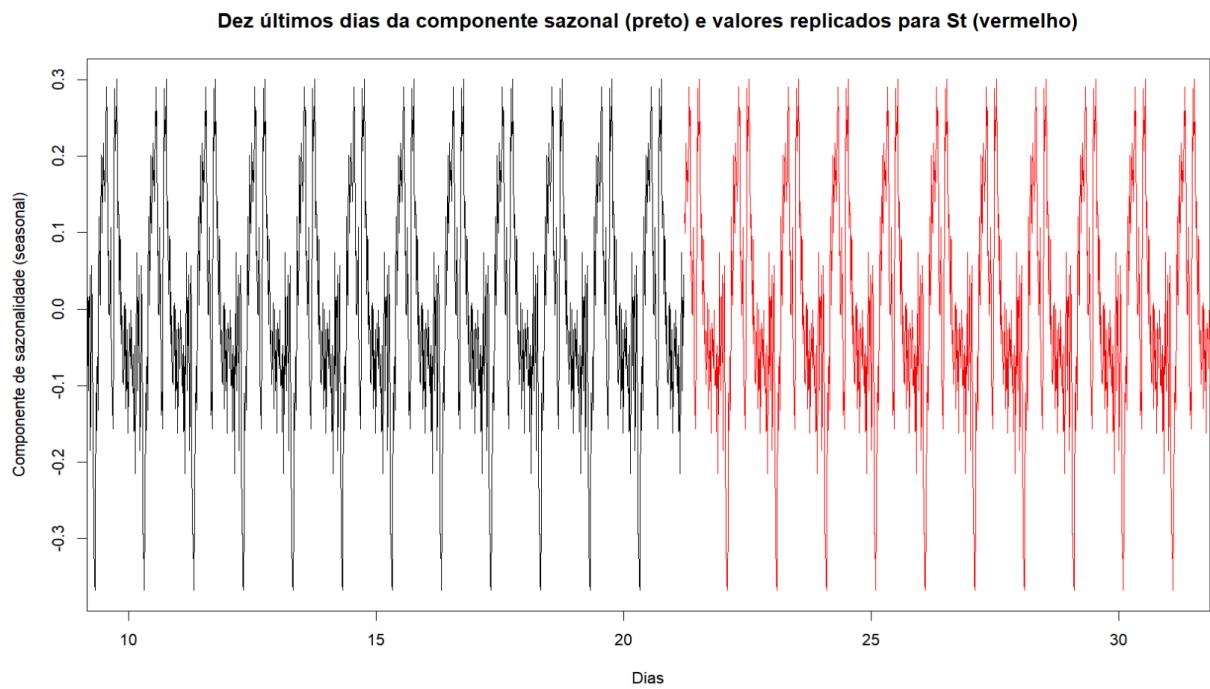


Figura 78 - Dez últimos dias de *seasonal* (preto) e valores replicados para S_t (efeito sazonal) (vermelho).

4 – Estimativa do método para a componente restante (*remainder*).

Os gráficos abaixo apresentam algumas características da componente (*remainder*) obtida sobre o conjunto de treinamento. A Figura 79 mostra a distribuição dessa componente.

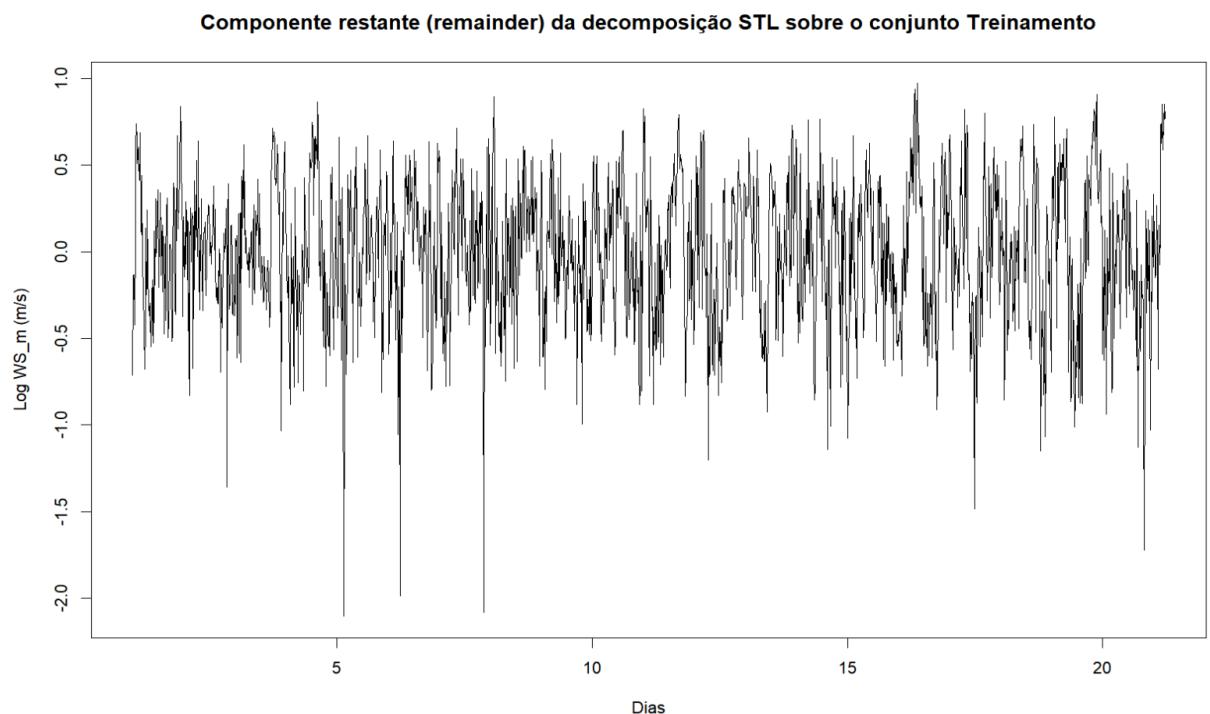


Figura 79 – Componente restante (*remainder*) obtida pela decomposição STL.

Os correlogramas FAC e FACP, apresentados na Figura 80 abaixo, sinalizam que muito provavelmente o processo que descreve a componente restante (*remainder*) seja um modelo ARMA(p,q). Foram realizados diversos testes envolvendo as ordens p e q na busca do modelo que apresentasse o menor valor de AICc e o modelo selecionado foi o ARMA(6,11).

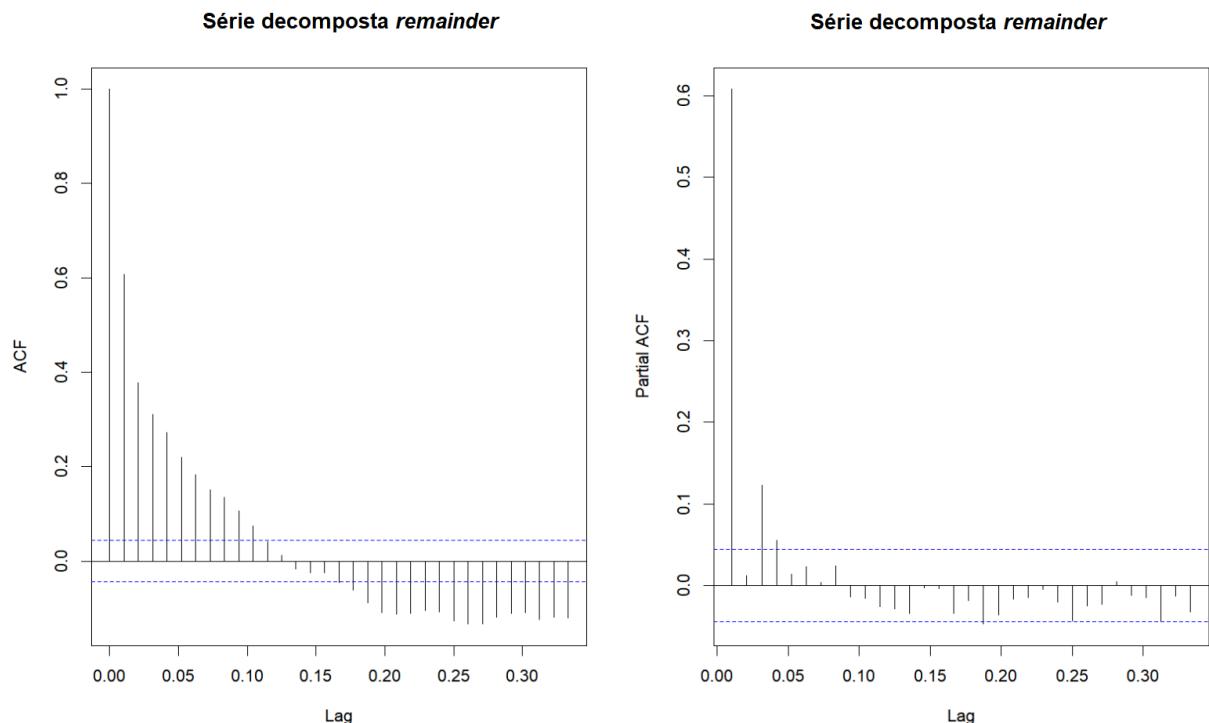


Figura 80 – Correlogramas da componente de restante (*remainder*) obtida pela decomposição STL.

Na Figura 81 observa-se que nenhuma raiz unitária desse modelo se encontra fora do círculo unitário, como desejado.

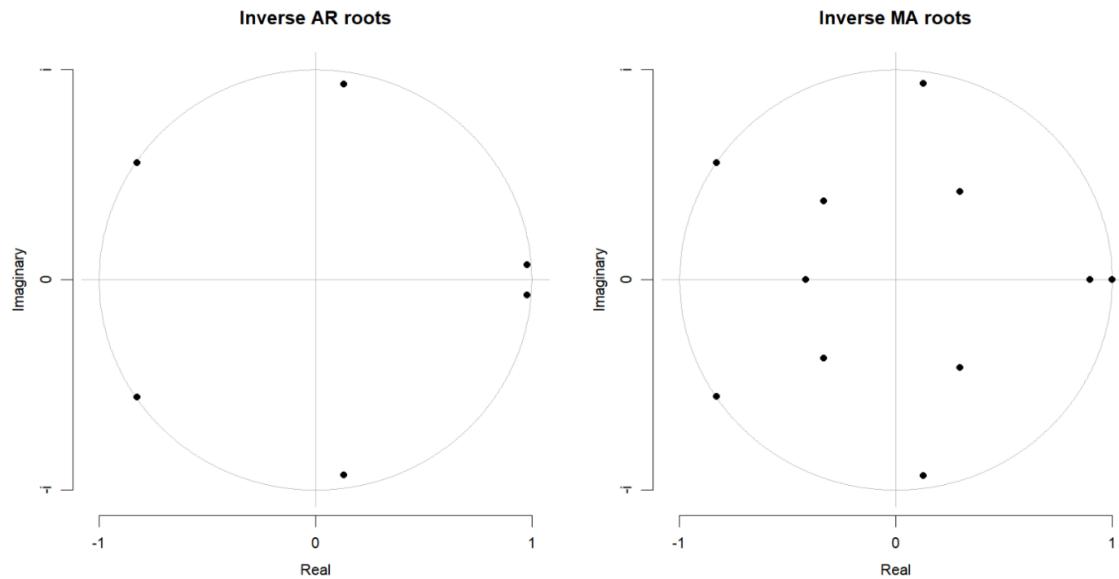


Figura 81 – Raízes unitárias do modelo ARMA(6,11).

O próximo passo consiste em analisar o correlograma FAC dos resíduos gerados pelo ajuste do modelo ARMA(6,11) sobre a componente restante (*remainder*). A Figura 82 exibe tal correlograma.

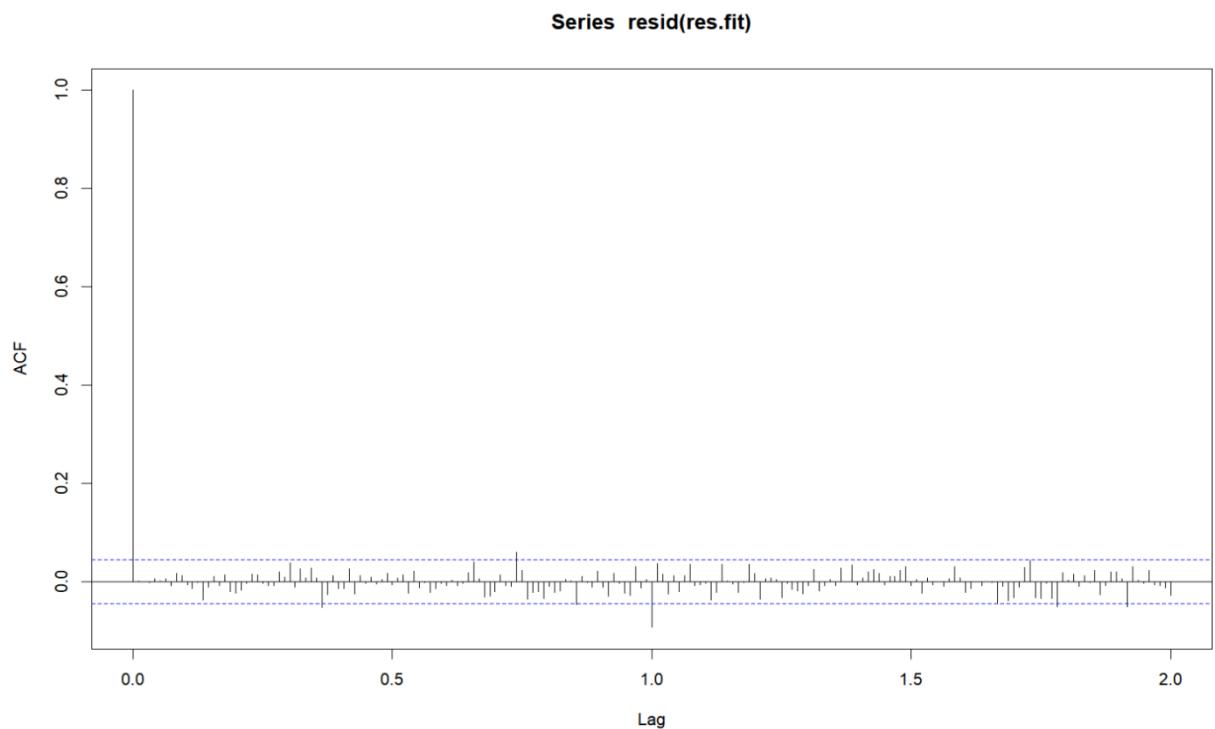


Figura 82 – FAC dos resíduos obtidos pelo ajuste do modelo ARMA(6,11).

Examinando a FAC, observa-se 5 *lags* ultrapassando os limites de confiança, indicando assim uma correlação. Conforme relatado acima, um percentual abaixo de 5% é considerado aceitável. Então, de posse de um ajuste aceitável sobre a componente restante (*remainder*), passa-se a efetuar a previsão para os próximos 1034 valores à frente sobre esse modelo ARMA, de modo a obter-se a parcela R_t (restante).

4 – Obtenção da série X_t da equação 1.

Ao somar-se as parcelas T_t (tendência) + S_t (efeito sazonal) + R_t (restante) obtidas acima, tem-se a série X_t , que contém a previsão realizada pelo método STL/SES/ARMA para o conjunto de teste.

O gráfico da Figura 83 esboça os dez últimos dias do conjunto de treinamento (vermelho) e a previsão STL/SES/ARMA obtida (azul) sobre a totalidade do conjunto de teste (verde). Observa-se que a previsão STL/SES/ARMA acompanhou relativamente bem a variabilidade presente no conjunto de teste e, como esperado, os primeiros valores de previsão são mais precisos, pois o aumento da janela preditiva implica em perda de precisão.

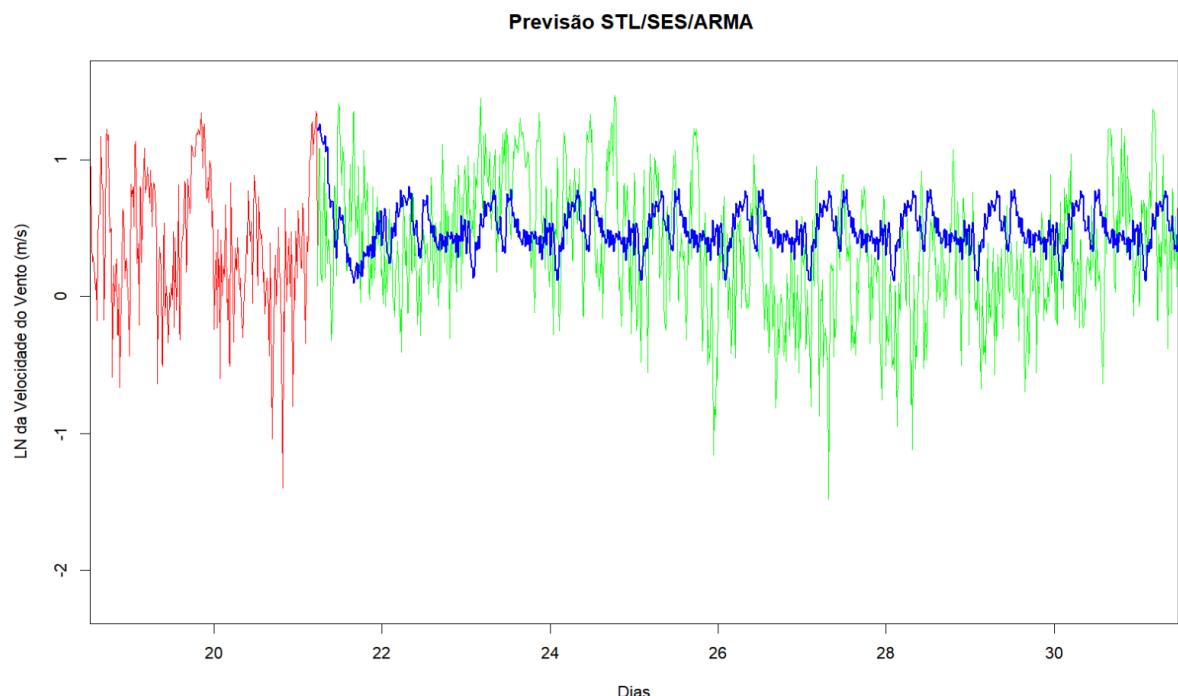


Figura 83 – Parte do conjunto de treinamento (vermelho), todo o conjunto de teste (verde) e a previsão STL/SES/ARMA sobre o conjunto de teste (azul).

A fim de facilitar uma melhor avaliação visual entre as previsões do SCA e do modelo STL/SES/ARMA, a Figura 84 esboça graficamente o último dia do conjunto de treinamento (vermelho), os oito primeiros valores do conjunto de teste (verde), os oito valores previstos pelo SCA (preto) e os oito primeiros valores previstos pelo STL/SES/ARMA (azul).

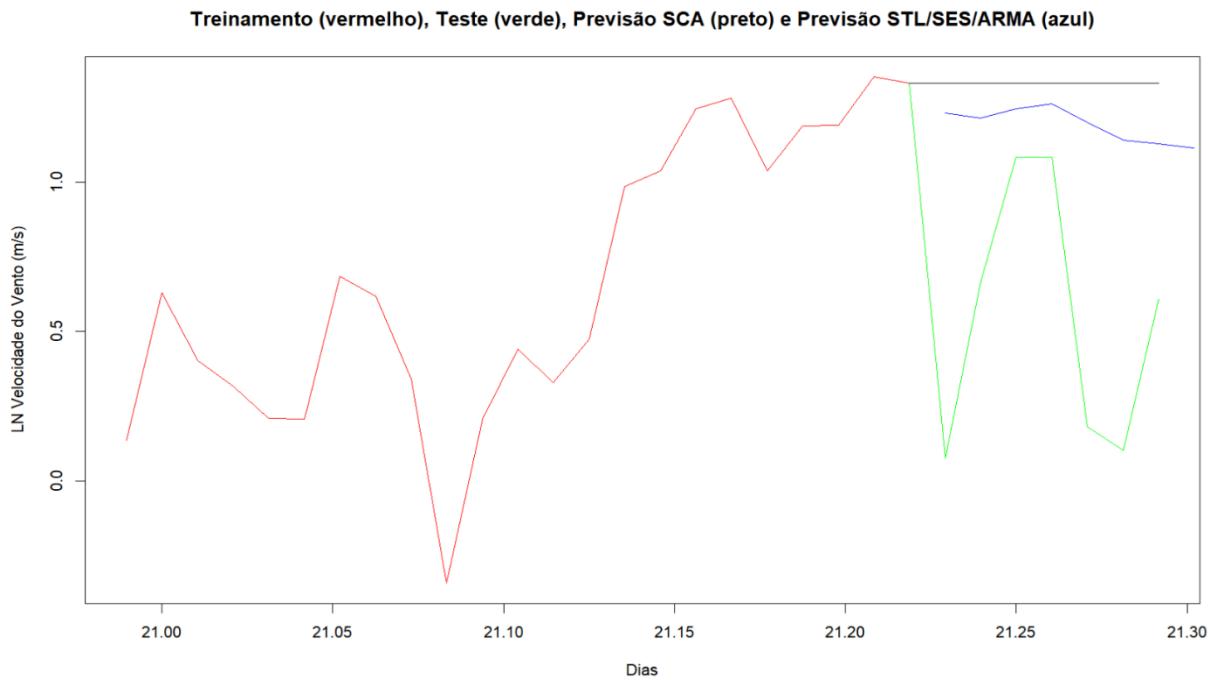


Figura 84 – Treinamento (vermelho), oito primeiros valores de teste (verde), previsão de oito valores do SCA (preto) e previsão de oito valores do STL/SES/ARMA (azul).

A Tabela 50 mostra as medidas de acurácia da previsão STL/SES/ARMA para os oito primeiros valores do conjunto de teste. Oito valores correspondem às duas primeiras horas que o SCA utiliza em suas previsões.

Tabela 50 – Medidas de acurácia da previsão STL/SES/ARMA sobre os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão STL/SES/ARMA	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,231408577	-0,15270858	0,15271	0,02331991	-14,1567236	14,156724	0,72999516	0,93847583	27,0349819	31,3477796
1,9438	1,213344877	0,730455123	0,73046	0,53356469	37,5787181	37,578718				
2,9541	1,243518416	1,710581584	1,71058	2,92608935	57,9053378	57,905338				
2,9527	1,26051735	1,69218265	1,69218	2,86348212	57,3096708	57,309671				
1,1976	1,199036264	-0,00143626	0,00144	2,0629E-06	-0,11992853	0,1199285				
1,1067	1,139619223	-0,03291922	0,03292	0,00108368	-2,97453902	2,974539				
1,8351	1,127960489	0,707139511	0,70714	0,50004629	38,5341132	38,534113				
1,9253	1,112761673	0,812538327	0,81254	0,66021853	42,2032061	42,203206				

A Tabela 51 exibe as medidas de acurácia da previsão STL/SES/ARMA para todo o conjunto de teste.

Tabela 51 – Medidas de acurácia da previsão STL/SES/ARMA sobre todo o conjunto de teste.

AH Valores Reais	AI Previsão STL/SES/ARMA	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
		Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
1,0787	1,231408577	-0,15270858	0,15270858	0,02331991	-14,1567236	14,1567236	1,07713296	1,71930284	61,150667	62,457247
1,9438	1,213344877	0,730455123	0,73045512	0,53356469	37,57871813	37,5787181				
2,9541	1,243518416	1,710581584	1,71058158	2,92608935	57,9053378	57,9053378				
2,9527	1,26051735	1,69218265	1,69218265	2,86348212	57,30967083	57,3096708				
1,1976	1,199036264	-0,00143626	0,00143626	2,0629E-06	-0,11992853	0,11992853				
1,1067	1,139619223	-0,03291922	0,03291922	0,00108368	-2,97453902	2,97453902				
1,8351	1,127960489	0,707139511	0,70713951	0,50004629	38,53411319	38,5341132				
1,9253	1,112761673	0,812538327	0,81253833	0,66021853	42,20320611	42,2032061				
2,7568	1,133166381	1,623633619	1,62363362	2,63618613	58,89558979	58,8955898				
1,1488	1,175720612	-0,02692061	0,02692061	0,00072472	-2,34336799	2,34336799				
2,1292	1,056324188	1,072875812	1,07287581	1,15106251	50,38868176	50,3886818				
1,2756	1,097246682	0,178353318	0,17835332	0,03180991	13,98191582	13,9819158				
1,5334	0,830888648	0,702511352	0,70251135	0,4935222	45,81396586	45,81396589				
1,4949	0,823488361	0,671411639	0,67141164	0,45079359	44,9134818	44,9134818				
1,3953	0,71432288	0,68097712	0,68097712	0,46372984	48,80506848	48,8050685				
1,1473	0,648913744	0,498386256	0,49838626	0,24838886	43,4399247	43,4399247				
0,7244	0,693537954	0,030862046	0,03086205	0,00095247	4,260359726	4,26035973				
0,8622	0,69761896	0,16458104	0,16458104	0,02708692	19,08849924	19,0884992				
1,3472	0,573353344	0,773846656	0,77384666	0,59883865	57,44111164	57,4411116				
2,2483	0,488943558	1,759356442	1,75935644	3,09533509	78,25274395	78,252744				
2,1999	0,361621485	1,838278515	1,83827852	3,3792679	83,56191259	83,5619126				
1,9508	0,405069437	1,545730563	1,54573056	2,38928297	79,23572704	79,235727				
1,166	0,278982415	0,887017585	0,88701759	0,7868002	76,07354933	76,0735493				
3,3066	0,412825077	2,893774923	2,89377492	8,3739333	87,51511894	87,5151189				
4,1093	0,570671823	3,538628177	3,53862818	12,5218894	86,11267556	86,1126756				
3,9172	0,656426672	3,260773328	3,26077333	10,6326427	83,24245196	83,242452				
2,7303	0,58873088	2,14156912	2,14156912	4,5863183	78,43713585	78,4371359				
2,6955	0,5575353819	2,137964181	2,13796418	4,57089084	79,31605199	79,316052				
2,2498	0,563077181	1,686722819	1,68672282	2,84503387	74,97212281	74,9721228				
2,9872	0,5991159313	2,388040687	2,38804069	5,70273832	79,942444	79,942444				
2,6069	0,464319399	2,142580601	2,1425806	4,59065163	82,18882969	82,1888297				
1,5367	0,395296741	1,141403259	1,14140326	1,3028014	74,27625814	74,2762581				
1,9246	0,373384817	1,551215183	1,55121518	2,40626855	80,59935485	80,5993548				
1,2238	0,371450856	0,852349144	0,85234914	0,72649906	69,64774836	69,6477484				
1,2072	0,249674105	0,957525895	0,9575259	0,91685584	79,31791709	79,3179171				
1,4227	0,216296604	1,107212306	1,10721234	1,22614206	77,7771570	77,7771570				

JARTOS | COMPARA | MAPEs | **Conjunto TEST** | MAPEs SES H, H-W | + | : |

5 – ANÁLISE DOS RESULTADOS

5.1 – ACURÁCIA DA PREVISÃO DO MODELO UTILIZADO PELO SCA

A cada nova execução, o SCA realiza uma previsão para uma e duas horas à frente utilizando o método de previsão conhecido por *naïve method* ou método ingênuo de previsão, o qual é largamente utilizado e se mostra bastante satisfatório nesse contexto. Dessa forma, a cada aquisição de dados atmosféricos o SCA realiza uma nova previsão de oito valores à frente, que corresponde a duas horas. O método ingênuo de previsão atende bem à medida que o próximo valor a ser coletado não difira substancialmente do último conhecido. Os dados avaliados nesse trabalho manipulam valores de velocidade do vento adquiridos pela torre C a cada quinze minutos referentes ao mês de maio de 2016, totalizando 2976 observações.

A fim de comparar a previsão realizada pelo SCA com alguns métodos aqui abordados, foi definido um conjunto de treinamento de modo que o seu último valor divirja consideravelmente do primeiro valor do conjunto de teste, pois esse fato caracteriza a vulnerabilidade do método ingênuo. Nesses termos, o conjunto de treinamento consiste em 1942 observações, correspondendo a aproximadamente 65.25% do total e o conjunto de teste corresponde a 1034 observações, equivalendo a aproximadamente 34.75%. O último valor do conjunto de treinamento indica uma velocidade do vento igual a 3.78 m/s e o primeiro valor do conjunto de teste aponta para uma velocidade de 1.07 m/s.

A avaliação de resultados consiste na comparação das medidas de acurácia obtidas entre a previsão realizada pelo SCA e a previsão obtida pelos métodos utilizados. No caso exemplo, série de maio de 2016, a previsão do SCA consiste na repetição do valor 3.78m/s oito vezes, correspondendo a duas horas. Ao comparar os valores previstos pelo SCA com os valores reais, que correspondem as oito primeiras observações do conjunto de teste, obtém-se os valores de acurácia da previsão do SCA, conforme mostrado na tabela abaixo.

Tabela 52 – Medidas de acurácia da previsão realizada pelo SCA para os 8 primeiros valores de teste.

Valor real da Velocidade do Vento	Previsão do SCA	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3.783										
1.0787	3.783	-2.7043	2.7043	7.313238	-250.7	250.6999	-1.90875	4.145576	-132.73	132.7303
1.9438	3.783	-1.8392	1.8392	3.382657	-94.6188	94.61879				
2.9541	3.783	-0.8289	0.8289	0.687075	-28.0593	28.05931				
2.9527	3.783	-0.8303	0.8303	0.689398	-28.12	28.12003				
1.1976	3.783	-2.5854	2.5854	6.684293	-215.882	215.8818				
1.1067	3.783	-2.6763	2.6763	7.162582	-241.827	241.8271				
1.8351	3.783	-1.9479	1.9479	3.794314	-106.147	106.1468				
1.9253	3.783	-1.8577	1.8577	3.451049	-96.4889	96.48886				

O valor do erro percentual médio absoluto (MAPE) foi o critério utilizado para a determinação do modelo candidato a atuar em conjunto com o método adotado pelo SCA.

Na tabela acima, observa-se que o MAPE da previsão do SCA corresponde a 132.73%. A seguir, será realizado o mesmo cálculo para as previsões abordadas nesse trabalho.

5.2 – ACURÁCIA DA PREVISÃO DO MODELO HOLT-WINTERS

A tabela abaixo repreSENTA as medidas de acurácia comparando os valores reais da velocidade do vento com os valores obtidos pelo método Holt-Winters.

Tabela 53 – Medidas de acurácia da previsão Holt-Winters para os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão Holt-Winters	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,091907983	-0,01320798	0,01321	0,00017445	-1,22443523	1,2244352	0,71758319	0,871605	24,3762683	31,429485
1,9438	1,266908871	0,676891129	0,67689	0,4581816	34,8230851	34,823085				
2,9541	1,321728555	1,632371445	1,63237	2,66463654	55,2578262	55,257826				
2,9527	1,294049538	1,658650462	1,65865	2,75112135	56,1740259	56,174026				
1,1976	1,371433989	-0,17383399	0,17383	0,03021826	-14,5151962	14,515196				
1,1067	1,244741294	-0,13804129	0,13804	0,0190554	-12,4732352	12,473235				
1,8351	1,139366134	0,695733866	0,69573	0,48404561	37,912586	37,912586				
1,9253	1,173364648	0,751935352	0,75194	0,56540677	39,0554901	39,05549				

O valor de MAPE obtido para previsão dos oito primeiros valores do conjunto de teste se mostrou bem aceitável, apesar desse método não ser o mais recomendável para previsão de séries temporais com alta frequência de aquisição.

5.3 – ACURÁCIA DA PREVISÃO DO MODELO SARIMA

A tabela abaixo reapresenta as medidas de acurácia comparando os valores reais da velocidade do vento com os valores obtidos pelo modelo SARIMA.

Tabela 54 – Medidas de acurácia da previsão SARIMA para os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão SARIMA	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,122090738	-0,04339074	0,04339	0,00188276	-4,02250279	4,0225028	0,98272178	1,46734065	44,0489668	45,0545925
1,9438	1,008348994	0,935451006	0,93545	0,87506859	48,1248588	48,124859				
2,9541	0,949455295	2,004644705	2,00464	4,01860039	67,8597442	67,859744				
2,9527	0,905058969	2,047641031	2,04764	4,19283379	69,3480893	69,348089				
1,1976	0,863215235	0,334384765	0,33438	0,11181317	27,9212396	27,92124				
1,1067	0,824671436	0,282028564	0,28203	0,07954011	25,4837412	25,483741				
1,8351	0,78951164	1,04558836	1,04559	1,09325502	56,9771871	56,977187				
1,9253	0,756654896	1,168645104	1,16865	1,36573138	60,6993769	60,699377				

Devido à estabilidade estatística, em termos de raiz unitária e uma aceitável autocorrelação entre os resíduos desse modelo, era esperado um melhor valor de MAPE. Contudo, o modelo SARIMA também não conta com a robustez matemática exigida para o trato de séries intradiárias.

5.4 – ACURÁCIA DA PREVISÃO DO MODELO STL/SES/ARMA

Finalmente, a tabela abaixo reapresenta o valor de MAPE obtido ao comparar-se os valores reais da velocidade do vento com os valores previstos pelo modelo híbrido STL/SES/ARMA aplicado sobre o conjunto de treinamento.

Tabela 55 – Medidas de acurácia da previsão STL/SES/ARMA para os 8 primeiros valores de teste.

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Valores Reais	Previsão STL/SES/ARMA	Erro	Erro	Erro^2	Erro%	Erro%	MAE	MSE	MPE	MAPE
3,783										
1,0787	1,231408577	-0,15270858	0,15271	0,02331991	-14,1567236	14,156724	0,72999516	0,93847583	27,0349819	31,3477796
1,9438	1,213344877	0,730455123	0,73046	0,53356469	37,5787181	37,578718				
2,9541	1,243518416	1,710581584	1,71058	2,92608935	57,9053378	57,905338				
2,9527	1,26051735	1,69218265	1,69218	2,86348212	57,3096708	57,309671				
1,1976	1,199036264	-0,00143626	0,00144	2,0629E-06	-0,11992853	0,1199285				
1,1067	1,139619223	-0,03291922	0,03292	0,00108368	-2,97453902	2,974539				
1,8351	1,127960489	0,707139511	0,70714	0,50004629	38,5341132	38,534113				
1,9253	1,112761673	0,812538327	0,81254	0,66021853	42,2032061	42,203206				

O valor de MAPE obtido é considerado aceitável e oferece as vantagens de o modelo não apresentar as restrições estatísticas dos modelos anteriores, além de ser o mais indicado para o tratamento de séries intradiárias.

6 – CONCLUSÕES

Este trabalho se destina a agregar maior precisão ao método de previsão utilizado pelo SCA para períodos de uma e duas horas à frente, de modo a mitigar a vulnerabilidade desse método diante de um valor futuro consideravelmente destoante em relação ao último valor conhecido. Para tal foi realizado um estudo comparativo entre as principais técnicas de previsão disponíveis na teoria de análise de séries temporais, de modo a desenvolver um modelo híbrido, que utilize um método matemático robusto o suficiente para lidar adequadamente com as características particulares das séries temporais intradiárias de velocidade do vento oriundas de um cenário geográfico complexo, como aquele onde se encontra a CNAAA. Uma vez estabelecido o modelo mais adequado para previsão de velocidade do vento, obtida de uma das torres meteorológicas do complexo da CNAAA, esse pode ser expandido para os outros dados coletados por essa torre, assim como para as demais torres pertencentes ao complexo.

A primeira técnica analisada, conhecida como modelo clássico de análise de séries temporais, não pôde ser concluída devido à limitação da massa de dados disponível. Essa, apesar de volumosa, totalizando mais de quinze milhões de registros, abrange somente oito anos e isso não foi suficiente para a determinação da componente cíclica de longo prazo conhecida como ciclo. Sendo assim a previsão utilizando este modelo ficou comprometida.

A segunda técnica considerada neste trabalho foi o modelo de suavização exponencial tripla, ou método de Holt-Winters. Os resíduos gerados por esse método não seguem uma distribuição normal e sua FAC mostra que o número de *lags* fora do intervalo de confiança superou o limite aceitável de 5%, indicando assim que os resíduos não são do tipo iid como esperado.

O terceiro modelo de previsão avaliado pertence à categoria de modelos conhecidos como autorregressivos. Devido à presença de um padrão sazonal na série WS_m, o modelo $SARIMA(p,d,q)(P,D,Q)m$ foi utilizado. O conjunto de parâmetros do modelo que apresentou menor valor de AICc e com suas raízes no interior do círculo unitário foi $(3,1,2)(0,0,1)[48]$. Os resíduos gerados pelo modelo também não seguem uma distribuição normal. A análise de sua FAC apresentou alguns *lags* com valores de autocorrelação fora do intervalo de confiança, porém não superando o limite aceitável de 5%.

A quarta categoria de modelagem estudada foi a decomposição STL. Esta metodologia oferece a robustez matemática desejada para o trato de séries temporais de previsão de velocidade do vento com alta frequência de aquisição (séries intradiárias) e proveniente de um relevo complexo. A decomposição STL ocorre sobre o conjunto de treinamento, gerando as componentes de tendência, sazonalidade e restante. A seguir, é definido o modelo estatístico que melhor represente cada componente, de modo a viabilizar uma previsão de 1034 valores à frente. Assim, relativo à componente tendência, o modelo de ajuste que se mostrou mais adequado foi o método de suavização exponencial simples. A sazonalidade diária tende a se manter constante durante o mês, então a previsão de seus valores ocorre replicando-se os valores dos dez últimos dias apresentados por essa componente. O modelo de ajuste mais adequado para a componente restante foi o modelo autorregressivo ARMA, sobre o qual foi realizada a previsão para os 1034 valores à frente. Finalmente, a previsão resultante de cada componente é somada (pois a série decomposta é do tipo aditiva) obtendo-se então a previsão desejada.

Todos os modelos aqui avaliados e que forneceram uma previsão para a velocidade de vento apresentaram um valor de MAPE menor que o obtido pelo modelo utilizado pelo SCA. Contudo, é pertinente ressaltar que tal fato ocorreu devido à situação especial selecionada, isto é, o último valor do conjunto de treinamento diferindo substancialmente do primeiro valor do conjunto de teste.

Além do modelo híbrido STL/SES/ARMA contar com a robustez matemática necessária no trato de séries temporais de alta frequência de aquisição e oriundas de um relevo complexo e costeiro, esse modelo foi o que apresentou o melhor resultado numérico, avaliado em termos de MAPE. Cabe também ressaltar que, independente desse fato, esse modelo seria o escolhido, visto que atende às características particulares das séries utilizadas neste trabalho, garantindo assim uma estabilidade estatística.

O ganho de acurácia desse modelo híbrido em relação ao modelo do SCA foi de 101,39% (132.73% – 31.34%). Assim sendo, o modelo sugerido mostrou-se estatisticamente confiável, factível e relevante em termos de ganho de acurácia diante de uma situação extrema.

7 – TRABALHOS FUTUROS

Considerando-se a limitação dos modelos tradicionais de previsão no tocante às séries intradiárias, a recomendação para trabalhos futuros se baseia nos modelos que não enfrentam tais restrições. Assim, pode-se enumerar as seguintes abordagens:

- 1 – Utilização de modelos TBATS sobre uma série que sabidamente apresente múltipla sazonalidade. Nesse sentido, pode-se considerar uma série semelhante a WS_m, porém abrangendo um ano ou mais.
- 2 – No mesmo senso do item anterior, seria a utilização de Regressão Harmônica Dinâmica com Vários Períodos Sazonais.
- 3 – Utilização de autorregressões vetoriais.
- 4 – Utilização de modernas arquiteturas de redes neurais profundas, por exemplo, LSTM e Transformers, que têm se demonstrado muito eficientes para previsão de séries temporais.
- 5 – Aplicação de análise bayesianas.
- 6 – Mudança do domínio temporal para o domínio de frequência.
- 7 – Expansão para um modelo multivalorado envolvendo estabilidade atmosférica, velocidade e direção do vento e considerando as outras torres.

REFERÊNCIAS:

BUENO, RODRIGO DE LOSO DA SILVEIRA. **Econometria de Séries Temporais 2^a edição.** São Paulo: Cengage Learning, 2015.

CAVALCANTI, I.F.A.; FERREIRA, N.J.; DIAS, M.A.F.S; SILVA, G.A.J. **Tempo e clima no Brasil.** 1.ed. São Paulo: Oficina de Textos, 2009.

CHUANJIN YU, YONGLE LI, MINGJIN ZHANG. **An improved Wavelet Transform using Singular Spectrum Analysis for wind speed forecasting based on Elman Neural Network,** Energy Conversion and Management, Volume 148, 2017, Pages 895-904, ISSN 0196-8904.

CLEVELAND WILLIAM. **Robust locally weighted regression and smoothing scatterplots.** Journal of the American Statistical Association, v.74, p. 829-836, 1979.

CLEVELAND, R. B., CLEVELAND, W. S., McRAE, J. E., & TERPENNING, I. J. (1990). **STL: A seasonal-trend decomposition procedure based on loess.** *Journal of Official Statistics*, 6(1), 3–33.

CLEVELAND W. S.; LOADER C. **Smoothing by local regression: Principles and methods.** Physica-Verlag, p. 10-49, 1996.

C.R. RAO AIMSCS, **HandBook of Statistics 30 Time Series Analysis: Methods and Applications,** First edition 2012, University of Hyderabad Campus, Hyderabad, India.

DE LIVERA, A. M., HYNDMAN, R. J., & SNYDER, R. D. **Forecasting time series with complex seasonal patterns using exponential smoothing.** J American Statistical Association, 2011, 106(496), 1513–1527.

G. W. CHANG, H. J. LU, Y. R. CHANG, Y. D. LEE. **An improved neural network-based approach for short-term wind speed and power forecast.** Renewable Energy, Volume 105, 2017, Pages 301-311, ISSN 0960-1481.

GEORGE E. P. BOX; GWILYM M. JENKINS; GREGORY C. REINSEL; GRETA M. LJUNG **Time Series Analysis Forecasting and Control** Fifth Edition, 2016, published by John Wiley & Sons, Inc., Hoboken, New Jersey.

HENRIQUE DO NASCIMENTO CAMELO, PAULO SÉRGIO LUCIO, JOÃO BOSCO VERÇOSA LEAL JUNIOR. **Modelagem de Média Mensal de Velocidade do Vento para Região Litorânea no Nordeste Brasileiro Através do Método Aditivo HOLT-WINTERS com Vias a Previsão de Geração Eólica**. Revista Brasileira de Energias Renováveis, v.5, n.4, p.587-604, 2016.

HUI LIU, CHAO CHEN, HONG-QI TIAN, YAN-FEI LI. **A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks**, Renewable Energy, Volume 48, 2012, Pages 545-556, ISSN 0960-1481.

HUI LIU, HONG-QI TIAN, CHAO CHEN, YAN-FEI LI. **An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization**, International Journal of Electrical Power & Energy Systems, Volume 52, 2013, Pages 161-173, ISSN 0142-0615.

HYNDMAN ROB J; GEORGE ATHANASOPOULOS **Forecasting Principles and Practice** Second print edition, May 2018, published by OTEXT.

INA KHANDELWAL, RATNADIP ADHIKARI, GHANSHYAM VERMA. **Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition**. Procedia Computer Science, Volume 48, 2015, Pages 173-179, ISSN 1877-0509.

JIANMING HU, JIANZHOU WANG, KAILIANG MA. **A hybrid technique for short-term wind speed prediction**, Energy, Volume 81, 2015, Pages 563-574, ISSN 0360-5442.

LOADER CLIVE. **Local regression and likelihood**. Springer-Verlag, 1999.

LOADER CLIVE. **Old faithful erupts: Bandwidth selection reviewed**. Working paper, ATT Bell Laboratory, 1995.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. **Forecasting: Methods and Applications**. 3.ed., New York: John Wiley & Sons, 1998.

MARQUETTI ADALMIR e VIALI LORÍ, **Princípios Básicos de Regressão Local**, Revista Análise Econômica, ano 22, n. 42, setembro 2014.

MORETTIN, A. PEDRO; TOLOI, C. M. CÉLIA, **Análise de Séries Temporais**, 3^a reimpressão, Editora Edgard Blücher – 2014.

NADARAYA E.A. **On estimating regression. Theory of Probability and its Applications**, v.9, p. 141-142, 1964.

NGAI HANG CHAN **Time Series Applications to Finance with R and S-Plus®** Second edition, 2010, A JOHN WILEY & SONS, INC., PUBLICATION.

NIELSEN, AILEEN. **Análise Prática de Séries Temporais**. Copyright @ 2021 da Starlin Alta Editora e Consultoria Eireli. ISBN: 978-85-5081-526-6.

PING JIANG, SHANSHAN QIN, JIE WU, BEIBEI SUN. **Time Series Analysis and Forecasting for Wind Speeds Using Support Vector Regression Coupled with Artificial Intelligent Algorithms**. Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 939305, 14 pages.

R Core Team (2022). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

REGINALDO NUNES DA SILVA, DARIO GERALDO FANTINI, RAFAEL CASTILHO FARIAS MENDES, ANTONIO CESAR PINHO BRASIL JUNIOR. **Previsão de Velocidade do Vento com Série Temporal Multivariada Utilizando Rede Neural Recorrente**. IX Congresso Brasileiro de Energia Solar – Florianópolis, 23 a 27 de maio de 2022.

SILVA, G. R. **Características de Vento da Região Nordeste** 2003. 131f. Dissertação (Mestrado Engenharia Mecânica) – Universidade Federal de Pernambuco, Recife, 2003.

STONE C. J. **Consistent nonparametric regression, with discussion.** The annals of Statistics, 5 p.549-645, 1977.

STONE C. J. **Optimal rates of convergence for nonparametric estimators.** The annals of Statistics, v.8 p.1348-1360, 1980.

TIAGO MENDES DANTAS, FERNANDO LUIZ CYRINO OLIVEIRA. **Previsão de Velocidade de Vento: Uma Abordagem Utilizando Bagging Holt-Winters com Decomposição STL.** XLVI Simpósio Brasileiro de Pesquisa Operacional –16 a 19 setembro de 2014 Salvador/BA.

TONG NIU, JIANZHOU WANG, KEQUAN ZHANG, PEI DU. **Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy.** Renewable Energy, Volume 118, 2018, Pages 213-229, ISSN 0960-1481.

VAREJÃO SILVA, M. A. **Meteorologia e Climatologia.** Versão digital 2 Recife 2006. https://icat.ufal.br/laboratorio/clima/data/uploads/pdf/METEOROLOGIA_E_CLIMATOLOGIA_VD2_Mar_2006.pdf (acesso realizado em novembro de 2022).

WATSON G.S. **Smooth regression analysis.** Sankhya, Series. A, v.26, p. 359-372, 1964.

WESSA P., 2017, **(Partial) Autocorrelation Function (v1.0.15) in Free Statistics Software (v1.2.1), Office for Research Development and Education, URL** http://www.wessa.net/rwasp_autocorrelation.wasp/ (acesso realizado em março de 2021).

YU JIANG, ZHE SONG, ANDREW KUSIAK. **Very short-term wind speed forecasting with Bayesian structural break model.** Renewable Energy, Volume 50, 2013, Pages 637-647, ISSN 0960-1481.

ZHENKUN LIU, PING JIANG, LIFANG ZHANG, XINSONG NIU. **A combined forecasting model for time series: Application to short-term wind speed forecasting.** Applied Energy, Volume 259, 2020, 114137, ISSN 0306-2619.