



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA**

EVA CABRAL DE PONTES

**ANÁLISE DE SÉRIES TEMPORAIS: MODELANDO DADOS DE IMPORTAÇÃO DE
GÁS NATURAL NO BRASIL**

CAMPINA GRANDE - PB

2021

EVA CABRAL DE PONTES

**ANÁLISE DE SÉRIES TEMPORAIS: MODELANDO DADOS DE IMPORTAÇÃO DE
GÁS NATURAL NO BRASIL**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr. Sílvio Fernando Alves Xavier Júnior

CAMPINA GRANDE - PB

2021

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

P814a Pontes, Eva Cabral de.

Análise de séries temporais [manuscrito] : modelando dados de importação de gás natural no Brasil / Eva Cabral de Pontes. - 2021.

41 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia , 2021.

"Orientação : Prof. Dr. Sílvio Fernando Alves Xavier Júnior , Departamento de Estatística - CCT."

1. Séries temporais. 2. Gás natural. 3. Modelos SARIMA.
4. Modelagem Box-Jenkins. I. Título

21. ed. CDD 519.55

EVA CABRAL DE PONTES

ANÁLISE DE SÉRIES TEMPORAIS: MODELANDO DADOS DE IMPORTAÇÃO DE GÁS
NATURAL NO BRASIL

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Trabalho aprovado em 05 de Outubro de 2021.

BANCA EXAMINADORA



Prof. Dr. Sílvio Fernando Alves Xavier Júnior
(Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba (UEPB)



Profa. Ma. Nyedja Fialho Morais Barbosa
Universidade Estadual da Paraíba (UEPB)

Dedico a Deus ao meu esposo.

AGRADECIMENTOS

Agradeço a Deus, porque sem Ele nada seria possível.

Ao meu esposo Adílio que acima de tudo é um grande amigo que fez parte da minha jornada acadêmica, sempre estando ao meu lado me apoiando, e nos momentos difíceis sempre tinha uma palavra de incentivo.

Ao meu orientador professor Sílvio por aceitar conduzir o meu trabalho de pesquisa.

Aos meus pais José Antônio e Severina pelo amor, incentivo e apoio.

Grata a todos os meus professores que contribuíram com a minha formação acadêmica.

E por último, mas não menos importante a todos os meus colegas de curso, pelos inúmeros conselhos, frases de motivação, risadas, enfim... muito obrigada.

"Por mais que as coisas fujam ao nosso controle, elas continuam rigorosamente sob o controle de Deus."
(Hernandes Dias Lopes)

RESUMO

O gás natural é uma substância composta por hidrocarbonetos leves que permanecem em estado gasoso nas condições atmosféricas normais. Atualmente o país produz mais de 130 milhões de m^3 de gás natural por dia o que representa cerca 13,7% do consumo interno. O Brasil importa gás natural de alguns países, onde o preço dessas importações aumentaram no primeiro trimestre de 2021, com relação ao primeiro trimestre de 2020. Diante do exposto, este trabalho tem por objetivo fazer previsões baseando-se em metodologias utilizadas na análise de séries temporais, tendo como finalidade prever os preços das importações para o Brasil até o 4º trimestre de 2021, com base nos dados do período de janeiro de 2000 a maio de 2021 da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). As previsões foram calculadas usando o algoritmo de alisamento exponencial de Holt-Winters, aditivo e multiplicativo, e também o método de modelagem de Box-Jenkins, SARIMA e SARIMAX. Todas as previsões foram geradas com 1, 4 e 7 passos a frente, e os valores das somas dos quadrados dos erros foram a forma de avaliar quais os melhores valores previstos. Na modelagem SARIMAX aplicou-se uma variável *dummy*, com valores 1 para os meses a partir de janeiro de 2020 até maio de 2021, que foi o período onde iniciou os primeiros casos de Covid-19 em outros países fora da China, e com o valor de 0 para os meses anteriores para verificar se afetariam de alguma forma em suas previsões. Para tal análise optou-se pelo uso função *auto.arima*, disponível no pacote *forecast* implementado no software R. Todos os modelos indicam uma boa capacidade de previsão, com intervalos de valores bem próximos, apresentando diferenças também não tão gritantes em seus erros relativos, entretanto, o modelo que obteve o melhor ajuste foi o modelo SARIMAX com o seu critério de informação (AIC) e erro mais baixo, obteve o melhor ajuste com relação aos erros das previsões com 7 passos a frente serem bem mais baixos.

Palavras-chave: Séries temporais. Importação. Gás Natural. Modelos SARIMA.

ABSTRACT

Natural gas is a substance composed of light hydrocarbons that remain in a gaseous state under normal atmospheric conditions. Currently, the country produces more than 130 million m^3 of natural gas per day, which represents about 13.7% of domestic consumption. Brazil imports natural gas from some countries, where the price of these imports increased in the first quarter of 2021, compared to the first quarter of 2020. In light of this, this paper aims to make forecasts based on methodologies used in the analysis of series with the purpose of forecasting the prices of imports into Brazil until the 4th quarter of 2021, based on data for the period from January 2000 to May 2021 from the National Agency for Petroleum, Natural Gas and Biofuels (ANP). Predictions were calculated using the Holt-Winters exponential smoothing algorithm, additive and multiplicative, as well as the Box-Jenkins, SARIMA and SARIMAX modeling method. All forecasts were generated 1, 4 and 7 steps ahead, and the values of sums of squares of the errors were the way to evaluate which the best predicted values were. In the SARIMAX modeling, a variable *dummy* was applied, with values of 1 for the months from January 2020 to May 2021, which was the period in which the first Covid-19 cases started in other countries outside China, and with a value of 0 for the previous months to see if they would affect your forecasts in any way. For this analysis, we chose to use the *auto.arima* function, available in the *forecast* package implemented in the R software. All models indicate a good predictive capacity, with very close value intervals, also showing differences not so glaring in their relative errors, however, the model that obtained the best fit was the SARIMAX model with its information criterion (AIC) and lowest error, it obtained the best fit with respect to forecast errors with 7 steps forward be much lower.

Keywords: Time series. Importation. Natural Gas. SARIMA Models.

LISTA DE ILUSTRAÇÕES

Figura 1 – Desvios da temperatura global média anual (1880-2015) em °C.	14
Figura 2 – Séries temporais caracterizadas por ruído branco, ruído rosa e ruído marron (à esquerda) e respectivos espectros de potência (à direita).	16
Figura 3 – Gráfico da série temporal de importação de gás natural para o período de jan/2000 a mai/2021.	28
Figura 4 – Gráfico da série, com transformação logarítmica dos valores de importação por m^3 no Brasil janeiro de 2000 a maio de 2021.	28
Figura 5 – Histograma da serie temporal da importação do gás natural	29
Figura 6 – Gráficos de autocorrelação amostral da série (ACF) e autocorrelação parcial da série (PACF)	29
Figura 7 – Gráficos de autocorrelação amostral da série (ACF) e autocorrelação parcial da série (PACF), aplicando a primeira diferença.	30
Figura 8 – Valores de importação de gás natural mil por m^3 no Brasil, dos períodos de janeiro de 2000 a maio de 2021	30
Figura 9 – Gráficos dos dados da série e seu ajuste de Holt-Winters.	32
Figura 10 – Gráficos da previsão 1 passos a frente.	33
Figura 11 – Gráficos da previsão 4 passos a frente.	34
Figura 12 – Gráficos da previsão 7 passos a frente.	34
Figura 13 – Gráficos de diagnóstico dos resíduos SARIMA com 4 passos a frente	37
Figura 14 – Gráficos de diagnóstico dos resíduos SARIMAX com 4 passos a frente . . .	37
Figura 15 – Gráficos de diagnóstico dos resíduos SARIMA com 7 passos a frente	38
Figura 16 – Gráficos de diagnóstico dos resíduos SARIMAX com 7 passos a frente . . .	38

LISTA DE TABELAS

Tabela 1 – Análises descritivas dos dados	27
Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e multiplicativo.	31
Tabela 3 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m^3))	32
Tabela 4 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m^3))	32
Tabela 5 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m^3))	33
Tabela 6 – Valores dos erros de previsão.	35
Tabela 7 – Valores dos critério de informação dos modelos da série e erros relativos, 1 passo à frente.	35
Tabela 8 – Previsões dos modelos SARIMA e SARIMAX estudados, 4 passo à frente. .	36
Tabela 9 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passo à frente. .	36
Tabela 10 – Erros relativos de previsão dos modelos 4 passos a frente	36
Tabela 11 – Erros relativos de previsão dos modelos 7 passos a frente	36

SUMÁRIO

1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Séries Temporais	13
2.2	Processos Estocásticos	14
2.3	Estacionariedade	14
2.4	Ruídos	15
2.5	Funções de Autocovariância e Funções de Autocorrelação	17
2.6	Suavização Exponencial	17
2.6.1	<i>Suavização Exponencial Simples</i>	18
2.6.2	<i>Alisamento Exponencial de Holt</i>	18
2.6.3	<i>Alisamento Exponencial de Holt-Winters</i>	19
2.7	Modelagem Box-Jenkins	20
2.7.1	<i>Modelo autorregressivo (AR)</i>	21
2.7.2	<i>Modelo autorregressivo (MA)</i>	21
2.7.3	<i>Modelo autorregressivo de médias móveis (ARMA)</i>	21
2.7.4	<i>Modelos autorregressivos integrados de médias móveis (ARIMA)</i>	22
2.7.5	<i>Modelos autorregressivos de médias móveis sazonal (SARIMA)</i>	22
2.7.6	<i>Modelos Autorregressivos de Médias Móveis Sazonal com Variáveis Exógenas (SARIMAX)</i>	22
2.7.7	<i>Função de autocorrelação (FAC)</i>	23
2.7.8	<i>Função de autocorrelação parcial(FACP)</i>	23
2.7.9	<i>Etapas do modelo SARIMA</i>	23
2.8	Avaliação de Previsão	25
3	APLICAÇÃO	27
3.1	Modelagem – Alisamento exponencial	31
3.2	Modelagem Box-Jekins	35
4	CONCLUSÃO	39
	REFERÊNCIAS	40

1 INTRODUÇÃO

Em meados da década de 1940 a exploração, produção e utilização de gás natural do Brasil foi iniciada. Ele é de suma importância para a economia e também para o meio ambiente. Começando por sua qualidade e capacidade de substituir qualquer combustível derivado do petróleo, como o carvão, diesel, óleo e álcool, serve também para uso de fertilizantes, e na produção de ureia e amônia. O gás natural é um combustível de origem fóssil, composto principalmente por hidrocarbonetos (carbono e hidrogênio), e é encontrado no estado físico gasoso. Segundo o ANP (2020) atualmente o país produz mais de 130 milhões de m^3 de gás natural por dia o que representa cerca 13,7% do consumo interno. No mundo o consumo desta importante fonte de energia vem crescendo e tem como uma de suas vantagens a baixa emissão de resíduos poluentes.

Em 1999 iniciaram as importações de gás natural no Brasil, com a entrada em operação do gasoduto Bolívia-Brasil. No ano seguinte iniciou com a entrada em operação dos dois primeiros trechos do gasoduto Uruguaina-Porto Alegre e assim também a importação de gás natural a Argentina, compreendendo duas importantes reservas da América do Sul ao mercado brasileiro (LAUREANO, 2005).

De acordo com a Petrobrás (2019), conforme citado pela ANP (2020) ao todo nove países são responsáveis por exportar o produto para o Brasil e entre eles, o principal é a Bolívia, a qual no ano de 2019 apresentou uma diminuição na importação, pois o Brasil vem comprando cada vez menos gás da Bolívia. Um dos motivos que explicam a redução das importações de gás boliviano é o preço menor do gás natural liquefeito (GNL) no mercado internacional de países como os Estados Unidos, isso levou o Brasil a reduzir a compra na Bolívia.

De acordo com o boletim trimestral da Agência Nacional de Petróleo, Gás natural e Biocombustível (ANP), as importações de gás natural totalizaram 39,6 milhões de metros cúbicos por dia no primeiro trimestre de 2021, volume que apresentou um aumento de 59,78% em relação ao primeiro trimestre de 2020. Na comparação com o trimestre anterior, houve um crescimento de 57,67% no volume importado. Em termos de valores, as importações de gás natural totalizaram um dispêndio de R\$ 798,1 milhões de Reais no primeiro trimestre de 2021, o valor que corresponde a um aumento de 86,68% em relação ao montante destinado às importações do produto no primeiro trimestre de 2020 (ANP, 2021).

Para a realização desse estudo verificou-se a viabilidade do Método de suavização exponencial de Holt-Winters que produz uma previsão para um período futuro, verificando se existe nível, tendência de crescimento e sazonalidade dos dados e utilizou-se a metodologia Box-Jenkins, na medida de previsão de erro MAE, MSE, MAPE, em comparação ao modelo

HoltWinters, que possui duas equações de previsão: aditivo e multiplicativo, nos valores de importação no Brasil . A análise dos dados foi realizada no período de janeiro de 2000 a maio de 2021, organizados na série temporal. O algoritmo de Box-Jenkins auxilia o usuário a escolher o melhor modelo baseado nos gráficos da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP).

O presente trabalho tem por finalidade utilizar a metodologia de Holt-Winters e Box-Jenkins para analisar os valores das importações no Brasil, tais como, as tendências. Através do alisamento exponencial de Hot-Winters, aditivo e multiplicativo, com previsões de 1, 4 e 7 passos à frente, prever como os valores de importações procederá até o 4º trimestre de 2021. Subsequente, utiliza-se a modelagem Box-Jenkins, com os modelos SARIMA e SARIMAX, onde busca-se o melhor modelo por meio dos dados.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Séries Temporais

Uma série temporal pode ser definida como uma coleção de variáveis aleatórias indexadas, de acordo com a ordem em que são obtidas por um período de tempo. Por exemplo, se coletarmos dados sobre temperaturas máximas diárias, podemos considerar a série temporal como uma sequência de variáveis aleatórias, Y_1, Y_2, Y_3, \dots , onde a variável aleatória Y_1 , indica a temperatura máxima no dia um, a variável Y_2 mostrando a temperatura máxima no segundo dia, Y_3 indicando o valor para o terceiro dia e assim por diante (SHUMWAY; STOFFER, 2017).

Este conjunto de observações podem ser discretas: $T = \{t_1, t_2, \dots, t_n\}$, quando acontecem em um tempo específico, por exemplo: Valores diários de poluição na cidade de São Paulo, ou contínuas: $T = \{t : t_1 < t < t_2\}$, quando as observações acontecem continuamente, por exemplo: Registro de marés no porto de Santos. Uma série temporal discreta também pode ser obtida através da amostragem de uma série temporal contínua em intervalos de tempo iguais (MORETTIN; TOLOI, 2004).

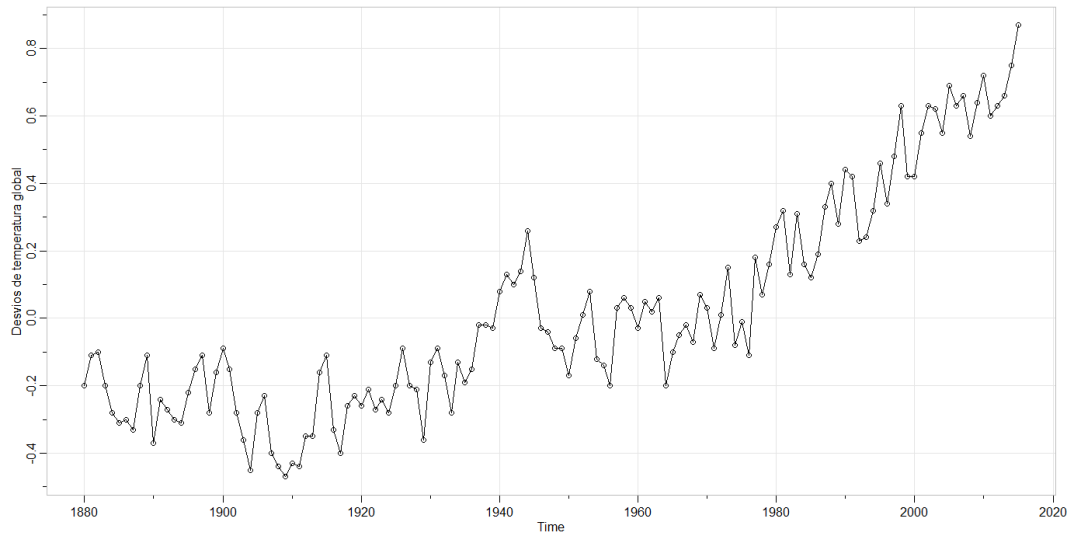
Para uma melhor compreensão do comportamento de uma série, faz-se necessário entender os seguintes componentes característicos, que são eles:

- **Sazonalidade:** movimentos padrões que se repetem em períodos idênticos de tempo.
- **Tendência:** comportamento a longo prazo apresentado pela série, que se reflete as mudanças, como diminuição ou aumento, que ocorrem nos dados de maneira consistente.
- **Ciclo:** equivale aos movimentos de elevação e queda, que ocorrem repetidamente, em torno da tendência.
- **Erro aleatório:** as variações que não são idênticas pelos demais componentes.

A seguir um exemplo de uma série temporal.

Considere o registro da série de temperatura global mostrado na Figura 1. Os dados são o índice global de temperatura média terra-oceano de 1880 a 2015, com o período base 1951-1980. Os valores são desvios (°C) de 1951-1980 média, atualizado de Hansen et al. (2006). A tendência de alta da série durante a última parte do século XX foi usado como um argumento para a hipótese da mudança climática.

Figura 1 – Desvios da temperatura global média anual (1880-2015) em °C.



2.2 Processos Estocásticos

O conjunto de valores $\{Y(t), t \in T\}$ é chamado de espaço de estados e os valores $Y(t)$ são chamados de estados. Seja T um conjunto arbitrário, tal que, $\forall t \in T$, $Y(t)$ é uma variável aleatória. Para cada t , $Y(t)$ tem uma distribuição de probabilidade. São através das seguintes funções, que os processos estocásticos são descritos (MORETTIN; TOLOI, 2004):

- **Função média:** $Y_t: E[Y_t] = \mu(t)$
- **Função auto covariância:** $\gamma_{t,s} = cov[Y_t, Y_s] = E[(Y_t - \mu_t)(Y_s - \mu_s)]$, sendo t e s tempos distintos.
- **Variância:** $Var[Y_t] = \sigma_t^2$

2.3 Estacionariedade

Para fazer inferências estatísticas sobre a estrutura de um processo estocástico com base em um registro observado desse processo, devemos geralmente fazer alguma simplificação suposições sobre essa estrutura. O mais importante desses pressupostos é o da estacionariedade. A ideia básica por trás do conceito de estacionariedade é que o processo encontra-se em equilíbrio estatístico, ou seja, em termos probabilísticos que regem o comportamento do processo não mudam com o tempo. Em certo sentido, o processo está em equilíbrio estatístico. Existem dois tipos de estacionariedade que são Estacionariedade Forte e Estacionariedade Fraca (MORETTIN;

TOLOI, 2004).

- **Estacionariedade Forte (fortemente estacionária):** Um processo estocástico $Y = Y_t, t \in T$ diz-se estritamente estacionário se todas as distribuições finito-dimensionais, ou seja, $F(y_1, \dots, y_n; t_1, \dots, t_n) = P\{Y_{t_1} \leq z_1, \dots, Y_{t_n} \leq z_n\}$, permanecem constantes no tempo.

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

$$E(t) = \mu, e \text{Var}(t) = \sigma^2, \forall t \in T. \quad (2.1)$$

- **Estacionariedade Fraca (fracamente estacionária.):** Um processo estocástico $Y = Y_t, t \in T$ diz-se fracamente estacionário, se e somente se:

1. $E[Y_t] = \mu$ (constante) $\forall t \in T$;
2. $E^2[Y_t] < \infty, t \in T$;
3. $\gamma_{t,s} = \text{Cov}[Y_t, Y_s]$ é uma função de $t - s$, chamado de defasagem.

2.4 Ruídos

Ruído é uma perturbação indesejada dentro de uma faixa de frequências de interesse originada por fontes artificiais ou naturais. Consiste num distúrbio que afecta e que pode distorcer a informação, originando um sinal aleatório com propriedades estatísticas conhecidas de amplitude, distribuição e densidade espectral. O ruído deriva de um sinal aleatório e pode apresentar diferentes propriedades estatísticas consoante a fonte de aleatoriedade. A densidade espectral é uma distribuição de energia no espectro de frequências e é uma propriedade que permite distinguir os diferentes tipos de ruído. (MOREIRA, 2013). A seguir tem-se a definição de 3 tipos de Ruídos.

Ruído Branco: é definido como uma sequência de variáveis aleatórias i.i.d.(independente e identicamente distribuída), sendo um processo estocástico não-correlacionado e estritamente estacionário, geralmente assumida como normal, tendo média zero e variância σ_e^2 .

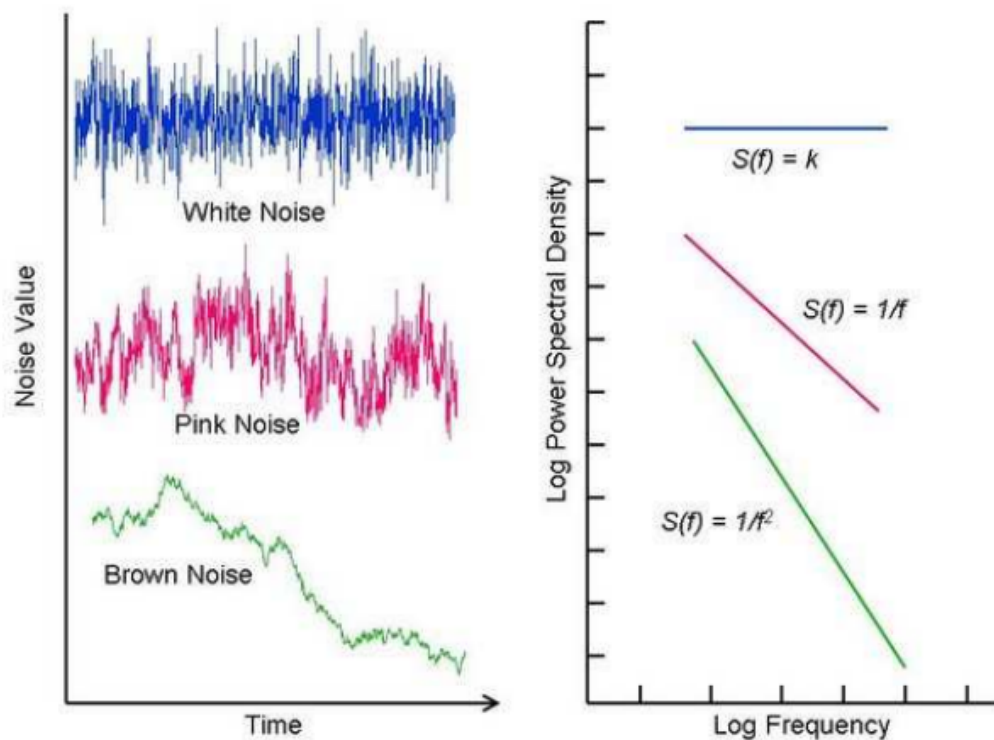
Ruído Rosa: distribui aleatoriamente na banda de baixa frequência com o espectral densidade $S(f)$ proporcional a $1/f^\gamma$ com o expoente γ sendo 1 (ZHOU et al., 2012). Para altas frequências o ruído rosa nunca é dominante, esse domínio é assegurado pelo ruído branco. Já

nas baixas frequências, o ruído rosa tem maior energia do que o ruído branco (MOREIRA, 2013).

Ruído Marron: Também conhecido por passeio aleatório, ele é um ruído produzido pelo movimento Browniano. A sua densidade espectral é inversamente proporcional a f^2 , indicando que tem mais energia nas baixas frequências, superior ao ruído rosa. Este tipo de ruído pode ser obtido através da integração do ruído branco. É um processo com tendência estocástica, que não manifesta condições de estacionaridade e que não é possível fazer previsões usando os registros da série (MOREIRA, 2013).

As diferentes definições de ruído assumem um sinal com componentes para todas as frequências, com uma densidade de potência espectral por unidade de largura de banda proporcional a f^k , onde k é o índice espectral ($k \leq 0$) como mostra a Figura 2.

Figura 2 – Séries temporais caracterizadas por ruído branco, ruído rosa e ruído marron (à esquerda) e respectivos espectros de potência (à direita).



Fonte: (MOREIRA, 2013)

2.5 Funções de Autocovariância e Funções de Autocorrelação

A Função de Autocovariância é utilizada para análise de dependência na estrutura estocástica de uma série temporal, para o conhecimento acerca da dependência entre as observações recentes e passadas.

De acordo com Brockwell e Davis (2016), uma série temporal Y_t estacionária, com $E[Y_t^2] < \infty$, apresenta uma função de autocovariância (FAC) dada por:

$$\gamma_k = Cov[Y_t, Y_{t-k}], \quad (2.2)$$

onde apresenta as seguintes propriedades:

1. $\gamma(0) > 0$;
2. $\gamma(-k) = \gamma(k), \forall t \in T$;
3. $|\gamma(k)| \leq \gamma(0), \forall t \in T$;

A função de autocorrelação (FAC) é originada por conta da desvantagem da dependência da unidade de medida de γ . A FAC é dada por:

$$\rho_k = \gamma(k)/\gamma(0), \quad \forall t \in T, \quad (2.3)$$

e apresenta propriedades análogas a $\gamma(k)$, exceto no caso $\rho(0) = 1$.

2.6 Suavização Exponencial

A suavização exponencial foi proposta no final da década de 1950 (BROWN, 1959); (HOLT, 2004); (WINTERS, 1960) e motivou alguns dos métodos de previsão de maior sucesso. As previsões produzidas usando métodos de suavização exponencial são médias ponderadas de observações anteriores, com os pesos decaindo exponencialmente à medida que as observações envelhecem. Em outras palavras, quanto mais recente a observação, maior o peso associado. Esta estrutura gera previsões confiáveis rapidamente e para uma ampla gama de séries temporais, o que é uma grande vantagem e de grande importância para aplicações (HYNDMAN; ATHANASSOPOULOS, 2018).

2.6.1 Suavização Exponencial Simples

O mais simples dos métodos de suavização exponencial é naturalmente chamado de suavização exponencial simples. Este método é adequado para dados de previsão sem tendência clara ou padrão sazonal. Podemos decompor a série da seguinte forma:

$$Y_t = N_t + \varepsilon_t, \quad (2.4)$$

onde $t=1, 2, \dots, N$, portanto o método simples assume que a observação mais recente é a única importante, e todas as observações anteriores não fornecem informações para o futuro. Isso pode ser considerado como uma média ponderada em que todo o peso é atribuído à última observação (HYNDMAN; ATHANASOPOULOS, 2018).

Frequentemente, queremos algo entre esses dois extremos. Por exemplo, pode ser sensato atribuir pesos maiores a observações mais recentes do que a observações do passado distante. Este é exatamente o conceito por trás da suavização exponencial simples. As previsões são calculadas usando médias ponderadas, onde os pesos diminuem exponencialmente conforme as observações vêm de mais longe no passado - os menores pesos estão associados às observações mais antigas (HYNDMAN; ATHANASOPOULOS, 2018), temos:

$$N_t = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + (1 - \alpha)^2 Y_{t-2} + \alpha(1 - \alpha)^3 Y_{t-3} + \dots, \quad (2.5)$$

Onde $0 \leq \alpha \leq 1$ é o parâmetro de suavização. A previsão de tempo um passo à frente $t + 1$ é uma média ponderada de todas as observações da série y_1, \dots, y_T . A taxa na qual os pesos diminuem é controlada pelo parâmetro α (HYNDMAN; ATHANASOPOULOS, 2018).

As representações da forma do componente dos métodos de suavização exponencial compreendem uma equação de previsão e uma equação de suavização para os componentes incluídos no método. A forma componente da suavização exponencial simples é dada por:

- Equação de previsão: $Y_t(h) = N_t$
- Equação de suavização: $N_t = \alpha Y_t + (1 - \alpha)N_{t-1}$

A equação de previsão mostra que o valor da previsão no tempo é o nível estimado no tempo. A equação de suavização para o nível, geralmente chamada de equação de nível, fornece o nível estimado da série em cada período.

2.6.2 Alisamento Exponencial de Holt

Holt (2004) estendeu a suavização exponencial simples para permitir a previsão de dados com uma tendência. Este método envolve uma equação de previsão e duas equações de suavização, uma para o nível e outra para a tendência.

- Equação de previsão: $Y_t(h) = N_t + hT_t$, onde $h = 1, 2, \dots$
- Equação de nível: $N_t = \alpha Y_t + (1 - \alpha)(N_{t-1} + T_{t-1})$
- Equação de tendência: $T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1}$

onde N_t denota uma estimativa do nível da série no momento t , T_t denota uma estimativa da tendência (inclinação) da série no tempo t , α é o parâmetro de suavização para o nível $0 \leq \alpha \leq 1$ e β é o parâmetro de suavização para a tendência $0 \leq \beta \leq 1$.

2.6.3 Alisamento Exponencial de Holt-Winters

Holt (2004) e Winters (1960) ampliaram o método de Holt para capturar a sazonalidade. O método sazonal de Holt-Winters compreende a equação de previsão e três equações de suavização, uma para o nível, uma para a tendência e uma para o componente sazonal, com os correspondentes parâmetros de suavização.

Existem duas variações desse método que diferem na natureza do componente sazonal. O método aditivo é preferido quando as variações sazonais são aproximadamente constantes ao longo da série, enquanto o método multiplicativo é preferido quando as variações sazonais estão mudando proporcionalmente ao nível da série.

- Modelo aditivo de Holt-Winters: Com o método aditivo, a componente sazonal é expressa em termos absolutos na escala das séries observadas, e na equação de nível a série é ajustada sazonalmente subtraindo a componente sazonal. Dentro de cada ano, o componente sazonal somará aproximadamente zero. As equações são dadas por:

$$N_t = \alpha(Y_t - F_{ts}) + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (2.6)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (2.7)$$

$$F_t = \gamma(Y_t - N_t) + (1 - \gamma)F_{t-s} \quad (2.8)$$

Com $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ e $0 \leq \gamma \leq 1$, sendo α , β e γ constantes.

- Modelo multiplicativo de Holt-Winters: Com o método multiplicativo, a componente sazonal é expressa em termos relativos (percentagens) e a série é ajustada sazonalmente pela divisão pela componente sazonal. Dentro de cada ano, m . A forma de recorrência deste algoritmo, é dada por:

$$N_t = \frac{\alpha Y_t}{F_{ts}} + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (2.9)$$

$$T_s = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (2.10)$$

$$F_t = \frac{t}{N_t} + (1 - \gamma)F_{t-s} \quad (2.11)$$

Com $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ e $0 \leq \gamma \leq 1$, sendo α , β e γ constantes.

2.7 Modelagem Box-Jenkins

Também conhecida como modelo Auto Regressivo Integrado de Médias Móveis (ARIMA), a modelagem Box-Jenkins é baseada em um ciclo iterativo, no qual a escolha do modelo é feita com base nos próprios dados, os estágios do ciclo são:

- **Identificação:** Nesta etapa encontram-se os filtros ARIMA e a ordem do modelo, representados pelos valores de p, d, e q do modelo;
- **Estimação:** Em seguida, estimam-se os parâmetros dos termos autorregressivos e de médias móveis que serão incluídos no modelo, por meio do Método de Mínimos Quadrados – MMQ;
- **Verificação:** Nesta fase verifica-se o modelo escolhido para ajustar adequadamente aos dados, isto é, analisar se os resíduos estimados se comportam como um ruído branco, em caso positivo, passar para a quarta etapa, e em caso negativo, deve-se recomeçar;
- **Previsão:** Esta é a fase final do processo, em que é estimada a previsão para a série temporal, em geral para um horizonte de curto prazo.

Caso o modelo não seja adequado o ciclo é repetido, voltando-se à fase de identificação. Um procedimento muito utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. Quando se obtém um modelo satisfatório, passa-se para a última etapa da metodologia de Box-Jenkins, que constitui o objetivo principal da metodologia: realizar previsões. O detalhamento dessas etapas se dará com a aplicação no estudo de caso (WERNER; RIBEIRO, 2003).

2.7.1 Modelo autorregressivo (AR)

Segundo Box (2015), um modelo autorregressivo de ordem p , denotado por $AR(p)$, a série Y_t é explicada pela média ponderada dos seus valores prévios e pelo ruído branco, um erro aleatório, tal que:

$$Y_t = e + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t, \quad (2.12)$$

onde u_t é um ruído branco.

Esse modelo pressupõe que o valor atual da série é uma combinação linear dos p valores passados da série e de um ruído branco u_t . A função de autocorrelação de um processo autorregressivo é constituída de uma mistura de polinômios, exponenciais e senóides amortecidas e é infinita em extensão. (BOX, 2015).

2.7.2 Modelo autorregressivo (MA)

O processo de médias móveis de ordem q é denotado por $MA(q)$ tal que, para um processo de média zero:

$$Y_t = e + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}, \quad (2.13)$$

onde u_t é um ruído branco.

Utiliza-se o modelo de médias móveis na análise de séries temporais para a modelagem de modelos das séries univariadas no tempo. Consequentemente, como a série Y_t é construída com base na combinação linear de seus erros aleatórios passados para gerar o valor mais recente da série, nesse modelo o valor atual da série é uma média ponderada dele próprio mais os q últimos valores de um processo ruído branco. (BOX, 2015).

2.7.3 Modelo autorregressivo de médias móveis (ARMA)

Para muitas séries encontradas na prática, se o objetivo for um modelo com um número reduzido de parâmetros, isto é, mais parcimonioso, a inclusão de termos tanto autorregressivos quanto de médias móveis é adequada (MORETTIN; TOLOI, 2004). O modelo $ARMA(p,q)$ pode ser escrito da seguinte forma:

$$Y_t = e + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}, \quad (2.14)$$

sendo os ϕ 's parâmetros autorregressivos e os θ 's parâmetros das médias móveis, respectivamente.

Tal modelagem é uma combinação linear da média ponderada de seus p valores prévios de sua série juntamente com seu ruído branco no tempo presente e nos q tempos passados (BOX, 2015).

2.7.4 Modelos autorregressivos integrados de médias móveis (ARIMA)

O modelo ARIMA (p,d,q) é um caso especial de processo integrado no sentido de que supõe que a d-ésima diferença de uma série não estacionária pode ser representada por um modelo ARMA(p, q), estacionário (SANTOS, 2012).

A diferença é aplicada na seguinte equação:

$$w_t = Y_t - Y_{t-1} = (1 - B)Y_t = \Delta Y_t. \quad (2.15)$$

Já para as diferenças de ordem superior ($d > 1$), temos o seguinte procedimento:

$$w_t = \Delta^d Y_t = \Delta^{d-1}(\Delta^1 Y_t), \quad (2.16)$$

onde Δ^d é o operador de defasagem.

2.7.5 Modelos autorregressivos de médias móveis sazonal (SARIMA)

Os modelos ARIMA sazonal (SARIMA), de ordem $(p, d, q) \times (P, D, Q)$ são constituídos por uma parte não-sazonal (p, d, q) e outra sazonal (P, D, Q) (SANTILI, 2015) onde:

- **P**: é o número de coeficientes sazonais autorregressivos referentes a estacionariedade.
- **D**: o número de diferenças sazonais, para tornar a série estacionária de acordo com os períodos.
- **Q**: o número de coeficientes sazonais de médias móveis referentes a invertibilidade.

O modelo pode ser escrito como

$$\phi(B)\Phi(B)[(1 - B)_d(1 - B)_D - \alpha]Y_t = \theta(B)\Theta(B)u_s, \quad (2.17)$$

onde s é um período de sazonalidade.

Esses modelos são os mais requisitados para descrição de séries temporais sazonais, demonstrando sucesso em suas aplicações.

2.7.6 Modelos Autorregressivos de Médias Móveis Sazonal com Variáveis Exógenas (SARIMAX)

O modelo SARIMAX (p, d, q, s) é uma variação sazonal de um modelo de regressão linear com adição de uma ou mais variáveis exógenas que é uma variável explicativa não

correlacionada com a perturbação aplicada ao modelo SARIMA (P, D, Q) . E a equação do modelo é dada por:

$$\phi(B)\Phi(B)(1-B)^d(1-B)^DY_t = \alpha + X\beta + \theta(B)\Theta(B^s)u_s, \quad (2.18)$$

sendo X a matriz $n \times k$ de regressores não aleatórios e β um vetor de parâmetros $k \times 1$.

2.7.7 Função de autocorrelação (FAC)

O Função de autocorrelação mede a previsibilidade linear da série no tempo t , digamos y_t , usando apenas o valor y_k . Podemos mostrar facilmente que $-1 < \rho_k < 1$ usando o Cauchy – Schwarz desigualdade. Se pudermos prever x_t perfeitamente a partir de x_k por meio de uma relação linear, $x_t = \beta_0 + \beta_1 x_k$, então a correlação será $+1$ quando $\beta_1 > 0$ e -1 quando $\beta_1 < 0$. Portanto, temos uma medida aproximada da capacidade de prever a série no tempo t a partir de o valor no tempo k . Muitas vezes, gostaríamos de medir a previsibilidade de outra série y_t do série x_k consultar (SHUMWAY; STOFFER, 2017).

A função de autocorrelação (FAC) é definida como:

$$\rho_k = \frac{Cov[y_t, y_{t-k}]}{Var[y_t]} = \frac{\gamma_k}{\gamma_0}. \quad (2.19)$$

Se temos um coeficiente de correlação ρ_1 , então verificamos a correlação do período t em relação ao período de defasagem $t - 1$, na qual é chamada de coeficiente de lag 1 (PELLEGRINI, 2000).

2.7.8 Função de autocorrelação parcial (FACP)

A função de autocorrelação parcial utilizada para facilitar o processo de identificação do modelo, denotada por ϕ_{kk} , analisa a correlação entre os pontos y_t e y_{t-k} , limitando a influência que $y_{t-1}, \dots, y_{t-k+1}$ exerce sobre y_t .

$$\left\{ \begin{array}{l} y_t = \phi_{11}y_{t-1} + u_t \\ y_t = \phi_{11}y_{t-1} + \phi_{22}y_{t-2} + u_t \\ \vdots \\ y_t = \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \dots + \phi_{kk}y_{t-k} + u_t \end{array} \right. \quad (2.20)$$

2.7.9 Etapas do modelo SARIMA

Na modelagem SARIMA tem-se três etapas: Identificação, Estimação e Verificação.

Na Etapa Identificação baseia-se em descobrir qual dentre as várias versões dos modelos de Box-Jenkins, sejam eles sazonais ou não, descreve o comportamento da série. Segundo

Morettin e Tolo (2004) a identificação é feita em dois passos:

1 - Obter estimativas preliminares α , β_j de μ , α_j e β_j onde $j = 1, \dots, 6$ por meio da análise de regressão de Y_t sobre 1 , $\text{sen} \frac{2}{12}$ e $\text{cos} \frac{2}{12}$, sendo $j = 1, \dots, 6$,

2 - Calcula os resíduos

$$N_t = Y_t - \mu - \sum_{j=1}^6 \left[\alpha_j \text{cos} \frac{2}{12} + \beta_j \text{sen} \frac{2}{12} \right] \quad (2.21)$$

e identificar as funções de autocorrelação e funções de autocorrelação parcial para verificar um modelo ARMA (p,q) para N_t .

Esta etapa compreende um conjunto de procedimentos a fim de proporcionar a determinação dos parâmetros. Inicia-se determinando o parâmetro de integração, parâmetro d , uma ideia aproximada da estrutura do modelo, sabe que para encontrar o d , tem que aplicar algum teste de estacionariedade com o objetivo de encontrar a ordem de integração. E este parâmetro pode assumir, normalmente, os valores $d = 0$, quando a série é estacionária e $d = 1$, ou quando a série não é estacionária, possuindo uma raiz unitária, mas, raramente o valor $d = 2$.

Teste de Estacionariedade de Dickey-Fuller

O teste Dickey-Fuller é utilizado para avaliar a presença de raiz unitária de uma série temporal.

$$\Delta Y_t = \gamma Y_{t-1} + u_t, \quad (2.22)$$

onde $\Delta = (Y_t - Y_{t-1})$ é o operador de diferença e $\gamma = \rho - 1$.

As hipóteses confrontadas no teste de Dickey-Fuller são:

- $H_0 : \gamma = 0$, existe pelo menos uma raiz unitária, não estacionária.
- $H_1 : \gamma < 0$, não te raiz unitária, e é fracamente estacionária.

Mas vamos utilizar a expressão alternativa do teste, onde se $\rho = 1$ a série possui uma raiz unitária e não estacionária, no entanto, se $|\rho| < 1$ a série não possui raiz unitária e é estacionária.

$$Y_t = \rho Y_{t-1} + u_t \quad (2.23)$$

onde u_t é o ruído branco.

Na Etapa Estimação baseia-se em estimar parâmetros ϕ e Φ do componente auto regressivo, os parâmetro θ e Θ do componente de média móvel e variância de e_i . Segundo Morettin e Tolo (2004), pode-se usar o procedimento iterativo de Mínimos Quadrados Condicionais

ou o Método de Máxima Verossimilhança para identificar um modelo provisório para a série e estimar seus parâmetros.

A função de máxima verossimilhança é dada por:

$$L(\phi_i, \theta_j, \sigma_\xi^2/w) = (2\pi\sigma_\xi^2)^{\frac{n}{2}} \left| \sum \right|^{\frac{1}{2}} e^{-\frac{1}{2}x' \Sigma^{-1} w}{\sigma_\xi^2} \quad (2.24)$$

Quando a função é maximizada tem-se as estimativas dos parâmetros ϕ_i , θ_j e σ_ξ^2 .

Na Etapa validação Utiliza-se a previsão para estimar o futuro valor da série temporal e para gerar intervalos de confiança para identificação do modelo ARIMA produzido pela etapa da estimação. Neste passo o objetivo é verificar se o modelo apresenta alguma inadequação e, assim, alcançar a melhoria do modelo. Segundo Box (2015) a partir do teste de Ljung-Box, pode-se verificar que os coeficientes de autocorrelação dos resíduos são estatisticamente iguais a zero. Entretanto, se faz necessário a verificação menos característica utilizada pelos resíduos do modelo que foi ajustado.

Os critérios a seguir são úteis na escolha do modelo mais acurado. Vale resaltar que por tenderem a selecionar modelos superparametrizados assintoticamente, os critérios de AIC não são consistentes. Entretanto, tem-se o BIC que é um critério bastante rigoroso, consequentemente consistente para a inclusão de parâmetros do modelo.

- FPE - Erro de Previsão Final

$$FPE = \sigma^2 x \frac{T+p}{T-p} \quad (2.25)$$

- AIC - Critério de Informação de Akaike

$$AIC = 2\log\hat{L} + 2(p+q), \quad (2.26)$$

em que L é o valor maximizado da verossimilhança.

- BIC - Critério de Informação Bayesiano

$$BIC = 2\log\hat{L} + (p+q)\log T \quad (2.27)$$

2.8 Avaliação de Previsão

Para verificar se as previsões estão com os valores próximos dos dados reais deve-se fazer a medição do erro, que neste caso é basicamente $[Y]_j$ sendo a previsão de Y_j .

$$\epsilon_t = Y_j + [Y]_j \quad (2.28)$$

As métricas a seguir são utilizadas para avaliar os erros de previsão (FARIAS, 2017).

- MPE = Erro Percentual Médio

$$MPE = \left(\frac{1}{n} \sum_{j=k+1}^M \frac{\varepsilon_j}{Y_t} \right) X 100 \quad (2.29)$$

- MSE = Erro Quadrado Médio

$$MSE = \frac{1}{n} \sum_{j=k+1}^M (\varepsilon_j)^2 \quad (2.30)$$

- MAE = Erro Absoluto Médio

$$MAE = \frac{1}{n} \sum_{j=k+1}^M |\varepsilon_j| \quad (2.31)$$

- ET = Erro Total

$$ET = \sum_{j=k+1}^M \varepsilon_j \quad (2.32)$$

- TAE = Erro Absoluto Total

$$TAE = \sum_{j=k+1}^M |\varepsilon_j| \quad (2.33)$$

- TPE = Erro Percentual Total

$$TPE = \left(\frac{\sum_{j=k+1}^M \varepsilon_j}{\sum_{j=k+1}^M Y_t} \right) X 100 \quad (2.34)$$

3 APLICAÇÃO

Os dados utilizados neste trabalho são provenientes da ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis) referentes importação de gás natural por mil m^3 no Brasil, de janeiro de 2000 a maio de 2021, estando disponível no Portal Brasileiro de Dados Abertos. A análise foi feita com auxílio do *software R* (TEAM, 2021), tais dados são compostos de 2 variáveis (meses e valor) e 257 observações. Será analisado o comportamento da série do valor mensal das importações de gás natural para o Brasil ao longo dos anos e fazer a previsão para os próximos meses.

Tabela 1 – Análises descritivas dos dados

Estatísticas	Valor (mil m^3)
Mínimo	101942
Máximo	2318656
1° Quadrante	624348
3° Quadrante	1063879
Média	866231
Mediana	822392
LCI Média	819215
LCS Média	9132473
Variância	1464924
Desvio Padrão	3827432
Assimetria	630000
Curtose	550000

Por meio da Tabela 1 ao aplicar o *summary* nos dados observa-se que o valor médio das importações foi de 866231 mil m^3 , apresentando mediana de 822392 mil m^3 , já seus valores extremos foram de 101942 mil m^3 em janeiro de 2000 sendo o valor mínimo e 2318656 milhões m^3 em janeiro de 2015 o valor máximo.

Na Figura 3 tem-se o gráfico da série em estudo, onde é possível perceber um pico no ano de 2015, onde obteve-se o maior valor já registrado. A transformação de dados é uma das possíveis formas de contornar o problema de dados que não obedecem os pressupostos da análise de variância. Na Figura 4 tem-se uma transformação logaritma na série. A série apresenta uma tendência de crescimento, e de acordo com o teste de Teste de Dickey-Fuller o p-valor = 0.08942, sendo superior a 0,01, representando que há uma evidência que a série tem um comportamento não estacionário.

Figura 3 – Gráfico da série temporal de importação de gás natural para o período de jan/2000 a mai/2021.

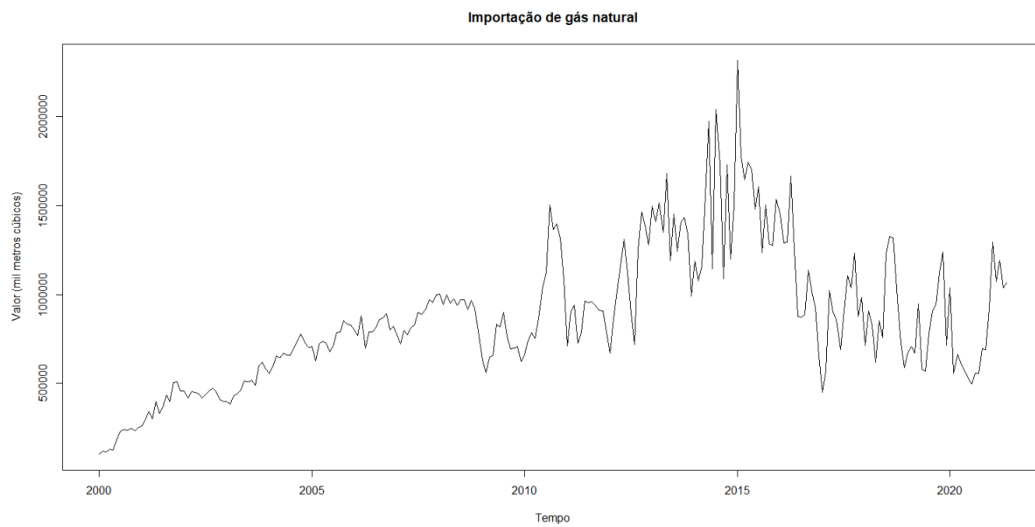
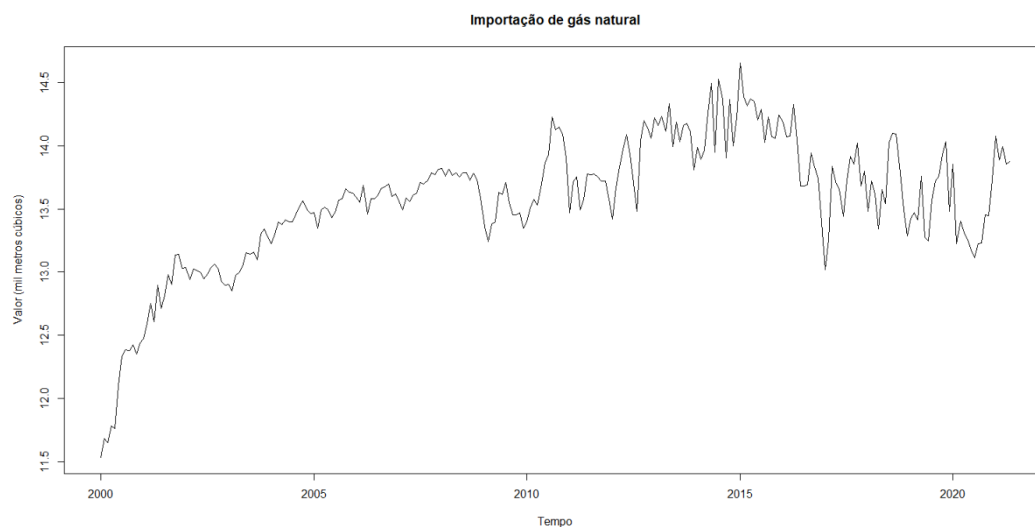
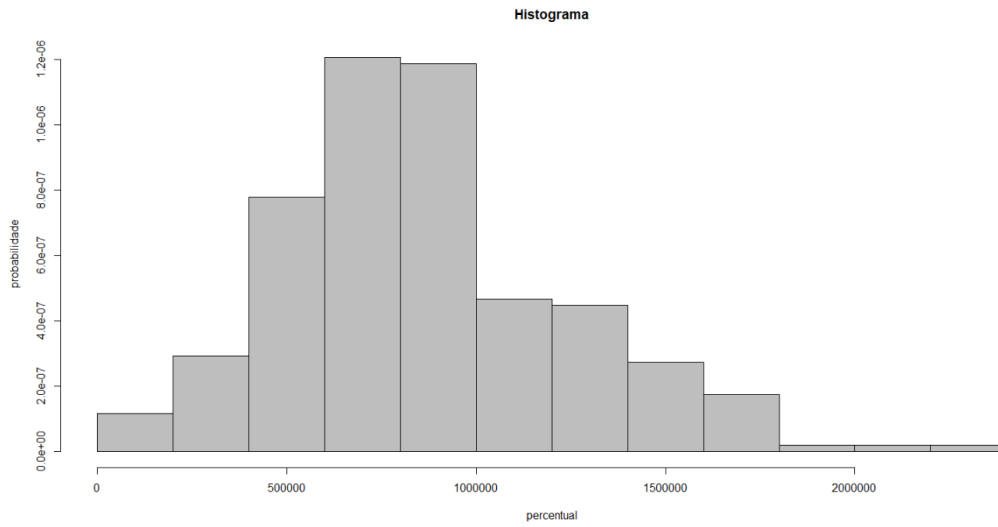


Figura 4 – Gráfico da série, com transformação logarítmica dos valores de importação por m^3 no Brasil janeiro de 2000 a maio de 2021.



Na Figura 5 apresenta-se o histograma da série em estudo, onde foi possível observar o comportamento assimétrico à direita, evidenciando a não normalidade dos dados. Tal resultado pôde ser validado por meio do teste de Shapiro-Wilk que apresentou $p\text{-valor}=0,0001056$.

Figura 5 – Histograma da serie temporal da importação do gás natural



Na Figura 6, a autocorrelação da série (ACF) apresenta um lento decaimento para zero; tal característica indica indícios de uma série não estacionária. O gráfico de autocorrelação parcial da série (PACF) apresenta apenas os lags 1 e 2 significativos.

Figura 6 – Gráficos de autocorrelação amostral da série (ACF) e autocorrelação parcial da série (PACF)

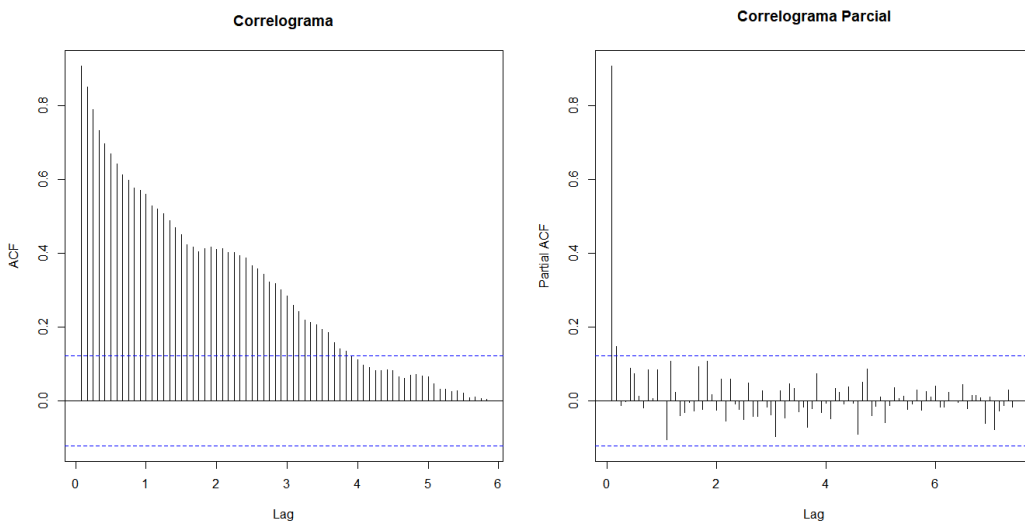
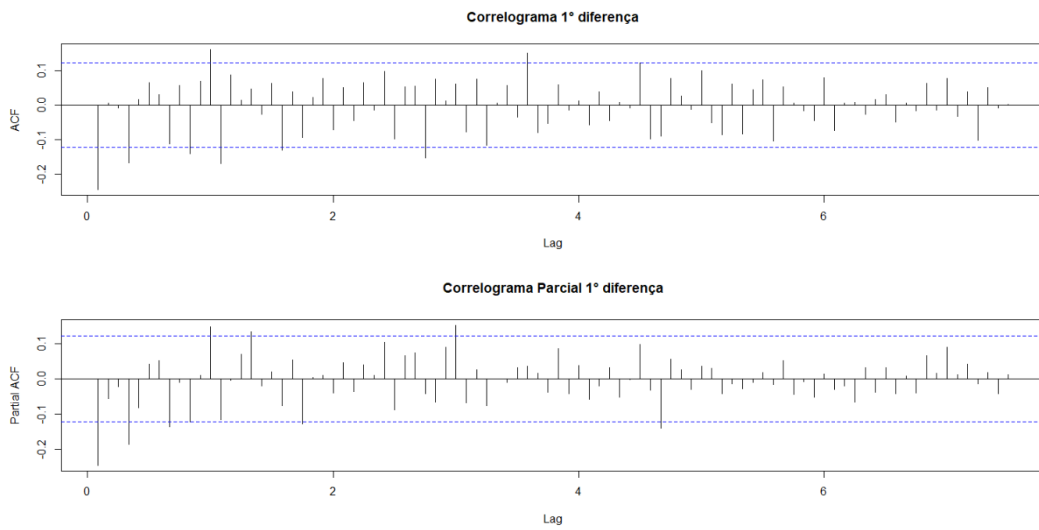
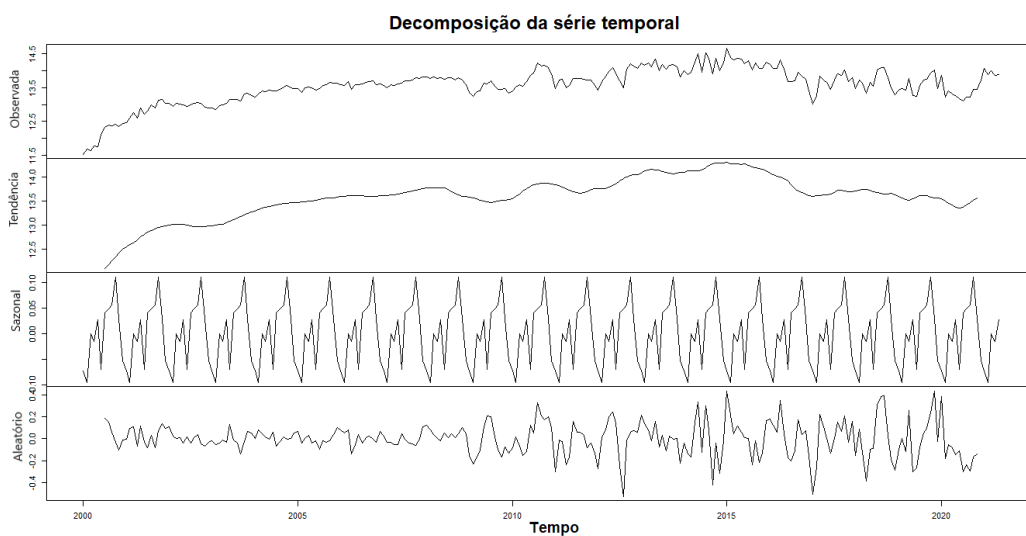


Figura 7 – Gráficos de autocorrelação amostral da série (ACF) e autocorrelação parcial da série (PACF), aplicando a primeira diferença.



Na Figura 7 aplicou-se o logaritmo nos dados dos devidos correlogramas, porém aplicando-se a primeira diferença ($d = 1$). Pode-se observar que os gráficos ACF e PACF apresentam alguns picos que ultrapassam os limites. Entretanto, no ACF da primeira diferença os picos que se repetem, sugerindo sazonalidade na série.

Figura 8 – Valores de importação de gás natural mil por m^3 no Brasil, dos períodos de janeiro de 2000 a maio de 2021



Para verificar a estacionariedade da série, novamente, aplicou-se a primeira diferença ($d = 1$) com os dados logaritmizados. A Figura 8, mostra que para o ACF e PACF, há picos que

ultrapassam os limites propostos. Porém, no correlograma da primeira diferença da série, nota-se que há picos de correlação, o que sugere uma leve sazonalidade na série.

3.1 Modelagem – Alisamento exponencial

Realizou-se a modelagem da série temporal da importação do gás natural no período de jan/2020 a maio/2021 via Alisamento Exponencial de Holt-Winters. Considerou-se previsões de 1, 4 e 7 passos a frente a fim de avaliar as capacidades preditivas do método de Holt-Winters em diferentes horizontes de previsão, no intuito de ver o comportamento da valor de importação no 3º e 4º trimestre de 2021 através do método aditivo e multiplicativo.

Primeiramente, faz-se importante a escolha dos parâmetros de suavização adequados para buscar ao máximo, minimizar a soma dos quadrados dos erros de previsão um passo à frente. Que estão dispostos na Tabela 2:

Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e multiplicativo.

MÉTODO	α	β	γ
HW Aditivo	0,722	0,0214	0,453
HW Multiplicativo	0,718	0,0214	0,491

As escolhas dos parâmetros de suavização são bem importantes para a previsão, pois ajudam a minimizar a soma dos quadrados dos erros. Como o valor de β aditivo e multiplicativo foram os que mais se aproximaram de 0, nos mostrando que ao longo do tempo existe um decréscimo lento das previsões dos valores de importação de gás natural. Os valores de previsão encontrado através dos algoritmos de Holt-Winters, um passo à frente foram de 975554.5 mil m^3 e 976606.9 mil m^3 para o aditivo com erro de 0,62% e o multiplicativo com erro de 0,61%, respectivamente. A essa previsão trouxe valores abaixo do valor observado, ou seja, uma subprevisão.

Na Figura 9 nota-se que os dados da série na pela linha preta, e seu ajuste por Holt-Winters de vermelho, e percebemos que de certa forma os dados previstos ficaram bem ajustados aos dados observados. Utilizando os mesmos métodos e modelos, realizou-se uma previsão da série com 1, 4 e 7 passos a frente, ou seja, até dezembro de 2021.

Figura 9 – Gráficos dos dados da série e seu ajuste de Holt-Winters.

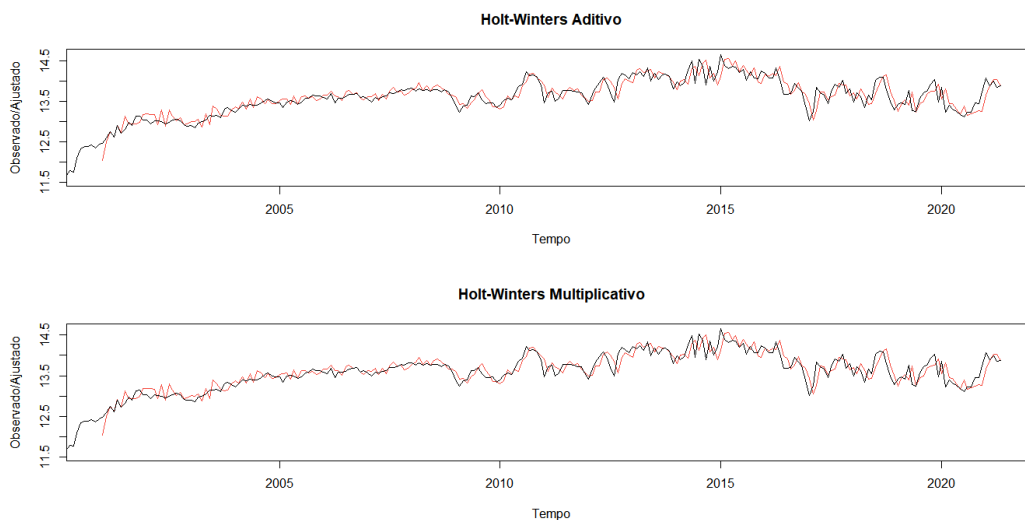


Tabela 3 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m^3))

MÊS/2021	Aditivo	Multiplicativo
JUNHO	975554,5	976606,9

Tabela 4 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m^3)))

MÊS/2021	Aditivo	Multiplicativo
JUNHO	975554,5	976606,9
JULHO	1183320,4	1191293,4
AGOSTO	1213631,9	1230987,9
SETEMBRO	1222909,9	1237647,1

Tabela 5 – Previsões intervalares do algoritmo Holt-Winters Aditivo e Multiplicativo (Importação de gás natural(mil/m³)))

MÊS/2021	Aditivo	Multiplicativo
JUNHO	975554,5	976606,9
JULHO	1183320,4	1191293,4
AGOSTO	1213631,9	1230987,9
SETEMBRO	1222909,9	1237647,1
OUTUBRO	1288722,6	1312188,2
NOVEMBRO	1146464,2	1168190,9
DEZEMBRO	1043786,5	1060843,8

Nas previsões com 1, 4 e 7 passos a frente estão dispostos nas Tabelas 3, 4 e 5, onde mostra os modelos Holt-Winters aditivo e multiplicativo, pode-se notar que as previsões obtiveram valores bem próximos e que até outubro houve um aumento, já nos meses seguintes os valores começam a baixar, sugerindo uma leve sazonalidade dos dados.

Figura 10 – Gráficos da previsão 1 passos a frente.

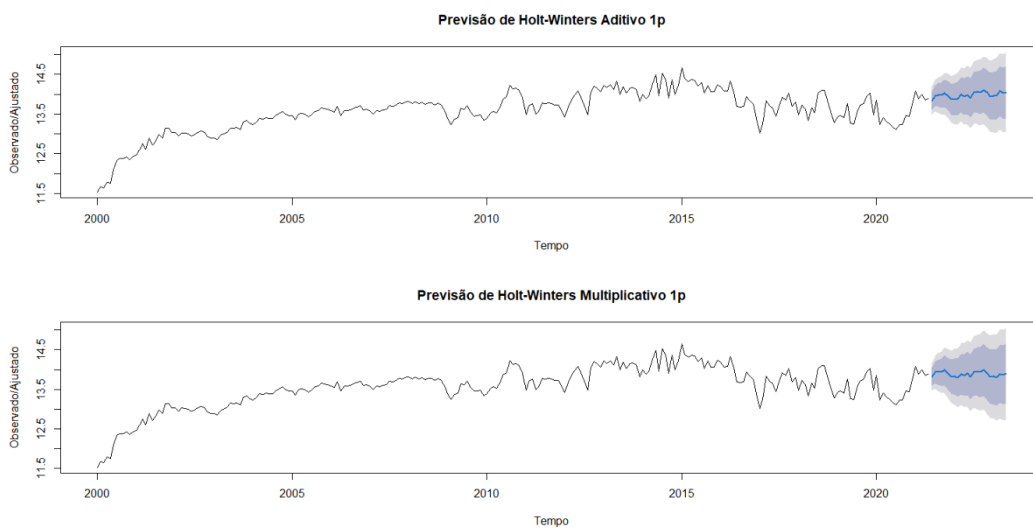


Figura 11 – Gráficos da previsão 4 passos a frente.

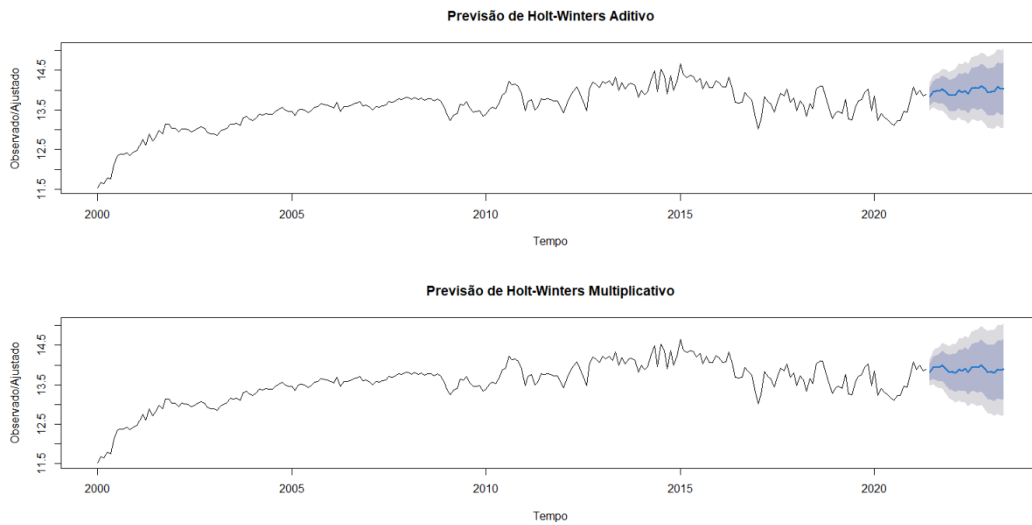
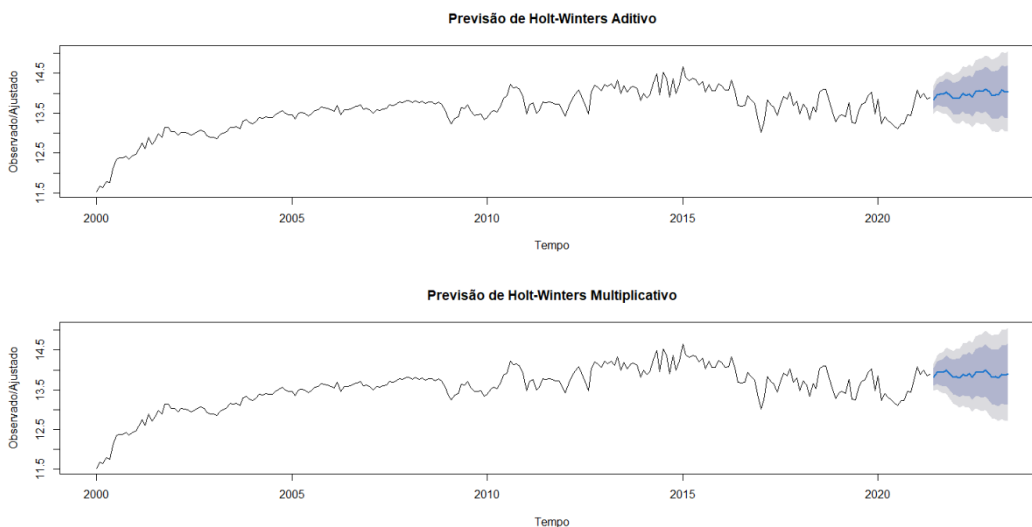


Figura 12 – Gráficos da previsão 7 passos a frente.



As sombras ao redor dos valores médios previstos são os intervalos de confiança: faixas de valores estimados onde a média tem uma dada probabilidade de ocorrer. Por padrão, os intervalos são de 80% e de 95% – existe 80% e 95% de probabilidade que seu intervalo contenha o valor correto. Nas Figuras 10, 11 e 12 demonstram que a tendência do valores de importação de gás natural tende a continuar e apresentam uma previsão contínua.

Na sequência é apresentado um teste de precisão da previsão ajustada com base nos dados da série, a mesma foi implementada na função *ets* no pacote *forecast*, no *software R* (TEAM, 2021), que são definidas em Hyndman e Khandakar (2008). Na Tabela 6 observa-se que o teste

de previsão resultou em erros baixos, indicando uma boa capacidade de previsão para o modelo.

Tabela 6 – Valores dos erros de previsão.

ME	RMSE	MAE	MPE	MAPE	MASE
0,002355549	0,1747921	0,1246954	0,006876941	0,9147988	0,4417601

3.2 Modelagem Box-Jekins

Para a modelagem Box-jekins foi utilizado a função *auto.arima*, disponível no pacote *forecast* implementado no *software R* (TEAM, 2021), para verificar a melhor previsão através dos modelos SARIMA e SARIMAX, com o modelo de ordem $(1, 1, 1) \times (1, 0, 0)_{12}$. Neste trabalho a modelagem SARIMA tem a finalidade de achar a previsão mais adequada para os dados analisados. Em seguida encontramos sua previsão de 975554,5 mil m^3 e 976606,9 mil m^3 e seus erros de 0,62% e 0,61%, respectivamente. (Como foi dito acima, o valor predito foi abaixo do observado.)

A variável *dummy* foi aplicada a partir de janeiro de 2020 até maio de 2021 (período onde iniciou os primeiros casos de Covid-19 em outros países fora da China) com o valor de 1 e com o valor de 0 para os meses anteriores, para assim prever os diferentes valores com os dados de importações no Brasil. O modelo que foi utilizado foi um SARIMA $(1, 1, 1) \times (1, 0, 0)_{12}$, tivemos a previsão de aproximadamente 975554,5 e seu erro de 0,42%, no caso do SARIMAX $(1, 1, 1) \times (1, 0, 0)_{12}$ obteve a previsão de 976606,9, mas o erro foi de 0,41%, fazendo a comparação dos modelos, podemos observar que houve uma diferença mínima entre eles.

Para a seleção do modelo foi calculado os critérios de informação AIC (Critério de informação Akaike) e o BIC (Critério de informação Baysiano) que são baseados no SQE (Soma dos Quadrados dos Erros). A partir da Tabela 7 observa-se os valores referentes aos critérios AIC e BIC, bem como, os erros relativos dos modelos:

Tabela 7 – Valores dos critério de informação dos modelos da série e erros relativos, 1 passo à frente.

MODELO	AIC	BIC	ERRO DE PREVISÃO(%)
SARIMA $(1, 1, 1) \times (1, 0, 0)_{12}$	-158,60	-144,42	0,42
SARIMAX $(1, 1, 1) \times (1, 0, 0)_{12}$	-165,49	-147,76	0,41

De acordo com os critérios de informação o modelo que tem o melhor ajuste é o que apresenta os menores valores, que no caso é o modelo SARIMAX. Na Tabela 8 tem-se as previsões para 4 passos a frente: junho-set/2021. Na Tabela 9 tem-se as previsões para 7 passos a frente: junho-dez/2021. Os valores das previsões foram próximos para ambos os modelos SARIMA e SARIMAX. Utilizou-se o *software R* (TEAM, 2021) para os modelos previstos pela função *auto.arima*, observando qual modelo fornecerá a melhor previsão.

Tabela 8 – Previsões dos modelos SARIMA e SARIMAX estudados, 4 passo à frente.

	MODELO	Jun/21	Jul/21	Ago/21	Set/21
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	1026454	1005491	1018579	1016482
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	1027833	1006806	1021927	1019915

Tabela 9 – Previsões dos modelos SARIMA e SARIMAX estudados, 7 passo à frente.

	MODELO	Jun/21	Jul/21	Ago/21	Set/21
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	1026454	1005491	1018579	1016482
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	1027833	1006806	1021927	1019915
	MODELO	Out/21	Nov/21	Dez/21	
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	1052426	1050858	1099558	
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	1059597	1057865	1111642	

Nas Tabelas 10 e 11 com erros de previsões com 4 e 7 passos a frente, observou-se que os erros com a variável *dummy* no modelo SARIMAX com 7 passos a frente apresentaram um melhor comportamento, do que com 4 passos a frente.

Tabela 10 – Erros relativos de previsão dos modelos 4 passos a frente

	MODELO	Jun/21	Jul/21	Ago/21	Set/21
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	0,42	0,57	0,48	0,49
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	0,41	0,56	0,45	0,47

Tabela 11 – Erros relativos de previsão dos modelos 7 passos a frente

	MODELO	Jun/21	Jul/21	Ago/21	Set/21
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	0,42	0,57	0,48	0,49
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	0,41	0,56	0,45	0,47
	MODELO	Out/21	Nov/21	Dez/21	
1	SARIMA (1, 1, 1)x(1, 0, 0) ₁₂	0,24	0,25	-0,07	
2	SARIMAX (1, 1, 1)x(1, 0, 0) ₁₂	0,19	0,20	-0,15	

As Figuras a seguir apresentam os gráficos do Teste de Ljung-Box para validação dos modelos SARIMA e SARIMAX com 4 e 7 passos a frente.

O modelo apresenta resíduos estacionários, com média zero e variância constante. Com relação ao ACF, tem-se que todos os lags (exceto 0, que sempre vai ser igual a 1) tenham valores dentro do intervalo de confiança, já os p-valores do teste de Ljung-Box estão todos acima de 5%, indicando que não se rejeita a hipótese nula de que as n primeiras auto-correlações dos erros são iguais a zero. assim, o modelo proposto mostrou-se adequado.

Nas Figuras abaixo apresentaram a função de Autocorrelação residual dos modelos ajustados SARIMA e SARIMAX dos dados de importação com 4 e 7 passos a frente. Estes resultados evidenciam que aparentemente os coeficientes de autocorrelação ficaram dentro dos limites de confiabilidade.

Figura 13 – Gráficos de diagnóstico dos resíduos SARIMA com 4 passos a frente

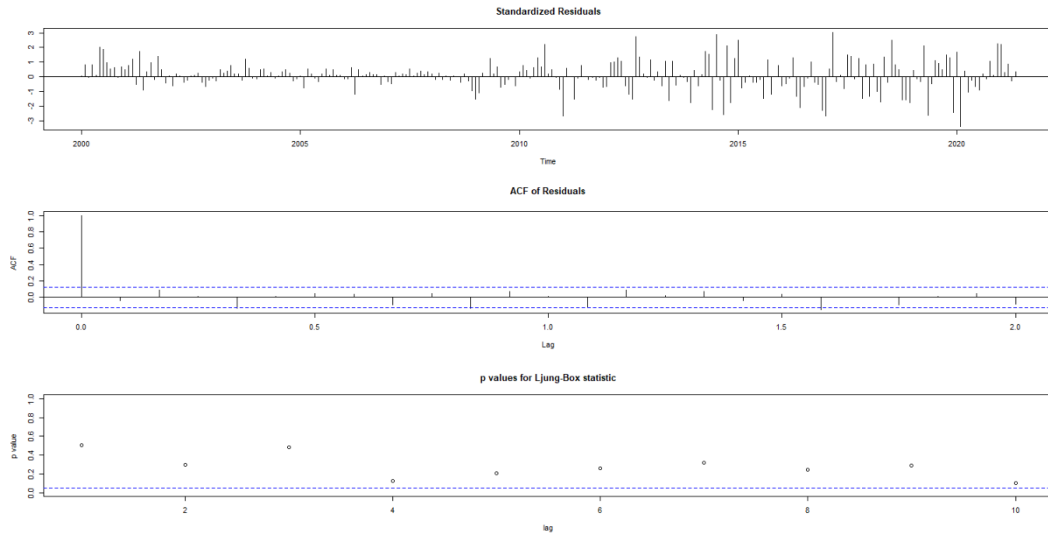


Figura 14 – Gráficos de diagnóstico dos resíduos SARIMAX com 4 passos a frente

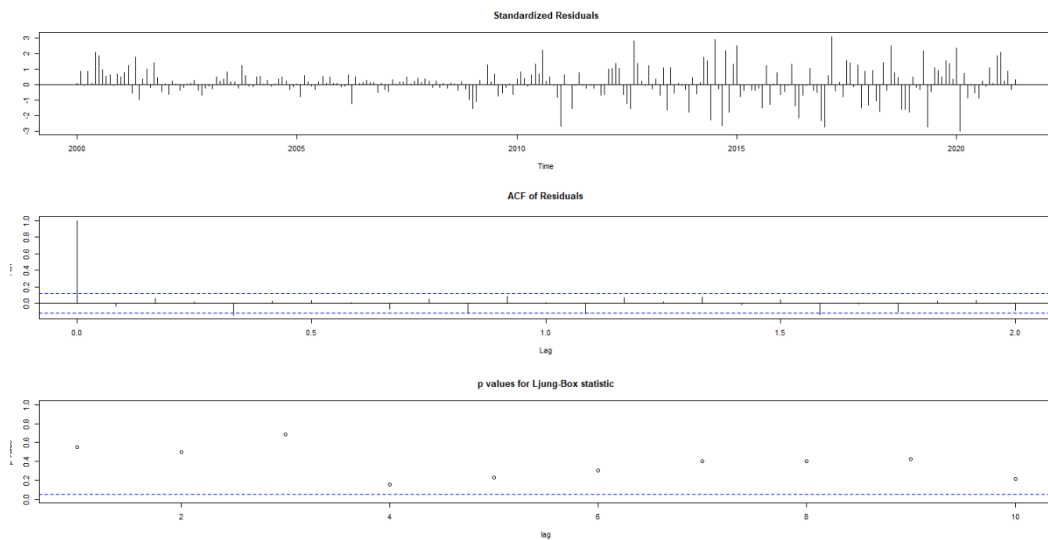


Figura 15 – Gráficos de diagnóstico dos resíduos SARIMA com 7 passos a frente

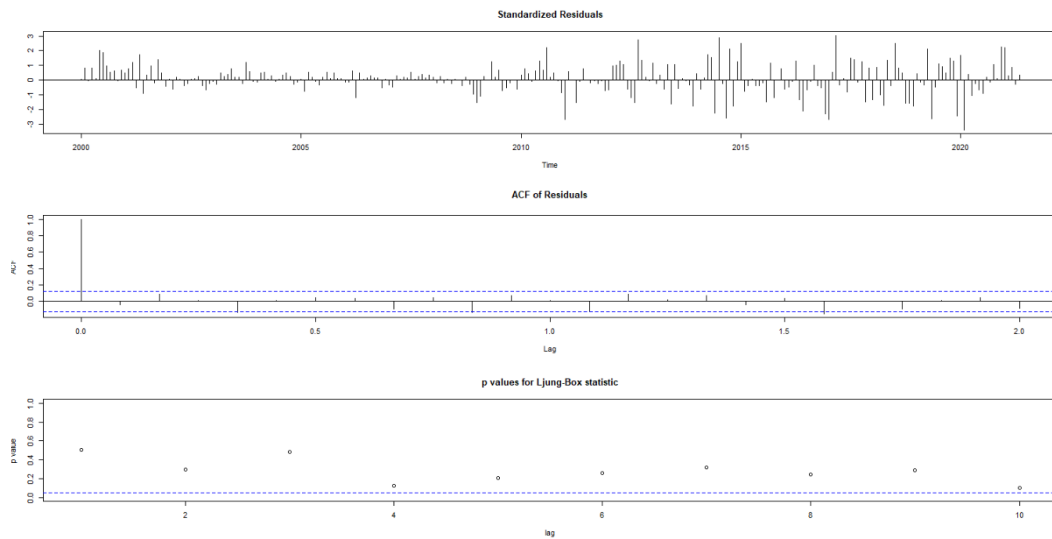
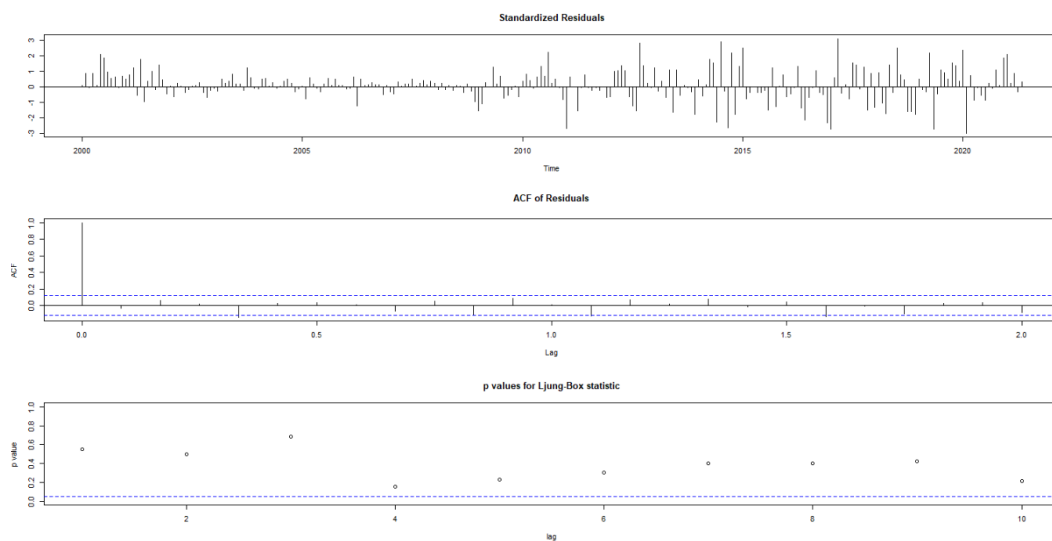


Figura 16 – Gráficos de diagnóstico dos resíduos SARIMAX com 7 passos a frente



Partindo-se para interpretação das Figuras 13, 14, 15 e 16, em relação aos p-valores do Teste de Ljung-Box apresentados graficamente, o modelo SARIMAX de 4 e 7 passos a frente apresentaram um melhor desempenho, onde todos os valores estão um pouco mais acima do limite de 5%, indicando que não rejeita a hipótese nula de que as n primeiras autocorrelações dos erros são iguais a zero. Desta maneira, conclui-se que os modelos descrevem adequadamente os dados, validando o seu uso para traçar previsões.

4 CONCLUSÃO

Neste trabalho foi descrito o comportamento das séries temporais referentes ao valor de importação de gás natural no Brasil de janeiro de 2000 até maio de 2021, buscando um modelo que possibilitasse uma boa previsão. Foram utilizados os algoritmos de Holt-Winters (Aditivo e Multiplicativo) e os modelos SARIMA e SARIMAX, onde neste último foi acrescentado uma variável *dummy*, com os valores de 1 para janeiro de 2020 (período onde iniciou os primeiros casos de Covid-19 em outros países fora da China) ao último mês da série, e 0 para os valores anteriores.

Com previsões de 1, 4 e 7 passos a frente e seus respectivos erros, foi feita a modelagem da série com o método Box-Jenkins ajustando o modelo SARIMA $(1, 1, 1) \times (1, 0, 0)_{12}$ e SARIMAX $(1, 1, 1) \times (1, 0, 0)_{12}$, à modo geral os modelos obtiveram previsões bem aproximadas, entretanto, o melhor ajuste foi o modelo SARIMAX com o seu critério de informação (AIC) e erro mais baixo.

O modelo SARIMAX $(1, 1, 1) \times (1, 0, 0)_{12}$ obteve o melhor ajuste com relação aos erros das previsões com 7 passos a frente serem bem mais baixos. Assim observou-se que a chegada da Covid-19 em alguns países afetaram o preço do gás natural, que no primeiro trimestre de 2021 houve um reajuste de 39% do mesmo ano, o que acarretou para as previsões serem tão próximas e contínuas.

Como proposta futura, através deste modelo proposto, podemos abranger este estudo acerca do preço mensal do gás natural, possibilitando estabelecer um comparativo que permita averiguar o comportamento do valor de importação com o preço do gás aqui no Brasil.

REFERÊNCIAS

- ANP. *Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis*. 2020. Disponível em: <<http://www.anp.gov.br/arquivos/central-conteudos/anuario-estatistico/2020/anuario-2020.pdf>>. Acesso em: 18 de agosto de 2021. Citado na página 11.
- ANP. *Boletim Trimestral de Preços e Volumes de Combustíveis*. 2021. Disponível em: <<https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/boletins-anp/btpvc-1/2021/boletim-trimestral-sdc-9-1t21.pdf>>. Acesso em: 16 de julho de 2021. Citado na página 11.
- BOX, G. E. P. *Time Series Analysis*. [S.l.]: John Wiley e Sons, 2015. ISBN 1118675029. Citado 2 vezes nas páginas 21 e 25.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to Time Series and Forecasting*. [S.l.]: Springer International Publishing, 2016. Citado na página 17.
- BROWN, R. G. *Statistical forecasting for inventory control*. [S.l.]: McGraw/Hill, 1959. Citado na página 17.
- FARIAS, J. F. Uma aplicação de séries temporais na arrecadação do icms do estado da paraíba. 2017. Citado na página 25.
- HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, Elsevier, v. 20, n. 1, p. 5–10, 2004. Citado 3 vezes nas páginas 17, 18 e 19.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018. Citado 2 vezes nas páginas 17 e 18.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, v. 27, n. 1, p. 1–22, 2008. Citado na página 34.
- LAUREANO, F. A indústria de gás natural e as relações contratuais uma análise do caso brasileiro. *Rio de Janeiro: COPPE*, 2005. Citado na página 11.
- MOREIRA, R. J. S. Análise das características de ruído em séries temporais gps. *FC - Dissertações de Mestrado*, 2013. Citado 2 vezes nas páginas 15 e 16.
- MORETTIN, P. A.; TOLOI, C. M. d. C. *Análise de séries temporais*. [S.l.]: Edgard Blucher, 2004. Citado 5 vezes nas páginas 13, 14, 15, 21 e 24.
- PELLEGRINI, F. R. Metodologia para implementação de sistemas de previsão de demanda. 2000. Citado na página 23.
- PETROBRÁS. *Petrobrás*. 2019. Disponível em: <<https://petrobras.com.br/pt/>>. Citado na página 11.
- SANTILI, M. A. Métodos de previsão auto-regressivo aplicado a uma série de volume de produção de caminhões. 2015. Citado na página 22.
- SANTOS, L. H. P. D. Uso de modelos autorregressivos e gráficos de controle para monitorar volatilidade de ativos financeiros. 2012. Disponível em: <http://pro.poli.usp.br/wp-content/uploads/2013/04/TF_Luiz_Henrique_Paiffer_dos_Santos_6480477.pdf>. Acesso em: 20 de julho de 2021. Citado na página 22.

SHUMWAY, R. H.; STOFFER, D. S. *Time Series Analysis and Its Applications*. [S.l.]: Springer International Publishing, 2017. ISBN 3319524518. Citado 2 vezes nas páginas 13 e 23.

TEAM, R. C. *R: A language and environment for statistical computing (R Version 4.0.3, R Foundation for Statistical Computing, Vienna, Austria, 2020)*. 2021. Disponível em: <<https://www.r-project.org/>>. Acesso em: 12 de Agosto de 2021. Citado 3 vezes nas páginas 27, 34 e 35.

WERNER, L.; RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos box-jenkins na área de assistência técnica de computadores pessoais. *Gestão & Produção*, FapUNIFESP (SciELO), v. 10, n. 1, p. 47–67, apr 2003. Citado na página 20.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management science*, INFORMS, v. 6, n. 3, p. 324–342, 1960. Citado 2 vezes nas páginas 17 e 19.

ZHOU, J. et al. Pink noise: Effect on complexity synchronization of brain activity and sleep consolidation. *Journal of Theoretical Biology*, v. 306, p. 68–72, 2012. ISSN 0022-5193. Citado na página 15.