

Ministério da Educação
Universidade Federal do Rio Grande - FURG
Programa de Pós-Graduação em Modelagem Computacional

Análise de séries temporais das emissões de dióxido de carbono equivalente no
estado do Amazonas: uma abordagem com os Métodos de Suavização
Exponencial e de Box-Jenkins

Francisco Soares Lima Filho

Rio Grande
2024

Ministério da Educação
Universidade Federal do Rio Grande - FURG
Programa de Pós-Graduação em Modelagem Computacional

Análise de séries temporais das emissões de dióxido de carbono equivalente no
estado do Amazonas: uma abordagem com os Métodos de Suavização
Exponencial e de Box-Jenkins

Francisco Soares Lima Filho

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como parte dos requisitos para obtenção do Grau de Mestre. Área de concentração: Modelagem Computacional.

Orientador: Profa. Dra. Viviane Leite Dias de Mattos

Rio Grande

2024

Francisco Soares Lima Filho

“ANÁLISE DE SÉRIES TEMPORAIS DAS EMISSÕES DE DIÓXIDO DE CARBONO EQUIVALENTE NO ESTADO DO AMAZONAS: UMA ABORDAGEM COM OS MÉTODOS DE SUAVIZAÇÃO EXPONENCIAL E DE BOX-JENKINS”

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestre. Área de concentração: Modelagem Computacional.

Aprovada em: 15/04/2024.

BANCA EXAMINADORA

Documento assinado digitalmente
 **VIVIANE LEITE DIAS DE MATTOS**
Data: 16/03/2024 10:08:54-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Viviane Leite Dias de Mattos (Orientadora – FURG)

Documento assinado digitalmente
 **CATIA MARIA DOS SANTOS MACHADO**
Data: 16/03/2024 09:01:36-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Catia Maria dos Santos Machado (FURG)

Documento assinado digitalmente
 **TIAGO DA CRUZ ASMUS**
Data: 15/03/2024 15:41:19-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Tiago da Cruz Asmus (FURG)

Documento assinado digitalmente
 **ANDREA CRISTINA KONRATH**
Data: 15/03/2024 11:42:11-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Andrea Cristina Konrath (UFSC)

Rio Grande - RS
2024

Este trabalho é dedicado às minhas filhas 'Sol' e Ada.

Agradecimentos

Agradeço primeiramente a Deus, cuja graça me sustentou ao longo desta jornada acadêmica. À minha amada família, expresso profunda gratidão pelo apoio incondicional, compreensão e amor constantes. Vocês são a raiz do meu sucesso e a fonte inesgotável de inspiração, em especial minha Mãe Rita Maria e meu irmão Marcos Antônio.

Por fim, à Profa. Dra. Viviane Leite Dias de Mattos, minha orientadora, suas orientações, sabedoria e apoio foram cruciais para o sucesso desta pesquisa. Sua dedicação inspiradora moldou não apenas esta pesquisa, mas também meu crescimento acadêmico.

A todos, meu sincero agradecimento.

*“Ensinar não é transferir conhecimento,
mas criar as possibilidades para a sua própria produção
ou a sua construção.”
(Paulo Freire)*

Resumo

A pesquisa aborda a questão das emissões de gases de efeito estufa (GEE) no estado do Amazonas, com ênfase na modelagem das séries temporais de emissões de CO_2e . Considerando os impactos das mudanças climáticas globais e a necessidade de reduzir as emissões de GEE. O objetivo do estudo é identificar o mecanismo por trás das emissões de CO_2e no Amazonas e comparar os modelos de Suavização Exponencial e Box-Jenkins para avaliar sua eficácia na modelagem dessas séries. A justificativa ressalta a importância da pesquisa diante da urgência de ações para mitigar os efeitos das mudanças climáticas, enfatizando a contribuição científica e prática do estudo. A metodologia começa com a revisão da literatura e segue com um estudo de caso. A área de estudo foi o estado do Amazonas. Os dados das emissões de CO_2e foram analisados no período de 1990 a 2021, abrangendo diversas fontes de emissão, como produção industrial, tratamento de resíduos, produção de energia, agropecuária e mudança do uso da terra e floresta (MUTF). Para a análise das séries temporais, foram empregados métodos de Suavização Exponencial e metodologia de Box-Jenkins (ARIMA). Os modelos foram ajustados e testados usando diferentes técnicas estatísticas para garantir sua precisão e adequação aos dados. Os critérios de informação como AIC, AICc e BIC foram utilizados na seleção do modelo mais apropriado. Além disso, métricas como RMSE, MAE e MAPE foram aplicadas para avaliar o desempenho dos modelos em termos de precisão das previsões. Em cada uma das fontes investigadas os modelos gerados por ambos os métodos apresentaram previsões bem semelhantes.

Palavras-chave: Gases de efeito estufa. Previsões. Suavização Exponencial. Box-Jenkins.

Abstract

The research addresses the issue of greenhouse gas emissions (GHG) in the state of Amazonas, with an emphasis on modeling time series of CO₂e emissions. Considering the impacts of global climate change and the need to reduce GHG emissions, the study aims to identify the mechanism behind CO₂e emissions in the Amazonas region and compare Exponential Smoothing and Box-Jenkins models to assess their effectiveness in modeling these series. The justification underscores the importance of the research in light of the urgency of actions to mitigate the effects of climate change, emphasizing the scientific and practical contribution of the study. The methodology begins with a literature review and proceeds with a case study. The study area was the state of Amazonas. CO₂e emissions data were analyzed from 1990 to 2021, covering various emission sources such as industrial production, waste treatment, energy production, agriculture, and land use change and forestry (LUCF). For the analysis of time series, Exponential Smoothing methods and Box-Jenkins methodology (ARIMA) were employed. The models were fitted and tested using different statistical techniques to ensure their accuracy and suitability to the data. Information criteria such as AIC, AICc, and BIC were used in selecting the most appropriate model. Additionally, metrics such as RMSE, MAE, and MAPE were applied to evaluate the models' performance in terms of prediction accuracy. In each of the investigated sources, the models generated by both methods showed very similar forecasts.

Keywords: Greenhouse gases. Predictions. Exponential Smoothing. Box-Jenkins.

Lista de figuras

Figura 3.1	Mapa do Brasil com destaque para os limites do estado do Amazonas	36
Figura 3.2	Processo geral para previsão usando um modelo ARIMA	38
Figura 4.1	Emissões de CO_2e distribuído por categoria de origem no estado do Amazonas, 1990-2021	40
Figura 4.2	Emissões de CO_2e distribuído por categoria de origem no Brasil, 1990-2021	40
Figura 4.3	Séries temporais das emissões de CO_2e no Amazonas e no Brasil	42
Figura 4.4	Emissões de dióxido de carbono equivalente no estado do Amazonas	44
Figura 4.5	Os modelos mais indicados para realizar previsões com Suavização Exponencial	52
Figura 4.6	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para processos industriais	53
Figura 4.7	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para tratamento de resíduos . .	54
Figura 4.8	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para produção de energia	54
Figura 4.9	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para agropecuária	55
Figura 4.10	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para MUTF	55
Figura 4.11	Autocorrelação (ACF) e a autocorrelação parcial (PACF) para a série total	56
Figura 4.12	Os modelos ARIMA selecionados para realizar previsões	62
Figura 4.13	Comparação entre os modelos de Suavização Exponencial (ETS) e ARIMA	65
Figura A.1	Quebra de estrutura, teste Chow	74

Lista de tabelas

Tabela 2.1	Índice de conversão de CO_2e	18
Tabela 2.2	Classificação dos métodos de suavização método	22
Tabela 2.3	Fórmulas para cálculos recursivos e previsões pontuais	23
Tabela 2.4	Características dos estudos selecionados	33
Tabela 2.5	Características dos estudos selecionados	34
Tabela 4.1	Distribuição do total das emissões de CO_2e por categoria, 1990-2021	39
Tabela 4.2	Medidas descritivas das emissões de CO_2e do estado do Amazonas, 1990-2021	43
Tabela 4.3	Resultados dos testes de tendência e quebra estrutural	46
Tabela 4.4	Coeficientes dos modelos estimados pelo Método de Suavização Exponencial, 1990-2015	47
Tabela 4.5	Resultados dos critérios de informação e dos testes de autocorrelação, heterocedasticidade e normalidade dos resíduos dos modelos, 1990-2015	49
Tabela 4.6	Capacidade preditiva dos modelos de suavização exponencial para emissões de CO_2e	50
Tabela 4.7	Modelos mais indicados para previsão da tendência de cada uma das séries	51
Tabela 4.8	Número de diferenças necessárias para uma série estacionária	53
Tabela 4.9	Modelos mais parcimoniosos da categoria processos industrial, 1990-2015	57
Tabela 4.10	Modelos mais parcimoniosos da categoria tratamento de resíduos, 1990-2015	57
Tabela 4.11	Modelos mais parcimoniosos da categoria produção de energia	58
Tabela 4.12	Modelos mais parcimoniosos da categoria agropecuária, 1990-2015	59
Tabela 4.13	Modelos mais parcimoniosos da categoria MUTF	59
Tabela 4.14	Modelos mais parcimoniosos da categoria total, 1990-2015	60
Tabela 4.15	Capacidade preditiva dos modelos ARIMA para emissões de CO_2e	60
Tabela 4.16	Modelos mais indicados para previsão da tendência de cada uma das séries	61
Tabela 4.17	Modelos selecionados dos métodos de Suavização Exponencial (ETS)	63
Tabela 4.18	Modelos selecionados das técnicas ARIMA	64

Lista de abreviaturas e siglas

<i>ADF</i>	Augmented Dickey-Fuller
<i>AIC</i>	Critério de informação de Akaike
<i>AICc</i>	Critério de informação de Akaike corrigido
<i>AED</i>	Análise exploratória dos dados
<i>AR</i>	Componente autoregressivo
<i>ARIMA</i>	Modelos autoregressivos integrados de média móveis
<i>BIC</i>	Critério de Informação Bayesiana de Schwarz
<i>ETS</i>	ExponenTial Smoothing
<i>GEE</i>	Gases do efeito estufa
<i>GWF</i>	Fator de Aquecimento Global
<i>GWP</i>	Potencial de aquecimento global
<i>IBGE</i>	Instituto Brasileiro de Geografia e Estatística
<i>I</i>	Diferenciação
<i>IPI</i>	Imposto sobre Produtos Industrializados
<i>IPCC</i>	Painel Intergovernamental sobre Mudanças Climáticas
<i>KPSS</i>	Kwiatkowski, Phillips, Schmidt, Shin
<i>MMTCDE</i>	Milhões de toneladas métricas de equivalentes de dióxido de carbono
<i>MMTCDE</i>	Milhões de toneladas de equivalentes de dióxido de carbono
<i>MUTF</i>	Mudança do uso da terra e floresta
<i>PP</i>	Phillips-Perron
<i>RMSE</i>	Raiz do Erro Médio Quadrático
<i>SEEG</i>	Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa
<i>SES</i>	Suavização exponencial simples
<i>UNEP</i>	Programa das Nações Unidas para o Meio Ambiente
<i>UNFCCC</i>	Convenção-Quadro das Nações Unidas sobre Mudanças Climáticas

Lista de símbolos

t	Tonelada
Mt	Um milhão de tonelada
Gt	Um bilhão de tonelada
A	Tendência aditiva
A_d	Tendência aditiva amortecida
AA	Tendência aditiva e erro aditivo
AA_d	Tendência aditiva amortecida e erro aditivo
CH_4	Metano
CO_2	Dióxido de carbono
CO_2e	Dióxido de carbono equivalente
cov	Covariância
lag	Refere-se ao intervalo de tempo entre duas observações
M	Tendência multiplicativa
M_d	Tendência multiplicativa amortecida
MA	Componente de médias móveis
MA_d	Tendência aditiva amortecida e erro multiplicativo
MM	Tendência multiplicativa e erro multiplicativo
N	Sem tendência
N_2O	Óxido nitroso
p	Ordem do componente autoregressivo
q	Ordem do componente de média móvel
α	Parâmetro de suavização para o nível
β	Parâmetro de suavização para a tendência
γ_t	Autocovariância
φ	Parâmetros de amortecimento da tendência
ρ_k	Coefficiente de correlação
σ^2	Variância
y_t	Série temporal

Sumário

1	Introdução	13
1.1	Apresentação do tema	13
1.2	O problema	15
1.3	Objetivos	15
1.3.1	Geral	15
1.3.2	Específicos	15
1.4	Justificativa	16
1.5	Estruturação da dissertação	16
2	Revisão da Literatura	18
2.1	Emissões de gases do efeito estufa	18
2.2	Modelagem de séries temporais	19
2.3	Modelos de previsão	21
2.3.1	Modelo de suavização exponencial	21
2.3.2	Metodologia de Box-Jenkins	24
2.4	Teste de tendência e de quebra estrutural	26
2.5	Teste de estacionariedade	27
2.6	Testes de diagnóstico de resíduos	29
2.7	CrITÉrios de seleção de modelos	30
2.8	Métricas de avaliação de desempenho	31
2.9	Estudos relacionados	32
3	Metodologia	36
3.1	Caracterização da Área de Estudo	36
3.2	Dados e sua análise	36
4	Resultados	39
4.1	Análise exploratória dos dados	39
4.2	Modelo de suavização exponencial	46
4.3	Metodologia de Box-Jenkins - ARIMA	52
4.4	Discussão entre as previsões realizadas pela Suavização Exponencial e ARIMA	63
5	Conclusão	66
	Referências	67
	Apêndices	73
	Apêndice A Quebra de estrutura	74

1 Introdução

1.1 Apresentação do tema

O efeito estufa é um mecanismo natural do planeta para manter a temperatura em equilíbrio, sendo potencializado por grandes emissões de gases poluentes (Aguiar; Fortes; Martins, 2016). Dentre os gases do efeito estufa (GEE) estão o gás carbônico (CO_2), o metano (CH_4) e o óxido nitroso (N_2O), que são emitidos quando a floresta é desmatada, tanto pela queimada quanto pela decomposição da biomassa que não queima (Canadell; Pedro, 2022).

As emissões de GEE estão conectadas a diversas áreas da economia global, pois são geradas por atividades econômicas que dependem das fontes de energia que os produzem. Por exemplo, a emissão de GEE pode ser gerada pela produção de eletricidade a partir de combustíveis fósseis, como o petróleo, o gás natural e o carvão, bem como pela produção de bens de consumo que envolvem o uso de energia intensiva, como o transporte e a indústria (Estevo, 2020; Bakay; Ağbulut, 2021).

Os efeitos generalizados e inevitáveis das mudanças climáticas se manifestam como temperaturas elevadas e crescentes, períodos de seca, variações nos padrões de chuva, elevação do nível médio do mar e inundações, dentre outros (Dankwa *et al.*, 2021). Estas mudanças exercem ameaças aos ecossistemas, recursos hídricos, segurança alimentar e saúde na maioria dos países, principalmente nos países em desenvolvimento (Hatfield *et al.*, 2011). De acordo com Kosoe e Ahmed (2022), as mudanças climáticas também afetam a produção de atividades econômicas, especialmente a agricultura.

A redução da emissão de GEE é fundamental para evitar o agravamento do superaquecimento do planeta e proteger o meio ambiente. Isso pode ser feito por medidas como: aumentar a eficiência energética, promover fontes de energia limpa (energia solar e eólica) e implementar políticas públicas para limitar sua emissão. Além disso, é importante que as grandes economias globais liderem pelo exemplo e estabeleçam metas ambiciosas para redução das emissões de GEE, visando proteger o meio ambiente e combater as mudanças climáticas (Giddens, 2010).

Tayra e Dos Reis (2020) afirmam que as causas das mudanças climáticas são as concentrações dos GEE na atmosfera devido ao desmatamento, à produção agropecuária e, principalmente, à queima de combustíveis fósseis. A redução das emissões de GEE é reconhecida como uma necessidade urgente, visto que os combustíveis fósseis são responsáveis por 90% das emissões desses gases. Essa urgência se reflete na necessidade de reduzir os lançamentos de CO_2 na atmosfera, que ocorrem principalmente devido à queima de combustíveis fósseis. As grandes economias globais, como a China, os Estados Unidos e a Índia, são responsáveis pela emissão de GEE em excesso, por dependerem dessas fontes de energia.

Contudo, o que se tem acompanhado nos últimos anos, é o aumento das emissões deste gás. Em 2017, elas cresceram 1,6%, e esperava-se um crescimento acima de 2% em 2018, graças ao aumento do uso de combustíveis fósseis, principalmente para geração de energia elétrica. Esse era o ritmo que vinha sendo observado até a crise sanitária de 2020 (Friedlingstein; O'Sullivan; Jones, 2022).

No ano de 2020, as emissões de CO_2 geradas por combustíveis fósseis caiu 5,4% em relação ao ano anterior, o maior declínio registrado até o momento. Isso ocorreu devido ao impacto da pandemia de COVID-19 nas atividades econômicas, que reduziram o uso de combustíveis fósseis em todo o mundo. A *British Petroleum* estimou um declínio de 4,5% no consumo de energia primária e um declínio de 6,3% nas emissões globais de CO_2 em 2020, em comparação com 2019. A demanda de energia primária aumentou 5,8% em 2021, superando os níveis de 2019 em 1,3%. Os combustíveis fósseis representaram 82% do uso de energia primária no ano

passado, abaixo dos 83% em 2019 e 85% há cinco anos. As emissões de dióxido de carbono equivalente (CO_2e) do uso de energia, processos industriais, queima e metano aumentaram 5,7% em 2021, com as emissões de dióxido de carbono da energia subindo 5,9%, perto dos níveis de 2019 (British Petroleum, 2023).

As emissões mundiais de GEE somavam entre 51 e 58 bilhões de toneladas (Gt) de CO_2e em 2021, segundo o relatório *Emissions Gap Report 2022*, publicado pelo Programa das Nações Unidas para o Meio Ambiente (UNEP). O Brasil contribuiu em 2021 com 2,42 Gt de CO_2e , o que corresponde a um valor entre 3,5% e 4,5% das emissões globais anuais (Crippa *et al.*, 2022). Esta participação já chegou a representar 6% em 2003 (Ipcc, 2021). Em 2020, a média per capita das emissões de CO_2 no Brasil foi de 10,2t brutas contra 6,7t da média mundial, 65,7% acima desta média (Observatório do Clima, 2023).

Mesmo com uma contribuição relativamente pequena em relação às emissões globais, o Brasil se mantém entre os maiores emissores do planeta. O país é o sétimo maior emissor de GEE do mundo, com 3% do total mundial, atrás de China (25,2%), EUA (12%), Índia (7%), União Europeia (6,6%), Rússia (4,1%) e Indonésia (4%), segundo o Observatório do Clima (2023). No entanto, como o desmatamento caiu na Indonésia nos últimos anos e a série de dados globais utilizada nesta análise só vai até 2019, é provável que o Brasil seja na realidade o sexto maior emissor. Mais de 70% da emissão dos gases causadores do efeito de estufa está relacionada com a mudança do uso das terras de floresta (MUTF) e a atividade agropecuária (Potenza *et al.*, 2021).

A Amazônia desempenha um papel importante no sistema climático global associado à produção de chuvas, que a definem como uma fonte de calor tropical. Por outro lado, passa por processos de muitas atividades humanas, como as mudanças no uso do solo através do desmatamento, que atingiram níveis preocupantes na década de 70, permanecendo até os dias atuais (Pinho, 2019). A Amazônia é o bioma que historicamente mais tem emitido GEE, decorrentes principalmente do avanço da pecuária sobre as florestas. Em relação às emissões estaduais, o estado do Amazonas emitiu 109,1 Mt de CO_2e por MUTF em 2020, a MUTF sozinha representa 60,5% da emissão do setor no Brasil (Potenza *et al.*, 2021).

O Brasil possui três inventários nacionais de GEE e quatro “relatórios de atualização” (MCTI, 2020), mas estes foram desenvolvidos sem periodicidade ou revisão internacional por pares. O primeiro inventário nacional foi lançado em 2005 e abrange os anos entre 1990 e 1994, enquanto o segundo saiu em 2010 abrangendo o período de 1990-2005. Já o terceiro inventário nacional de GEE foi desenvolvido cobrindo o período de 1990-2010 (Scolastrici; Villela, 2016). Essas lacunas de tempo na geração de inventários e um atraso de cobertura de quase 5 anos podem deixar de captar e entender as tendências de emissões de GEE no Brasil e, em última análise, o desenho de políticas efetivas para descarbonizar sua economia e cumprir com as promessas assumidas no acordo de Paris, ocorrido na Convenção-Quadro das Nações Unidas sobre Mudanças Climáticas (UNFCCC) de 2015.

Diante do exposto, torna-se evidente que algo deve ser feito para amenizar os efeitos dos GEE no meio ambiente. O estudo e a compreensão dos impactos das mudanças climáticas globais sobre cada região do país são imprescindíveis para o planejamento estratégico e para a tomada de decisões por parte do poder público e da sociedade (Copertino *et al.*, 2017). Para isto, são necessários estudos e processamentos dos dados, como a modelagem de séries temporais para fornecer informações relacionadas às tendências das emissões de CO_2e para fornecer subsídios para adoção de políticas públicas responsáveis.

Diversos métodos e modelos estatísticos, matemáticos, computacionais ou híbridos podem ser utilizados na modelagem de séries temporais, permitindo prever quantitativamente uma variável e medir a qualidade da previsão estabelecida. Para o ajuste de tais modelos, é necessário estudar o comportamento da série a fim de descobrir suas principais características e o método de modelagem mais adequado.

De acordo com Junqueira *et al.* (2022) os métodos estatísticos de previsão por séries temporais têm sido cada vez mais utilizados, uma vez que representam um diferencial para antecipar cenários futuros. Bowersox e

Closs (2007) classificam as técnicas de previsão em três tipos:

1. Técnica qualitativa: envolve a utilização de julgamentos subjetivos e experiência para prever resultados futuros, sendo adequada para situações em que os dados quantitativos são escassos ou incertos.
2. Técnica causal: envolve a identificação de relações causais entre variáveis e o uso dessas relações para prever resultados futuros, sendo mais adequada para situações em que existem dados quantitativos suficientes e é possível estabelecer relações causais entre as variáveis.
3. Técnica de série temporal: envolve a análise de séries de dados ao longo do tempo para prever resultados futuros, sendo usada nas situações em que existem dados quantitativos suficientes e é possível identificar padrões e tendências ao longo do tempo.

O objetivo da análise de séries temporais é encontrar modelos que permitam descrever de forma adequada um processo de geração de dados, que consiga prever o comportamento futuro de uma determinada variável para um determinado horizonte de previsão (Veríssimo *et al.*, 2013). Entre os métodos estatísticos univariados estão o de Suavização Exponencial e os de Box-Jenkins.

O modelo de Suavização Exponencial é uma técnica que permite ajustar os dados da série temporal para refletir a tendência ao longo do tempo, sendo bastante útil para modelar séries temporais que apresentam tendências significativas, crescente ou decrescente. No entanto, é importante lembrar que esse método pode não ser a melhor opção para todas as séries temporais, sendo sempre importante considerar vários métodos de previsão e escolher o que melhor se adapta aos dados e às necessidades específicas de cada situação (Ragsdale, 2021).

Os modelos de séries temporais têm sido amplamente empregados para previsões em diversas áreas como Finanças, Economia, Meteorologia e Engenharia (Tsay, 2010). Reconhecidos por sua confiabilidade e precisão, esses modelos, exemplificados pelo trabalho de Box-Jenkins, oferecem uma flexibilidade notável, incorporando tanto termos de médias móveis quanto termos autorregressivos (Naim; Mahara, 2018). Eles se destacam pela capacidade de ajuste a diferentes tipos de dados, tornando-se uma escolha confiável para previsões em séries temporais.

1.2 O problema

Os métodos de suavização exponencial e os métodos de Box-Jenkins já produziram excelentes resultados na modelagem de séries temporais em vários contextos (Alam; AlArjani, 2021; Thabani; Hapanyengwi; Mutongi, 2019; Nyoni; Bonga, 2019). Contudo, para a modelagem de séries temporais de emissões de CO_2e no estado do Amazonas, qual abordagem pode produzir resultados mais representativos e esclarecedores? No presente trabalho propõe-se avaliar qual destes modelos pode ser mais elucidativo sobre o mecanismo gerador de dados das emissões de CO_2e no estado do Amazonas.

1.3 Objetivos

1.3.1 Geral

- Identificar o mecanismo gerador dos dados que compõem a série temporal de emissão de CO_2e no estado do Amazonas, fornecendo subsídios para elaboração de políticas públicas preventivas.

1.3.2 Específicos

- Fazer uma análise exploratória da série temporal de emissão de CO_2e no estado do Amazonas e no Brasil.
- Modelar a série temporal analisada, por meio da técnica de suavização exponencial.

- Modelar a série temporal analisada, mediante a metodologia de Box-Jenkins.
- Comparar os modelos para verificar qual apresenta o melhor desempenho.
- Fazer uma análise crítica das técnicas usadas para avaliar o mecanismo gerador dos dados das séries analisada, discutindo os achados.

1.4 Justificativa

A compreensão e mitigação das emissões de gases de efeito estufa (GEE) se reveste de uma relevância ambiental e social incontestável. As emissões desses gases têm impactos diretos nas mudanças climáticas globais, afetando ecossistemas, recursos hídricos, segurança alimentar e saúde humana, especialmente em regiões vulneráveis como o estado do Amazonas. Este estudo visa contribuir para a proteção do meio ambiente e o bem-estar da população ao fornecer um entendimento mais profundo dessas emissões.

Do ponto de vista científico, a pesquisa proposta preenche uma lacuna na literatura, uma vez que há uma escassez de estudos que se aprofundem na modelagem das emissões de CO_2e no estado do Amazonas, usando diferentes abordagens estatísticas. A comparação entre os modelos de Suavização Exponencial e Box-Jenkins promete fornecer informações valiosas sobre o mecanismo gerador dessas emissões, contribuindo para um corpo de conhecimento mais abrangente.

O ano de 2023 apresentou eventos climáticos extremos, com as mudanças climáticas agravadas pelo fenômeno *El Niño*, intensificado pelo aumento médio de temperaturas nos oceanos. Registrou-se, ainda, as mais altas temperaturas dos últimos 120 anos (Calvin *et al.*, 2023). A realidade preocupante se manifesta com o menor nível do rio Negro em Manaus-AM, atingindo a marca mínima histórica (ANA, 2023). As chuvas intensas nos estados de Santa Catarina e Rio Grande do Sul já resultaram em mais de 50 vítimas e milhares de desabrigados (Defesa Civil-RS, 2023). Tais eventos enfatizam a urgência de ações concretas na compreensão e enfrentamento das emissões de GEE.

Além disso, esta pesquisa não se limita ao aspecto teórico, ao ter implicações práticas significativas. Os resultados deste estudo podem ser aplicados na formulação de políticas públicas, no planejamento estratégico e na tomada de decisões, tanto ao nível estadual quanto nacional. Ao compreendermos melhor as emissões de CO_2e no contexto do Amazonas, podemos trabalhar para reduzi-las e, assim, promover um desenvolvimento sustentável e a preservação do meio ambiente para as gerações futuras.

1.5 Estruturação da dissertação

Este trabalho está estruturado em 5 (cinco) capítulos, sendo que o capítulo 1 apresenta a temática central sobre as emissões de GEE no estado do Amazonas. O problema central, focado na comparação entre os métodos de suavização exponencial e Box-Jenkins na modelagem de séries temporais de emissões de CO_2e , é delineado. Aqui, os objetivos geral e específicos são estabelecidos, destacando a relevância da identificação do mecanismo gerador desses dados para embasar a formulação de políticas públicas preventivas.

O capítulo 2 aborda conceitos fundamentais de séries temporais, os métodos de suavização exponencial e a metodologia de Box-Jenkins. A revisão contempla também discussões sobre estacionariedade das séries, testes de heterocedasticidade, autocorrelação e normalidade, bem como critérios de precisão.

No capítulo 3, são detalhados os métodos e procedimentos adotados para a análise das séries temporais relacionadas às emissões de CO_2e . A metodologia abrange desde a análise exploratória das séries até a comparação crítica de desempenho entre os modelos propostos.

Os resultados obtidos são apresentados e discutidos no capítulo 4. Destacam-se os modelos mais adequados para cada categoria de emissões, evidenciando o desempenho dos modelos selecionados. Além disso, enfatiza-se a aplicabilidade prática dos resultados na formulação de políticas públicas e tomada de

decisões estratégicas.

O último capítulo conclui a pesquisa, reforçando a importância da compreensão das emissões de GEE não apenas como um aspecto teórico, mas como uma base essencial para ações práticas voltadas ao desenvolvimento sustentável e à preservação do meio ambiente. Aqui, são sintetizados os principais achados, destacando a contribuição da pesquisa para o conhecimento na área e delineando possíveis direções para estudos futuros.

2 Revisão da Literatura

Este capítulo visa fornecer uma base teórica para compreender o contexto e os métodos utilizados nesta pesquisa, permitindo orientar o leitor através dos conceitos e metodologias usadas no desenvolvimento e análise de modelos de previsão em séries temporais. Inicialmente, são examinadas as emissões de GEE. Em seguida, é explorada a modelagem de séries temporais, oferecendo uma visão ampla das técnicas e abordagens empregadas para entender e prever padrões temporais em dados. Os modelos de previsão, especificamente a suavização exponencial e a metodologia de Box-Jenkins, proporcionando uma compreensão das ferramentas disponíveis para realizar projeções precisas. Além disso, são tratados os testes de tendência, estacionariedade e quebra estrutural, juntamente com os testes de diagnóstico de resíduos e os critérios de seleção de modelos. Por fim, são apresentadas métricas de avaliação de desempenho e estudos relacionados que contribuem para fundamentar e contextualizar os resultados desta pesquisa.

2.1 Emissões de gases do efeito estufa

O dióxido de carbono equivalente (CO_2e) é uma medida internacional cuja finalidade é estabelecer a equivalência entre todos os GEE e o CO_2 . Assim, os demais GEE são convertidos em CO_2 para ser mais fácil analisar os impactos das emissões para o aquecimento global, sendo esta a maneira usual de medir a quantidade de GEE que uma atividade emite para o meio ambiente (Observatório do Clima, 2023). O CO_2e é calculado multiplicando a quantidade de um GEE específico (como dióxido de carbono, metano ou óxido nitroso) por um "fator de conversão" que indica o quão potente esse gás é comparado ao CO_2 (Lintunen; Rautiainen, 2021).

Os gases equivalentes de dióxido de carbono são comumente expressos como milhões de toneladas de equivalentes de dióxido de carbono (abreviado como MMTCDE), mas muitas vezes apresenta-se em quilos, também. Para quantificar o CO_2e , multiplica-se a quantidade de GEE pelo seu potencial de aquecimento global (GWP), em inglês, *Global Warming Potential*, ou Fator de Aquecimento Global (GWF), em inglês, *Global Warming Factor* (ICBE, 2000).

O CO_2 tem um fator de conversão igual a 1 (um), visto que é a base de comparação conforme a Tabela 2.1. O fator de conversão para todos os outros GEE é o número de vezes o potencial de aquecimento que eles causam quando comparados ao CO_2 . Assim, 1kg de metano causa 28 vezes mais aquecimento ao longo de 100 anos em comparação com 1 kg de CO_2 , e é por isso que a pontuação do metano é de 28 (Estevo, 2020). Desta forma, se 1kg de metano for emitido, ele pode ser expresso como 28 kg de CO_2e ($1kg CH_4 \times 28 = 28kg CO_2e$).

Tabela 2.1 – Índice de conversão de CO_2e

Gás	Valores de GWP para horizonte de tempo de 100 anos
Dióxido de carbono (CO_2)	1
Metano (CH_4)	28
Óxido nitroso (N_2O)	265

Fonte: Relatório de avaliação relacionado ao CO_2 - AR5 (IPCC, 2013).

No setor de Processos Industriais e Uso de Produtos (PIUP) são contabilizadas as emissões provenientes de transformações físicas ou químicas de materiais na indústria. A combustão de combustível industrial e as emissões de resíduos são contabilizadas nos setores de produção de energia e tratamento de resíduos, respectivamente (Azevedo *et al.*, 2018).

As emissões oriundas do tratamento de resíduos, abrange as estimativas de emissões de GEE de resíduos sólidos e tratamento e descarga de águas residuais, exceto as emissões relacionadas com os resíduos gerados

nas atividades agrícolas e pecuárias, contabilizadas no setor agropecuária. As emissões de metano (CH_4) e dióxido de carbono (CO_2) em aterros sanitários são os principais emissores de GEE em regiões metropolitanas (Azevedo *et al.*, 2018).

As emissões do setor de energia são provenientes da queima de combustíveis em atividades como transportes, indústria e geração de eletricidade, além das chamadas emissões fugitivas, causadas pelo escape de GEE durante a produção de combustíveis (como a fuga de metano na exploração de petróleo e gás natural) (SEEG, 2021).

No setor de Agropecuária são contabilizadas as emissões provenientes da digestão dos animais ruminantes, que emite metano – o popular “arroto” do boi – (fermentação entérica), do tratamento e disposição que os dejetos desses animais recebem (manejo de dejetos), do cultivo de arroz irrigado (cultivo de arroz), da queima dos resíduos agrícolas do cultivo de cana-de-açúcar e algodão (queima de resíduos agrícolas) e as originadas pela forma de como os solos agrícolas são manejados, considerando o incremento de nitrogênio via utilização de insumos e operações agrícolas (solos manejados) (SEEG, 2021). A agricultura contribui para a emissão de GEE por meio de várias atividades, entre elas o uso de combustíveis fósseis, queimadas, utilização de insumos e conversão do uso da terra (Pachauri; Meyer, 2014).

As MUTF foram responsáveis pela emissão de 655.9 Mt de CO_2 em 2020 no Brasil (Observatório do Clima, 2023). Isso consolida o setor como a maior fonte de emissão bruta de GEE do país, representando 46% do total nacional naquele ano. As mudanças na cobertura da terra representam os principais níveis de emissões e remoções de CO_2 . Este setor também estima as emissões de resíduos florestais queimados (CH_4 e N_2O) e calagem (Azevedo *et al.*, 2018). Quando consideradas as remoções por áreas protegidas, vegetação secundária e outras mudanças de uso da terra 636 Mt de CO_2 e, o setor apresentou uma emissão líquida de 362 Mt CO_2 e no último ano, cerca de 24% das emissões líquidas brasileiras (Quintana; Cardoso; Potenza, 2023).

2.2 Modelagem de séries temporais

Conforme Box *et al.* (2015), o conceito de séries temporais engloba um conjunto de observações que se desdobram sequencialmente ao longo do tempo. No contexto da modelagem de séries temporais, a atenção se volta principalmente a uma categoria específica de processos estocásticos, conhecidos como processos estocásticos estacionários. Estes processos são definidos pela sua característica de equilíbrio constante, ou seja, independentemente do ponto temporal ou da janela de observação, tanto a média quanto a variância da série permanecem inalteradas.

A verificação da estacionariedade de uma série temporal é uma importante propriedade a ser considerada na modelagem de um processo estocástico. As séries temporais podem ser estritamente estacionárias ou fracamente estacionárias. Em geral, quando se verifica a estacionariedade de uma série, faz-se referência a estacionariedade fraca, também conhecida como estacionariedade de segunda ordem (Morettin; Toloi, 2018).

O propósito da análise de séries temporais reside na formulação de modelos matemáticos que ofereçam descrições para os dados amostrais. Com o intuito de fornecer uma estrutura estatística para descrever a natureza dos dados que exibem flutuações aparentemente aleatórias ao longo do tempo, parte-se do pressuposto de que uma série temporal pode ser concebida como uma coleção de variáveis aleatórias indexadas de acordo com sua ordem de obtenção no tempo (Robert; David, 2017).

Graficamente, uma série temporal é representada por um gráfico de linha no qual o eixo x refere-se ao tempo e o eixo y corresponde a variável (Castanheira, 2013). Sua modelagem permite identificar um padrão de comportamento para a variável aleatória, possibilitando analisar e entender o passado para prever o futuro (Morettin; Toloi, 2018). O tipo de modelo, portanto, depende de como cada uma dessas componentes é considerada.

Cabe aqui incluir os principais conceitos sobre séries temporais usados nessa pesquisa. Matematica-

mente, a série temporal é definida como:

Discreta: $y_1, y_2, \dots, y_T$ ou y_t , onde $t = 1, 2, \dots, T \in \mathbb{N}$

Contínua: $y(t)$, onde $t \in T \subseteq \mathbb{R}$.

Seja y_t uma sequência de variáveis aleatórias indexadas pelo tempo t . De acordo com Zivot e Wang (2007), esta sequência de variáveis aleatórias é denominada série temporal. A série temporal y_t é considerada estacionária se satisfizer as seguintes condições:

$$E(y_t) = \mu, \text{ para } \forall t,$$

$$\text{cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \text{ para } \forall t \text{ e para } \forall k,$$

onde a expressão $E(y_t)$ representa a expectativa matemática de y_t , o que equivale à média da série temporal. A constante μ denota essa média, que se mantém inalterada ao longo da série. A covariância entre y_t e y_{t-k} , dada por $\text{cov}(y_t, y_{t-k})$, reflete a relação entre as observações em momentos temporais distintos, indicando como elas variam conjuntamente. Por fim, γ_k representa a autocovariância para um deslocamento temporal de k , revelando a natureza da autocorrelação na série temporal, ou seja, como as observações estão correlacionadas com elas mesmas em intervalos específicos de tempo. Esses elementos são essenciais para compreender a estacionariedade de uma série temporal.

Se a média de y_t for constante para todos os instantes de tempo t , onde μ representa essa média, e se a covariância entre y_t e y_{t-k} for constante para todos os instantes de tempo t , onde γ_k é a autocovariância de ordem k ou defasagem k de y_t , para séries temporais estacionárias, então a autocovariância γ_k depende apenas de k . Ela mede a dependência linear entre y_t e y_{t-k} , ou seja, a dependência dinâmica de y_t em seu passado y_{t-k} .

A análise de séries temporais lineares se concentra em estudar a dependência dinâmica da série y_t . O coeficiente de correlação entre y_t e y_{t-k} , denotado por ρ_k , é chamado de autocorrelação de defasagem k ou lag k . Formalmente, é definido pela Equação 2.1:

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t)\text{var}(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}, \quad (2.1)$$

onde a variância $\text{var}(y_t) = \text{var}(y_{t-k}) = \gamma_0$, pois y_t é estacionária. A partir da Equação 2.1, temos que $\rho_0 = 1$, $\rho_k = \rho_{-k}$, $-1 < \rho_k < 1$ e que o valor de ρ_k independe da escala de medição da série temporal. A coleção de autocorrelações, $\{\rho_k\}$, é chamada de função de autocorrelação (FAC).

De acordo com Box *et al.* (2015), o fenômeno que evolui no decorrer do tempo sob a influência de leis probabilísticas é denominado um processo estocástico. Em muitas ocasiões, refere-se a isso simplesmente como um "processo", deixando de lado o termo "estocástico".

Segundo Zivot e Wang (2007), uma série temporal estacionária y_t é ergódica se os momentos amostrais convergem em probabilidade (p) para os momentos populacionais, isto é, se $\bar{y} \xrightarrow{p} \mu$, $\hat{\gamma}_j \xrightarrow{p} \gamma_j$ e $\hat{\rho}_j \xrightarrow{p} \rho_j$. Aqui, as médias amostrais $\bar{y}$ convergem em probabilidade para a média populacional μ , as estimativas amostrais de covariância $\hat{\gamma}_j$ convergem em probabilidade para as covariâncias populacionais γ_j , e as estimativas amostrais de correlação $\hat{\rho}_j$ convergem em probabilidade para as correlações populacionais ρ_j .

Conforme Tsay (2014), uma série temporal é classificada como ruído branco se for composta por uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid), as quais possuem média zero e variância finita. No contexto específico em que tais variáveis a_t seguem uma distribuição normal com média 0 e variância σ_a^2 , a série é referida como ruído branco gaussiano.

Conforme descrito por Tsay (2014), uma série temporal y_t é considerada linear se puder ser expressa na forma da Equação 2.2:

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad (2.2)$$

onde, μ representa o termo de interceptação e ψ_i são os coeficientes associados aos termos defasados a_{t-i} da série temporal, $\psi_0 = \mu + \sum_{i=0}^{\infty} \psi_i^2 < \infty$ variância finita, e a_t é uma série de ruído branco. Uma vez que a_t denota a nova informação no instante t da série temporal é muitas vezes referida como a inovação ou choque no tempo t e será denotado por $a_t \sim RB(0, \sigma_a^2)$. Segundo Zivot e Wang (2007), o teorema da decomposição de Wold, afirma que qualquer série temporal estacionária, y_t , tem uma representação média móvel de ordem finita na forma da Equação 2.2. Esta equação representa a soma de uma constante μ mais um processo ruído branco gaussiano a_t .

2.3 Modelos de previsão

Conforme descrito por Hyndman *et al.* (2023), um método de previsão é um algoritmo utilizado para obter uma previsão pontual para algum tempo futuro. Já um modelo estatístico estrutura o processo estocástico gerador de dados, utilizado para obter uma distribuição de probabilidade para algum tempo futuro. A partir desse modelo, pode-se utilizar a distribuição de probabilidade estimada para obter previsões pontuais e intervalares.

A maioria dos métodos de previsão baseia-se na ideia de que as observações passadas contêm informações sobre padrões de comportamento da série temporal. O propósito dos métodos é diferenciar o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros da série (Morettin; Toloi, 2018).

Os modelos de Suavização Exponencial e Box-Jenkins estão entre os mais utilizados para a previsão por meio da análise de séries temporais. Eles fornecem abordagens complementares para a resolução de um determinado problema (Taneja *et al.*, 2016; Thabani; Hapanyengwi; Mutongi, 2019; Nyoni; Bonga, 2019). Segundo Martins, Machado e Esteves (2004), os modelos de suavização exponencial são baseados em uma descrição da tendência e sazonalidade dos dados, enquanto os modelos ARIMA visam descrever as autocorrelações dos dados (Box; Al, 2015) e ambos os modelos são muito aplicados em previsões do mercado financeiro (Khashei; Bijari; Ardali, 2009).

A seguir, serão expostos os conceitos concernentes aos métodos de suavização exponencial, seguidos pelos conceitos das abordagens de Box-Jenkins, ambos voltados para a identificação da tendência por meio da análise de séries temporais sem sazonalidade.

2.3.1 Modelo de suavização exponencial

Ao longo de quase meio século, os métodos de suavização exponencial têm sido amplamente conhecidos e populares entre os preditores, sendo sua origem associada ao trabalho seminal de Brown (1956). Hyndman *et al.* (2002) introduziram o modelo ETS (Erro, Tendência e Sazonalidade), que incorpora a suavização exponencial em uma estrutura de espaço de estado, proporcionando uma base estatística sólida. A notação ETS ($\cdot, \cdot, \cdot$) ajuda a lembrar a ordem na qual os componentes são especificados. ETS também é a abreviação de *Exponential Smoothing*.

O método de suavização exponencial atribui pesos decrescentes exponencialmente aos dados históricos, dando maior peso aos dados mais recentes em comparação com os mais antigos. Essa atribuição exponencial de pesos significa que os valores mais recentes exercem uma influência maior nas previsões, conferindo à suavização exponencial uma maior sensibilidade às mudanças recentes na série temporal (Morettin; Toloi, 2018).

A suavização exponencial é um método utilizado para prever tendências futuras em séries temporais (Svetunkov; Kourentzes; Ord, 2022). Nesse processo, inicia-se considerando a componente de tendência, que é uma combinação de um termo de nível (ℓ) e um termo de crescimento (b). A combinação entre o nível e o crescimento pode ser feita de várias maneiras, resultando em cinco tipos diferentes de padrões de tendência

futura, apresentados na Tabela 2.2.

O primeiro tipo é chamado 'nenhum padrão de tendência' (N). Nesse caso, a série temporal não apresenta crescimento ou declínio significativo, o que significa que y_t é igual a ℓ existente. O segundo tipo é denominado 'tendência aditiva' (A). Aqui, a y_t é calculada somando o ℓ com b_t projetado para os próximos períodos. Outro padrão é a 'tendência aditiva amortecida' (A_d). Nessa abordagem, uma componente de amortecimento (φ) é introduzida para controlar o impacto do crescimento sobre o nível ao longo do tempo. Aqui a y_t é obtida somando ℓ com um termo amortecido, calculado como a soma dos termos de amortecimento.

Tabela 2.2 – Classificação dos métodos de suavização método

Tendência	Equações
N	$y_t = \ell$
A	$y_t = \ell + b_t$
A_d	$y_t = \ell + (\varphi + \varphi^2 + \dots \varphi^t)$
M	$y_t = \ell b^t$
M_d	$y_t = \ell \cdot (\varphi + \varphi^2 + \dots \varphi^t)$

N nenhuma, A aditiva, M multiplicativa, A_d aditiva amortecida, M_d multiplicativa amortecida
 y_t valor observado no período t (o valor real), ℓ nível da série, b_t tendência (taxa de crescimento ou decrescimento),
 φ amortecimento que controla a redução da tendência ao longo de t ($0 < \varphi < 1$), e_t erro (ou resíduo).

Fonte: Adaptado de Hyndman e Athanasopoulos (2023).

Já o quarto tipo é a 'tendência multiplicativa' (M). Nessa situação, y_t é calculada multiplicando o ℓ pelo b_t projetado para os períodos futuros. Por fim, o quinto padrão é a 'tendência multiplicativa amortecida' (M_d), que é semelhante ao caso aditivo amortecido. Aqui, a y_t é obtida multiplicando o ℓ pelo crescimento amortecido, calculado como o produto do b com a soma dos termos de amortecimento.

Essas diferentes abordagens permitem adaptar a suavização exponencial a diversas situações e padrões encontrados em séries temporais, tornando-a versátil e eficaz para análise e previsão de dados. Esta classificação de métodos originou-se com a taxonomia de Pegels (1969). Mais tarde, isso foi manipulado por Gardner (1985), modificado por Hyndman *et al.* (2002), e ampliado novamente por Taylor (2003).

Um método de tendência amortecida é apropriado quando há uma tendência na série temporal, mas acredita-se que é improvável que a taxa de crescimento no final dos dados históricos continue por mais de um curto período no futuro. As equações para tendência amortecida fazem o que o nome indica: amortecem a tendência à medida que o comprimento do horizonte de previsão aumenta. Isso geralmente melhora a precisão da previsão, especialmente em longos prazos de entrega.

A seguir apresenta-se uma introdução simples de alguns dos métodos de suavização exponencial mais conhecidos: suavização exponencial simples (N,N), método linear de Holt (A,N) e método de tendência amortecida (A_d,N). Considerando a série temporal observada por $y_1, y_2, \dots, y_n$. Uma previsão de y_{t+h} baseada em todos os dados até o tempo t é denotada por $\hat{y}_{t+h|t}$, para previsões de uma etapa, usa-se a notação mais simples $\hat{y}_{t+1} \equiv \hat{y}_{t+h|t}$. Normalmente, as previsões requerem alguns parâmetros para serem estimados, mas, para simplificar, assume-se que os valores de todos os parâmetros relevantes são conhecidos.

Para cada método, existem dois modelos, um modelo com erros aditivos e um modelo com erros multiplicativos. O método de suavização exponencial simples (SES) é usado para fazer previsões de dados que não mostram uma tendência clara ou um padrão sazonal. Quando a observação y_t se torna disponível, o erro de previsão é $y_t - \hat{y}_t$. O método SES, devido ao trabalho de Brown em meados da década de 1950 e publicado em Brown (1956), utiliza a previsão do período anterior e a ajusta usando o erro de previsão. Ou seja, a previsão para o próximo período é dada pela Equação 2.3:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t), \quad (2.3)$$

onde α é uma constante entre 0 e 1.

Observa-se que a atualização da previsão é basicamente a previsão anterior somada a um ajuste que considera o erro ocorrido na última previsão. Se o parâmetro α estiver próximo de 1, o ajuste para o erro da previsão anterior será significativo na nova previsão. Por outro lado, quando α se aproxima de 0, o ajuste na nova previsão será mínimo.

Para previsões de intervalo mais longo, assume-se que a função de previsão é "plana". Uma função de previsão plana é usada porque a suavização exponencial simples funciona melhor para dados que não têm tendência, sazonalidade ou outros padrões subjacentes.

Outra maneira de expressar este conceito é $\ell_t = \hat{y}_{t+1}$. Então $\hat{y}_{t+h|t} = t$ e $t = \alpha y_t + (1 - \alpha)\ell_{t-1}$. O valor de t é uma medida do "nível" da série no tempo t . Embora isso possa parecer uma maneira complicada de expressar o método, ele fornece uma base para generalizar a suavização exponencial para permitir tendência e sazonalidade.

Para calcular as previsões usando SES, precisamos especificar o valor inicial $0 = \hat{y}_1$ e o valor do parâmetro α . Tradicionalmente (particularmente na era pré-computador), $\hat{y}_1$ era definido como igual à primeira observação e α era especificado como um número pequeno, geralmente 0,2. Esses valores iniciais eram escolhidos com base em considerações práticas e experiência empírica.

Holt (1957) estendeu a SES para permitir a previsão de dados em séries com uma tendência. Este método envolve uma equação de previsão e duas equações de suavização, uma para o nível e outra para a tendência. As equações são apresentadas na Tabela 2.3, onde α é o parâmetro de suavização para o nível, β é o parâmetro de suavização para a tendência, ambos α e β variam entre 0 e 1.

Tabela 2.3 – Fórmulas para cálculos recursivos e previsões pontuais

Tendência	Equações
Nenhuma (N)	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$ $\hat{y}_{t+h t} = \ell_t$
Aditiva (A)	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$ $\hat{y}_{t+h t} = \ell_t + hb_t$
Aditiva amortecida (A_d)	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \varphi b_{t-1})$ $b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\varphi b_{t-1}$ $\hat{y}_{t+h t} = \ell_t + \varphi h b_t$
Multiplicativa (M)	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} b_{t-1})$ $b_t = \beta(\frac{\ell_t}{\ell_{t-1}}) + (1 - \beta)\varphi b_{t-1}$ $\hat{y}_{t+h t} = \ell_t + b_t^h$
Multiplicativa amortecida (M_d)	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} b_{t-1}^\varphi$ $b_t = \beta(\frac{\ell_t}{\ell_{t-1}}) + (1 - \beta)\varphi b_{t-1}^\varphi$ $\hat{y}_{t+h t} = \ell_t b_t^\varphi$

$\hat{y}_{t+h|t}$ valor observado no período t , ℓ_t nível (nível da série), b_t tendência (taxa de crescimento ou decrescimento), α suavização para o nível ($0 < \alpha < 1$), β suavização para a tendência ($0 < \beta < 1$), φ amortecimento que controla a redução da tendência ao longo do tempo ($0 < \varphi < 1$), e_t erro (ou resíduo).

Fonte: Adaptado de Hyndman e Athanasopoulos (2023).

Quando $\alpha = \beta$, o método de Holt equivale à 'suavização exponencial dupla' de Brown (Brown; Cane, 1959). Brown empregou um argumento de desconto para derivar suas equações de previsão, onde $1 - \alpha$ representa o fator de desconto comum aplicado aos componentes de nível e tendência.

Contudo, o método linear de Holt e sua variação exponencial apresentam a desvantagem de superestimar previsões, principalmente a longo prazo. Para resolver esta desvantagem, Gardner e McKenzie (1985)

propuseram uma modificação do método linear de Holt para permitir o "amortecimento" das tendências, apresentado na terceira linha da Tabela 2.3. Em conjunto com os parâmetros de suavização α e β este método também inclui um parâmetro de amortecimento $0 \leq \varphi \leq 1$. Se $\varphi = 1$, o método é idêntico ao método linear de Holt. Para valores entre 0 e 1, φ amortece a tendência para que ela se aproxime de uma constante em algum momento no futuro. De fato, as previsões convergem para $\ell_t + \varphi b_t / (1 - \varphi)$ como $h \rightarrow \infty$ para qualquer valor $0 \leq \varphi \leq 1$. Isso significa que as previsões de curto prazo são tendenciosas, enquanto as previsões de longo prazo são constantes (Hyndman; Athanasopoulos, 2023).

2.3.2 Metodologia de Box-Jenkins

A metodologia consiste em propor e ajustar modelos lineares estacionários ou não estacionários a uma série temporal (Albuquerque, 2019). A estratégia para a construção do modelo é baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseado nos próprios dados.

2.3.2.1 Modelos Autorregressivos(AR)

Conforme descrito por Padilha *et al.* (2013), o modelo autoregressivo (AR) é concebido com a ideia de que a observação atual em uma série temporal, representada por y_t , pode ser explicada como uma função das p observações anteriores, $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$. Nesse contexto, p determina o número de intervalos entre as observações passadas e a previsão da próxima. A formulação geral desse componente autoregressivo é expressa pela Equação 2.4:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + a_t \quad (2.4)$$

onde, φ_i denota os parâmetros da estrutura autoregressiva, com $i = 1, \dots, p$ indicando a ordem da estrutura. O termo a_t refere-se ao ruído branco associado ao período t , caracterizado por uma média zero e uma variância σ_a^2 . Em síntese, a equação modela y_t como uma combinação linear das p observações precedentes, cada uma multiplicada por seu peso correspondente, com a adição do termo de erro a_t . Essa abordagem captura a relação autorregressiva na série temporal, onde as observações passadas influenciam a predição da próxima observação.

2.3.2.2 Modelos de Médias Móveis (MA)

Segundo Souza *et al.* (2006), os modelos de médias móveis são formados por combinações lineares de ruído branco, a_t , ocorridos no período corrente e nos períodos passados. A estrutura de médias móveis geral é expressa pela Equação 2.5:

$$y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} + \epsilon_t \quad (2.5)$$

onde θ_i , representando a ordem da estrutura, com $i = 1, \dots, q$. A variável de interesse y_t no tempo t é modelada como a soma do ruído branco a_t e termos ponderados por θ_i de observações passadas. O ruído branco a_t possui média zero e variância constante σ_a^2 . A influência das observações passadas em y_t é determinada pelos parâmetros θ_i , onde um valor absoluto maior implica em um impacto mais significativo de a_{t-i} no valor de y_t . A ordem da estrutura, q , indica quantos períodos passados são considerados no modelo.

2.3.2.3 Modelos Autorregressivos e Média Móveis (ARMA)

Box e Jenkins (2008) formalizam a teoria da utilização de componentes autorregressivos e de médias móveis na modelagem de séries temporais utilizando-se de duas ideias básicas na criação de sua metodologia de construção de modelos: a parcimônia que é a utilização do menor número possível de parâmetros para se obter uma representação adequada do fenômeno em estudo e a interatividade, ou seja, a informação empírica é analisada teoricamente e o resultado deste estágio é confrontado com a prática sucessivas vezes, até que o modelo obtido seja satisfatório.

Conforme Figueredo (2008), o modelo ARMA é uma combinação do modelo autoregressivo (AR) com o modelo de médias móveis (MA), em que y_t é descrito pelos seus valores passados e pelo ruído branco corrente

e passado. A estrutura geral do modelo ARMA(p,q) é expressa pela Equação 2.6:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q} \quad (2.6)$$

onde φ_i são os parâmetros da estrutura autoregressiva, com $i = 1, \dots, p$, e θ_i são os parâmetros da estrutura de médias móveis, com $i = 1, \dots, q$. O termo a_t representa o ruído branco. Esse modelo permite capturar tanto a dependência linear de observações passadas quanto a influência de componentes de médias móveis na série temporal y_t .

Segundo Morettin e Toloi (2018), modelos com essa característica são bastante utilizados em algumas áreas, como em Economia, onde é natural o valor de alguma variável no instante t como função de valores defasados da mesma variável. Em outras áreas, como em Ciências, Física e Geografia, o interesse em modelos autorregressivos reside em outro aspecto que não seja o da previsão: deseja-se estimar o espectro do processo e os estimadores autorregressivos são utilizados para tal fim.

2.3.2.4 Modelos Autorregressivos Integrado de Média Móveis (ARIMA)

Conforme Bayer (2010), a classe de modelos mais utilizada na análise de séries temporais é a dos modelos ARIMA. Essa metodologia consiste em ajustar modelos autoregressivos integrados de médias móveis (ARIMA) a um conjunto de dados. Para y_t segue um processo autoregressivo e de médias móveis de ordens ARMA(p, q), dados pela Equação :

$$\varphi(B)y_t = c + \theta(B)a_t \quad (2.7)$$

onde a_t é um ruído branco com média zero e variância σ^2 , c é um parâmetro que permite ao processo ter média diferente de zero, o polinômio $B^k y_t = y_{t-k}$ é o operador retroativo, $\varphi(B)$ e $\theta(B)$ são os polinômios autoregressivo e de médias móveis, respectivamente, apresentados nas Equações 2.8 e 2.9:

$$\varphi(B)y_t = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \quad (2.8)$$

$$\theta(B)y_t = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.9)$$

Nestas equações, B representa o operador de defasagem, B^k denota um deslocamento no tempo de k períodos, e φ_i e θ_i são os parâmetros dos polinômios autoregressivo e de médias móveis, respectivamente. Essa abordagem permite modelar a dependência temporal da série temporal y_t e incorporar componentes de médias móveis e autoregressivas para uma análise mais abrangente.

Uma série que não apresenta característica de estacionariedade pode tornar-se estacionária pela utilização de um processo de diferenciação. Sendo assim, se uma série se tornar estacionária após d diferenças, a série é dita integrada (I) de ordem d . Com isso, o modelo ARMA integrado passa a ser denominado de modelo ARIMA. O modelo ARIMA(p,d,q) tem a forma da Equação 2.10:

$$\varphi(B)(1 - B)^d y_t = \theta(B)\epsilon_t \quad (2.10)$$

em que d é a ordem de integração, ou seja, o número de diferenças necessárias para torna a série estacionária (FAVA, 2000).

O modelo ARIMA foi proposto por Box-Jenkins em 1970 (Box; Al, 2015) sendo usado para processos não estacionários observáveis y_t que possuem algumas tendências claramente identificáveis: uma tendência constante (ou seja, média zero) é modelada por $d = 0$, uma tendência linear (ou seja, comportamento de crescimento linear) é modelada por $d = 1$ e uma tendência quadrática (ou seja, comportamento de crescimento quadrático) é modelada por $d = 2$. Um modelo ARIMA não sazonal é classificado como um “ARIMA (p, d, q)”, onde:

- p é o número de termos autorregressivos;

- d é o número de diferenças não sazonais; e
- q é o número de erros de previsão defasados na equação de previsão.

Nestes casos, o modelo ARIMA pode ser visto como uma “cascata” de dois modelos conforme mostrado pelas Equações 2.11 e 2.12. O primeiro é não estacionário

$$Y_t = (1 - L)^d X^t \quad (2.11)$$

enquanto o segundo é estacionário de sentido amplo:

$$\left(1 - \sum_{i=0}^p \varphi_i L^i\right) Y^t = \left(1 + \sum_{i=0}^p \varphi_i L^i\right) \varepsilon^t \quad (2.12)$$

onde Y_t é o valor real no período t , L_i é o operador de atraso, f é o parâmetro da parte autorregressiva e y é o parâmetro da parte da média móvel.

Uma vantagem do modelo ARIMA são suas propriedades estatísticas e a conhecida metodologia Box-Jenkins no processo de construção do modelo (Box; Al, 2015). Além disso, os modelos ARIMA podem implementar vários modelos de suavização exponencial. No entanto, as vantagens dos modelos ARIMA de previsão precisa ocorrem em um curto período, com as duas limitações básicas a seguir (Khashei; Bijari; Ardali, 2009):

1. Limitação linear: o valor futuro de uma variável é assumido como uma função linear de várias observações passadas e erros aleatórios em modelos ARIMA. Essa limitação linear faz com que a aproximação dos modelos ARIMA seja totalmente inadequada para um mecanismo subjacente não linear. Como os sistemas do mundo real são muitas vezes não lineares (Zhang; Patuwo; Hu, 1998), não é razoável supor que uma realização particular de uma determinada série temporal seja gerada por um processo linear.
2. Limitação de dados: os modelos ARIMA demandam grandes volumes de dados históricos para alcançar os resultados desejados. De acordo com essa restrição, é recomendável ter pelo menos 50, ou preferencialmente 100, ou mais, dados para obter resultados confiáveis (Box; Al, 2015). As incertezas decorrentes do ambiente dinâmico e do rápido desenvolvimento de novas tecnologias têm levado à necessidade de prever ocorrências futuras utilizando conjuntos de dados menores em curtos períodos.

Em um modelo ARIMA, o valor futuro de uma variável é assumido como uma função linear de várias observações passadas e erros aleatórios. Ou seja, o processo subjacente que gera a série temporal dada pela Equação 2.13:

$$y'_t = c + \varphi_1 y'_{t-1} + \varphi_2 y'_{t-2} + \cdots + \varphi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (2.13)$$

onde y'_t e ε_t são o valor real e o erro aleatório no período t , respectivamente, φ_i ($i = 1, 2, y, p$) são parâmetros autorregressivos e θ_j ($j = 0, 1, 2, y, q$) os de média móvel, p e q são números inteiros e muitas vezes referidos como ordens do modelo. Os erros aleatórios, ε_t , são considerados independentes e identicamente distribuídos com uma média de zero e uma variância constante de σ_2 .

A equação geral do modelo ARIMA envolve vários casos especiais importantes. Se $q = 0$, então ela torna-se um modelo *AR* de ordem p . Quando $p = 0$, o modelo se reduz a um modelo *MA* de ordem q . Uma tarefa central da construção do modelo ARIMA é determinar a ordem apropriada do modelo (p, q).

Ao combinar a diferenciação com a autorregressão e um modelo de média móvel, obtêm-se um modelo ARIMA não sazonal. Na Equação 2.13 y' é a série diferenciada (pode ter sido diferenciada mais de uma vez). Os “preditores” no lado direito incluem ambos os valores defasados de y_t e erros atrasados.

2.4 Teste de tendência e de quebra estrutural

O teste de Cox-Stuart é um dos testes recomendados para a verificação da existência de tendência, isto é, crescentes ou decrescentes, em séries temporais (Morettin; Toloi, 2018). O teste se baseia nas seguintes

hipóteses:

$$H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \forall_i : \text{não existe tendência};$$

$$H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \forall_i : \text{existe tendência};$$

A aplicação do teste consiste inicialmente em agrupar os dados em pares $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_n, Z_{n+c})$, tal que, n é o número de observações da série e $c = \frac{n}{2}$ para n par, ou $c = \frac{n+1}{2}$ caso n seja ímpar. Calcula-se as diferenças $(Z_{1+c} - Z_1), (Z_{2+c} - Z_2), \dots, (Z_{n+c} - Z_n)$ associando os sinais $-$ para valores negativos e $+$ para valores positivos das diferenças. Para haver tendência positiva, espera-se um número maior de valores positivos T_2 . Caso $T_2 > n - t$, rejeita-se H_0 , indicando a presença de tendência. Para $n \leq 20$, o valor t é obtido da distribuição binomial com parâmetro $p = \frac{1}{2}$. Para $n > 20$, utiliza-se a aproximação normal com média np e variância npq . O teste de Cox-Stuart é uma alternativa para a verificação de tendência de fácil interpretação. É um teste bastante utilizado na literatura, em diferentes áreas aplicadas (Nascimento *et al.*, 2008; Costa; Sáfiadi, 2010; Nogueira; Sáfiadi, 2011).

Na metodologia para determinar mudanças estruturais conforme defendido pelo autor, no modelo abrangente com k variáveis explicativas e um termo interceptador, emerge a existência de dois grupos distintos, representados por $g = 1$ e $g = 2$. O objetivo é examinar a igualdade das intercepções e inclinações entre esses dois grupos. A representação deste modelo é dada pela Equação 2.14:

$$\gamma = \beta_{g,0} + \beta_{g,1}x_1 + \beta_{g,2}x_2 + \dots + \beta_{g,k}x_k + u \quad (2.14)$$

onde, γ representa a variável dependente, $x_1, x_2, \dots, x_k$ são as variáveis explicativas, $\beta_{g,0}, \beta_{g,1}, \dots, \beta_{g,k}$ são os coeficientes associados a cada variável explicativa no grupo g , e u é o termo de erro.

A ideia principal é que a soma dos resíduos quadrados do modelo irrestrito pode ser obtida a partir de duas regressões separadas, uma para cada grupo: seja SRC_1 a soma dos resíduos quadrados obtidos pela estimativa para o primeiro grupo (n_1 observações estão envolvidas aqui); seja SRC_2 a soma dos resíduos quadrados obtidos da estimação do modelo usando o segundo grupo (n_2 observações estão envolvidas aqui). Agora, a soma dos resíduos quadrados para o modelo irrestrito é simplesmente $SRC_{nr} = SRC_1 + SRC_2$. A soma restrita dos resíduos quadrados é simplesmente o SRC de reunir os grupos e estimar uma única equação, SRC_p . Uma vez obtidos, a estatística F é calculada pela Equação 2.15:

$$F = \frac{[SRC_p - (SRC_1 + SRC_2)]}{SRC_1 + SRC_2} \cdot \frac{n - 2(k + 1)}{k + 1}. \quad (2.15)$$

Essa estatística é utilizada para determinar se há evidência estatística para rejeitar a hipótese nula de que os coeficientes das variáveis independentes são iguais nos dois grupos.

2.5 Teste de estacionariedade

Uma série temporal estacionária é aquela cujas propriedades estatísticas não dependem do momento em que a série é observada. Ao longo do tempo, a série mantém uma média constante e uma variância constante, sem tendências ou sazonalidades previsíveis a longo prazo. No entanto, pode haver comportamento cíclico, desde que esse ciclo não tenha uma duração fixa e previsível. Em resumo, uma série temporal estacionária exibe estabilidade em suas propriedades estatísticas ao longo do tempo.

De acordo com Hyndman e Athanasopoulos (2023), a diferenciação em séries temporais envolve o cálculo das diferenças entre observações consecutivas, sendo empregada para converter uma série que apresenta tendências ou sazonalidades em uma série estacionária. Essa técnica contribui para estabilizar a média da série, eliminando as alterações no nível que podem ser causadas por tendências ou sazonalidades. Dessa forma, a diferenciação é uma ferramenta para reduzir ou eliminar tendências e sazonalidades, transformando a série

em uma série estacionária, cujas propriedades estatísticas permanecem consistentes independentemente do momento das observações.

A série diferenciada é a mudança entre observações consecutivas na série original e pode ser escrita como a Equação 2.16:

$$y'_t = y_t - y_{t-1}. \quad (2.16)$$

A série diferenciada terá apenas $T - 1$ valores (T = número de observações), uma vez que não é possível calcular uma diferença e'_t para a primeira observação.

Quando a série diferenciada é um ruído branco, o modelo para a série original pode ser escrito como a Equação 2.17:

$$y_t - y_{t-1} = \varepsilon_t, \quad (2.17)$$

onde ε_t denota ruído branco. Reorganizar isso leva ao modelo de “caminhada aleatória”, dada pela Equação 2.18:

$$y_t = y_{t-1} + \varepsilon_t. \quad (2.18)$$

Os modelos de passeio aleatório são amplamente utilizados para dados não estacionários, particularmente dados financeiros e econômicos. Passeios aleatórios normalmente têm:

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

As previsões de um modelo de passeio aleatório são iguais à última observação, uma vez que os movimentos futuros são imprevisíveis e têm a mesma probabilidade de serem ascendentes ou descendentes.

Um modelo intimamente relacionado permite que as diferenças tenham uma média diferente de zero. Então:

$$y_t - y_{t-1} = c + \varepsilon_t \text{ ou } y_t = c + y_{t-1} + \varepsilon_t,$$

onde c é a média das mudanças entre observações consecutivas. Se c é positivo, então a variação média é um aumento no valor de y_t . Por isso, y_t tenderá a flutuar para cima. No entanto, se c é negativo, y_t tenderá a derivar para baixo.

Ocasionalmente, os dados diferenciados não parecerão estacionários e pode ser necessário diferenciar os dados uma segunda vez para obter uma série estacionária:

$$y''_t = y'_t - y'_{t-1}$$

$$y''_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

$$y''_t = y_t - 2y_{t-1} + y_{t-2}.$$

Nesse caso, y''_t terá $T - 2$ valores. Então, modelaríamos a “mudança nas mudanças” dos dados originais. Na prática, quase nunca é necessário ir além das diferenças de segunda ordem.

O Teste de Dickey-Fuller Aumentado (ADF) é uma ferramenta estatística essencial para avaliar a presença de raiz unitária em séries temporais, indicando se a tendência estocástica da série impede a convergência para um valor constante ao longo do tempo. A base matemática do teste ADF está fundamentada no conceito de um processo autoregressivo (AR) de primeira ordem, representado pela Equação 2.19:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad (2.19)$$

onde Y_t é a observação no tempo t , φ é o coeficiente de autoregressão, Y_{t-1} é a observação anterior, e ε_t é o termo de erro no tempo t .

A hipótese nula do teste ADF é que $\varphi = 1$, indicando a presença de raiz unitária e a não estacionariedade da série temporal. A hipótese alternativa, $\varphi < 1$, sugere que a série torna-se estacionária após diferenciação. Este teste aprimora a estatística do Dickey-Fuller clássico ao incluir termos defasados adicionais para controlar possíveis correlações residuais. A estatística de teste é então comparada a valores críticos de distribuição para decidir se a hipótese nula pode ser rejeitada.

O teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) é uma ferramenta estatística utilizada para testar a estacionariedade de uma série temporal. A estacionariedade é uma propriedade importante em análise de séries temporais, indicando que as propriedades estatísticas da série não mudam ao longo do tempo. O teste KPSS é projetado para complementar o teste de Dickey-Fuller, sendo mais comumente usado para testar a presença de raiz unitária (não estacionariedade) em uma série temporal.

H_0 : A série é estacionária ao redor de uma tendência determinística. H_A : A série não é estacionária.

A estatística de teste do KPSS é baseada em uma regressão que modela a série temporal como uma tendência determinística mais um componente estocástico. A forma básica da regressão é dada pela Equação 2.20:

$$Y_t = \mu_t + \varepsilon_t, \quad (2.20)$$

onde Y_t é o valor da série temporal no tempo t , μ_t é a tendência determinística e ε_t é o componente estocástico.

A estatística de teste KPSS é então calculada comparando a soma cumulativa ponderada dos quadrados dos resíduos (erros) da regressão com uma distribuição crítica. Se a estatística de teste é maior que o valor crítico, rejeitamos a hipótese nula de estacionariedade.

2.6 Testes de diagnóstico de resíduos

Os testes de análise dos resíduos verificam se um modelo estatístico se ajusta adequadamente aos dados, examinando as diferenças entre os valores observados e previstos. Eles identificam padrões nos resíduos que podem indicar violações das suposições do modelo, incluindo normalidade, homocedasticidade e independência. A combinação dos testes de Ljung-Box, ARCH e Jarque-Bera permite avaliar a adequação do modelo estatístico e a qualidade dos resíduos em uma análise de séries temporais ou modelagem de dados (Mattos *et al.*, 2021).

O teste de Ljung-Box é utilizado para avaliar a presença de autocorrelação dos resíduos de um modelo estatístico ou séries temporais. A hipótese nula é de que os resíduos são independentes e não autocorrelacionados até um certo número de lags (atrasos). A estatística do teste é calculada somando os quadrados das autocorrelações parciais dos resíduos até o atraso k . A fórmula do teste de Ljung-Box é dada pela Equação 2.21:

$$Q = T(T+2) \sum_{k=1}^{\ell} (T-k)^{-1} r_k^2, \quad (2.21)$$

onde n é o tamanho da amostra (número de observações), h é o número de lags considerados e ρ^2 é a autocorrelação parcial dos resíduos no lag k . Quando o p-valor $< 0,05$ indica que os resíduos não são independentes, ou seja, ainda possuem autocorrelação, sugerindo que o modelo pode não estar capturando totalmente a estrutura temporal dos dados (Ljung; Box, 1978).

O teste de Heterocedasticidade Condicional Autorregressiva (ARCH) é usado para verificar a presença de heteroscedasticidade condicional nos resíduos de um modelo. A hipótese nula é de que os resíduos têm variância constante ao longo do tempo. O teste é realizado ajustando um modelo de regressão dos quadrados dos resíduos em relação aos quadrados dos resíduos atrasados até o lag k , com fórmula dada pela Equação 2.22:

$$ARCH(k) : \hat{e}_t^2 = \gamma_0 + \sum_{i=1}^k \gamma_i \hat{e}_{t-i}^2, \quad (2.22)$$

onde k é o número de lags considerados, $\hat{e}_t$ é o resíduo no período t e $\gamma_0, \gamma_1, \dots, \gamma_k$ são os coeficientes do modelo de regressão estimados. Se o p-valor for maior que o nível de significância, podemos inferir que os resíduos são homoscedásticos, caso contrário, podemos inferir que são heteroscedásticos (Engle, 1982).

O teste de Jarque-Bera é utilizado para verificar a normalidade dos resíduos de um modelo estatístico ou séries temporais. A hipótese nula é de que os resíduos têm distribuição normal. O teste é baseado nas medidas de assimetria e curtose dos resíduos. A fórmula é dada pela Equação 2.23:

$$JB = \frac{1}{n} \left(S^2 + \frac{1}{4} K^2 \right), \quad (2.23)$$

onde n é o tamanho da amostra (número de observações), S é a medida de assimetria dos resíduos, K é a medida de curtose dos resíduos. Quando o valor-p associado ao teste for maior que um determinado nível de significância (geralmente 0,05 ou 5%), então não há evidências suficientes para rejeitar a hipótese nula, e pode-se afirmar que os dados se assemelham a uma distribuição normal (Jarque; Bera, 1987).

2.7 Critérios de seleção de modelos

Os critérios de informação permitem comparar modelos com diferentes complexidades, balanceando o ajuste do modelo com sua complexidade. De acordo com Hyndman e Athanasopoulos (2023), o Critério de Informação de Akaike (AIC) é uma medida estatística intimamente relacionada com a seleção de modelos de previsão. O AIC é calculado usando a Equação 2.24:

$$AIC = T \cdot \left(\frac{SSE}{T} \right) + 2 \cdot (k + 2) \quad (2.24)$$

onde T representa o número de observações utilizadas para a estimativa, e k é o número de preditores no modelo. A variável SSE denota o erro quadrático médio obtido pelo modelo com T observações.

É importante destacar que diferentes pacotes de computador podem adotar definições levemente distintas para o AIC, porém, todos devem conduzir à seleção do mesmo modelo. O termo $k + 2$ na equação surge devido aos k coeficientes para os preditores, mais um intercepto e a variância dos resíduos. Essa adição tem o propósito de penalizar o ajuste do modelo (SSE) em relação ao número de parâmetros que precisam ser estimados.

Geralmente, o modelo com o menor valor de AIC é considerado o mais adequado para fins de previsão. Notavelmente, em situações com um grande número de observações (T), minimizar o AIC equivale a minimizar o Erro Quadrático Médio (CV) do modelo. Portanto, o AIC representa uma ferramenta valiosa para selecionar o modelo mais apropriado, considerando tanto a complexidade quanto a capacidade de previsão.

Para situações em que o número de observações (T) é pequeno, o Critério de Informação de Akaike (AIC) pode tender a selecionar modelos com um número excessivo de preditores. Com o intuito de corrigir esse viés, foi desenvolvida uma versão ajustada do AIC, denominada AICc. Essa correção considera o tamanho da amostra e é calculada conforme a Equação 2.25:

$$AICc = AIC + \frac{2(k + 2)(k + 3)}{T - k - 3}, \quad (2.25)$$

onde k representa o número de preditores no modelo, e AIC é o Critério de Informação de Akaike calculado anteriormente.

Assim como no AIC, o objetivo é minimizar o valor do AICc para selecionar o modelo mais apropriado para a previsão. A correção oferecida pelo AICc é especialmente relevante em situações em que o conjunto de

dados é limitado, permitindo uma escolha mais adequada entre a complexidade do modelo e sua capacidade de generalização. Dessa forma, o AICc se torna uma medida valiosa para a seleção criteriosa de modelos em cenários com amostras pequenas.

Outra medida relacionada é o Critério de Informação Bayesiana de Schwarz (BIC), frequentemente abreviado como SBIC ou SC. O BIC é definido pela Equação 2.26:

$$BIC = T \cdot \log\left(\frac{SSE}{T}\right) + (k + 2) \cdot \log(T), \quad (2.26)$$

onde T representa o número de observações usadas para estimativa e k é o número de preditores no modelo. Similarmente ao AIC, o objetivo é minimizar o valor do BIC para selecionar o melhor modelo.

O BIC apresenta uma penalização mais rigorosa para o número de parâmetros do que o AIC. Como resultado, o modelo escolhido pelo BIC é geralmente o mesmo selecionado pelo AIC ou um com menos termos. Isso ocorre porque o BIC coloca uma maior ênfase na simplicidade do modelo, evitando a inclusão excessiva de preditores. Em situações em que o número de observações (T) é grande, minimizar o BIC é aproximadamente equivalente a realizar v validações cruzadas quando $v = T \left(1 - \frac{1}{\ln(T)-1}\right)$ (Emiliano, 2013).

Portanto, assim como o AIC e o AICc, o BIC é uma ferramenta útil para a seleção adequada de modelos, fornecendo uma abordagem mais rigorosa para a escolha de um modelo com base na compensação entre ajuste e complexidade. Os procedimentos recomendados por Hyndman e Athanasopoulos (2023) para selecionar preditores são úteis quando o modelo é usado para previsão. Consequentemente, há recomendações que uma das estatísticas AICc, AIC ou BIC seja usada, cada uma das quais tendo como objetivo a previsão. Se o valor de T é grande o suficiente, todos eles levarão ao mesmo modelo. Caso contrário, o valor AICc deve ser usado para selecionar o modelo de previsão (Pérez, 2023).

2.8 Métricas de avaliação de desempenho

Na avaliação do desempenho preditivo, diversas medidas são comumente utilizadas para avaliar a precisão de um método de previsão. Essas medidas são baseadas no período de validação, o qual oferece uma base mais objetiva para avaliar a precisão preditiva, pois não é utilizado para selecionar preditores ou estimar parâmetros do modelo. As métricas usadas nesta pesquisa para comparar o desempenho dos modelos são o Erro Médio Absoluto (MAE), a Raiz do Erro Médio Quadrático (RMSE) e o Erro Percentual Médio Absoluto (MAPE), dadas respectivamente pelas Equações 2.27, 2.28 e 2.29:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (2.27)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (2.28)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100\%, \quad (2.29)$$

onde n representa o número de observações, y_i as observações reais e $\hat{y}_i$ as previsões. O MAE e o RMSE são medidas que dependem da escala, sendo o MAE mais fácil de interpretar, pois minimizá-lo leva a previsões da mediana, enquanto minimizar o RMSE leva a previsões da média.

O MAPE é provavelmente a medida de erro mais amplamente empregada (Tsolkas *et al.*, 2023; Alamri; Khan, 2023; Javanmard; Ghaderi, 2022), e ao contrário do MAE e do RMSE, não depende da escala porque se baseia em erros percentuais. Assim, tem a vantagem de ser uma métrica sem unidade. No entanto, também requer algumas considerações críticas. Por exemplo, pode levar a previsões tendenciosas porque fornece resultados infinitos ou indefinidos quando um ou mais pontos de dados da série temporal são iguais a 0 e

aplica uma penalidade mais pesada a erros negativos (ou seja, quando os valores previstos são maiores que os valores reais) do que em erros positivos (Perone, 2022).

2.9 Estudos relacionados

A revisão de literatura visa identificar, selecionar, avaliar e sintetizar as evidências relevantes disponíveis na área de conhecimento definida. A primeira etapa da coleta de dados consistiu na escolha da base de dados online a ser utilizada: *Scopus*. Após foi realizada a escolha das palavras-chave em inglês para uma melhor abrangência de resultados, sendo elas: *time series*, *forecast*, *ARIMA*, *exponential smoothing*, *GHG* e *CO₂ emissions*.

A busca, realizada no dia 06/06/2023, foi feita no título, resumo e palavras-chave de artigos ou conferências publicados nos últimos 10 anos, sendo utilizadas as *queries*:

- ('time series' OR 'forecast') AND ('exponential smoothing' OR 'ARIMA'), que conduziu a 621 documentos, e,
- ('time series' OR 'forecast') AND 'exponential smoothing' AND 'ARIMA' AND ('GHG' OR 'CO₂ emissions'), que conduziu a 96 documentos.

Na listagem de documentos selecionados por cada *query* foram escolhidos os três documentos mais citados e os três documentos mais recentes que tivessem relação com o estudo desenvolvido, conforme Tabela 2.4.

Os estudos apresentados na Tabela 2.4 têm como objetivo principal realizar previsões utilizando técnicas aplicadas a conjuntos de dados históricos em diversas áreas. Destaca-se o uso dos modelos de suavização exponencial e ARIMA, amplamente empregados na análise e previsão de séries temporais. Os exemplos de aplicação dessas técnicas abrangem a previsão de séries temporais relacionadas às emissões de GEE. Essas abordagens têm se mostrado eficazes para obter *insights* valiosos sobre as tendências atuais e futuras dos indicadores de GEE, contribuindo para a tomada de decisões em relação à política ambiental. A seguir apresenta-se mais detalhes sobre os estudos relacionados a esta pesquisa.

Os autores Liu *et al.* (2014) investigaram se a China poderia alcançar sua meta de reduzir a intensidade de emissão de CO₂ por unidade do PIB até 2020. Eles construíram um modelo combinado de previsão, utilizando a modelagem cinzenta (GM), ARIMA e regressão polinomial de segunda ordem (SOPR). O modelo foi otimizado com o algoritmo de otimização de enxame de partículas (PSO). Os resultados mostraram que, se o setor termelétrico a carvão da China continuar expandindo na taxa atual, será extremamente desafiador atingir as metas estabelecidas pelo governo chinês. O estudo também identificou uma faixa de geração de energia termelétrica a carvão que permitiria atender às metas de redução de emissões.

Os autores Kale e Pohekar (2014) abordaram a importância da previsão da demanda de eletricidade para fornecer soluções sustentáveis no setor de energia. Eles utilizaram o modelo LEAP para prever a demanda de eletricidade para o estado de Maharashtra, na Índia, até 2030. O método de suavização exponencial de Holt foi aplicado para obter taxas de crescimento adequadas. Três cenários foram gerados, incluindo *Business as Usual* (BAU), *Energy Conservation* (EC) e *Renewable Energy* (REN). A análise subsequente considerou aspectos como energia, impacto ambiental e custo. Os resultados destacaram o crescimento projetado da demanda de eletricidade e as emissões de GEE para cada cenário, fornecendo informações valiosas para políticas energéticas e ambientais.

O estudo dos autores Sen, Roy e Pal (2016) explorou a previsão de consumo de energia e emissão de GEE para uma organização de produção de ferro gusa na Índia. Eles aplicaram o modelo ARIMA para prever o consumo de energia e a emissão de GEE, identificando os melhores modelos ajustados. Os resultados revelaram a utilidade desses modelos para fornecer informações (percepções) sobre as tendências atuais e futuras desses indicadores, o que pode ajudar os gestores a tomarem decisões informadas para uma melhor política ambiental.

Tabela 2.4 – Características dos estudos selecionados

Autor, ano	Citações	Objetivo	Dados	Técnicas	Conclusões
KALE; POHEKAR, 2014	41	Prever a demanda de eletricidade para o ano-alvo de 2030 no estado de Maharashtra, na Índia, a fim de fornecer soluções sustentáveis para os problemas de eletricidade.	Demanda de eletricidade em Maharashtra nos anos base 2012 e alvo 2030, bem como o PIB e a elasticidade da demanda.	O modelo LEAP é utilizado para a previsão da demanda de eletricidade. A técnica de suavização exponencial de Holt é empregada para determinar as taxas de crescimento adequadas.	Em 2030, a demanda projetada de eletricidade para os cenários BAU e REN aumentou em 107,3% em relação a 2012, enquanto a demanda de eletricidade no cenário EC aumentou 54,3%. As estimativas de emissões de GEE em 2030 são 245,2% maiores para o cenário BAU, 152,4% maiores para o cenário EC e 46,2% menores para o cenário REN, em comparação com o ano base. Uma análise de sensibilidade foi realizada para estudar o efeito sobre o custo total dos diferentes cenários.
LIU <i>et al.</i> , 2014	130	Prever a intensidade das emissões de CO ₂ per capita em 2020 na China, a fim de avaliar a possibilidade de alcançar a meta de redução de intensidade de CO ₂ por unidade do PIB anunciada na Conferência de Copenhague, com foco no desenvolvimento de usinas termelétricas a carvão da China.	Relacionados à geração de energia termelétrica a carvão na China e às emissões de CO ₂ .	Foi construído um modelo combinado de previsão com um modelo de previsão cinza GM(1,1), um modelo ARIMA e um modelo de SOPR. A precisão da previsão foi aprimorada otimizando três coeficientes dos modelos individuais mencionados acima com a técnica de PSO.	Os resultados mostram que, até 2020, a geração de energia termelétrica atingirá 7258,83 bilhões de kW h, as emissões de CO ₂ alcançarão 17379,90 milhões de toneladas e a intensidade de emissão de CO ₂ será de 0,21 kg por unidade de Yuan, o que representa quase o dobro de 40-45% do nível de 2005. É alertado que a situação de cumprimento das metas estabelecidas pelo governo chinês é extremamente séria se a geração de energia termelétrica a carvão da China continuar expandindo na taxa atual. Para atender à meta de redução de emissões, a geração de energia termelétrica de 2020 deve ser controlada em um intervalo entre 3801,45 e 4492,62 bi/kW h.
SEN; ROY; PAL, 2016	142	Prever o consumo de energia e as emissões de GEE para uma organização de fabricação de ferro-gusa na Índia, a fim de fornecer informações relevantes aos gestores e auxiliar na formulação de políticas ambientais eficientes.	Relacionados ao consumo de energia e emissões de GEE para o setor de fabricação de ferro e aço na Índia.	Foi aplicado o ARIMA, com a identificação do modelo ARIMA (1,0,0) × (0,1,1) como o melhor ajustado. Também foi utilizado o modelo ARIMA, sendo o ARIMA (0,1,4) × (0,1,1) identificado como o melhor ajustado.	Os modelos ARIMA permite uma previsão precisa do consumo de energia e das emissões de GEE, fornecendo informações relevantes para a prática de gestão ambiental eficiente. Os resultados indicam que os modelos ARIMA selecionados se assemelham aos de um modelo sazonal de tendência aleatória, mas com uma suavização mais eficiente dos padrões sazonais e de tendência tanto para o consumo de energia quanto para as emissões de GEE
JAVANMARD; GHADERI, 2022	13	Apresentar uma abordagem híbrida de aprendizado de máquina e programação matemática para prever a emissão de GEE, com alta precisão de previsão e utilizando poucos dados.	Setor de energia que afetam as emissões de GEE de 1990 a 2018 no Irã.	A previsão de emissão de cada gás foi realizada por nove algoritmos de ANN, AR, ARIMA, SARIMA, SARIMAX, RF, SVR, KNN e LSTM.	A emissão de GEE no Irã excederá 1096 Mt/ano em 2028. A melhoria foi de pelo menos 31,7% com o algoritmo PSO e 12,8% com o algoritmo GWO. Com base nessas previsões de alta precisão, o estudo conclui que as emissões de GEE estufa no Irã deverão ultrapassar 1096 Mt/ano em 2028.
YÖRÜK <i>et al.</i> , 2023		Examinar o planejamento energético da Turquia em termos de planejamento estratégico, política energética, planejamento de produção de eletricidade, seleção de tecnologia e políticas ambientais.	Dados históricos e projeções futuras de demanda de eletricidade na Turquia.	Diversas técnicas de previsão são utilizadas para estimar a demanda de eletricidade ao longo do horizonte de planejamento, incluindo métodos de regressão, suavização exponencial, método de Winter e modelos ARIMA, sendo escolhido o melhor método com base em diversas medidas de erro.	Um aumento significativo no uso de opções de geração de energia renovável, como solar, eólica e hidrelétrica, enquanto o uso de combustíveis fósseis na geração de energia diminuirá consideravelmente. As descobertas sugerem que são necessários investimentos graduais em estratégias de produção de eletricidade baseadas em energias renováveis para eventualmente substituir as alternativas de combustíveis fósseis. Essa mudança não apenas reduz custos de investimento, operação e manutenção, mas também diminui as emissões a longo prazo.
ALAMRI; KHAN, 2023	4	Analisar e investigar as emissões de CO ₂ e desenvolver um modelo ótimo para prever a emissão de CO ₂ . Além disso, fornecer recomendações e estratégias para mitigar as mudanças climáticas através da redução das emissões de CO ₂ .	Coletados e sintetizados a partir de várias fontes usando um web crawler.	Diversas técnicas de aprendizado de máquina e modelagem estatística foram implementadas e avaliadas, incluindo Suavização Exponencial, Transformers, Rede Convolutacional Temporal (TCN) e análise de expansão neural para séries temporais interpretáveis.	O modelo N-BEATS demonstrou um desempenho geral melhor para prever as emissões de CO ₂ na Arábia Saudita em comparação com os outros modelos avaliados. Além disso, o artigo fornece recomendações e estratégias para mitigar as mudanças climáticas, reduzindo as emissões de CO ₂ .

PIB - Produto Interno Bruto, GM - Modelagem Cinzenta, SOPR - Regressão Polinomial de Segunda Ordem, PSO - Otimização de Enxame de Partículas, BAU - Business as Usual, REN - Energia Renovável, EC - Conservação de Energia, TCN - Temporal Convolutional Networ, ANN - Rede Neural Artificial, AR - Autorregressão, ARIMA - Média Móvel Integrada Autorregressiva, SARIMA - Média Móvel Integrada Autorregressiva Sazonal, SARIMAX - Média Móvel Integrada Autorregressiva Sazonal com Covariáveis Exógenas, RF - Floresta Aleatória, SVR - Regressão de Vetor de Suporte, KNN - K-Vizinhos Mais Próximos e LSTM - Memória de Longo Prazo.

Fonte: Elaboração própria a partir dos estudos selecionados na revisão de literatura.

Dankwa *et al.* (2021) modelam as tendências e os padrões de precipitação previstos de 2017 a 2035 na área de Manga-Bawku, nordeste de Gana, usando dados de precipitação de 1976 a 2016. Os modelos de suavização exponencial sazonal simples e ARIMA (0,1,1) foram empregados. Os resultados obtidos no teste de eficiência do modelo de previsão mostraram um valor R-quadrado estacionário de 0,698 e 0,669, RMSE de 48,775 e 50,717, e BIC normalizado de 7,800 e 7,904 para os modelos de suavização exponencial e ARIMA (0,1,1), respectivamente. Indicando que os modelos são adequados para a análise de séries temporais. O estudo mostrou ainda que a componente sazonal foi uniforme numa base anual, o que significa que foi um impacto aditivo em vez de um efeito multiplicativo. A previsão de chuva prevê um declínio nos níveis de chuva nos próximos 19 anos.

Yörük *et al.* (2023) propõem um modelo de otimização misto especificamente adaptado para o planejamento estratégico de eletricidade na Turquia, visando não apenas a minimização de custos, mas também incorporando diferentes políticas de nível estratégico, como redução de emissão de CO_2 . Para prever com precisão a demanda de eletricidade, os pesquisadores empregam várias técnicas de previsão, incluindo métodos ARIMA e suavização exponencial, selecionando o método mais adequado com base em diferentes medidas de erro. Os resultados destacam um aumento significativo na utilização de fontes de energia renováveis, como alternativas solares, eólicas e hidrelétricas, juntamente com uma diminuição acentuada na dependência de combustíveis fósseis para geração de energia.

O estudo de Alamri e Khan (2023), investigou a previsão de emissões de CO_2 na Arábia Saudita e desenvolveu um modelo ótimo para prever essas emissões. Os pesquisadores implementaram e avaliaram várias técnicas de modelagem estatística e de aprendizado de máquina, os modelos como suavização exponencial e ARIMA, para prever futuras emissões de CO_2 . O modelo N-BEATS demonstrou um desempenho geralmente melhor para a previsão de emissões de CO_2 na Arábia Saudita em comparação com outros modelos. O estudo também ofereceu recomendações e estratégias para mitigar as mudanças climáticas, reduzindo as emissões de CO_2 .

Uma nova busca utilizando as mesmas *queries* localizou-se estes dois artigos, apresentados na Tabela 2.5, que apresentam características semelhantes a esta pesquisa. Ambos comparam modelos utilizados para realizar previsões de séries temporais de dados das emissões de CO_2 .

Tabela 2.5 – Características dos estudos selecionados

Autor, ano	Citações	Objetivo	Dados	Técnicas	Conclusões
Alam, Alarjani 2021	9	Comparar modelos de previsão de emissões de CO_2 nos países do Golfo para fornecer informações sobre o estado atual das previsões e auxiliar no desenvolvimento de planos estratégicos.	Emissões per capita de CO_2 dos países do Golfo, coletados da EIA dos EUA entre 1960 e 2014.	Análise de séries temporais, modelagem estatística e coleta de dados. Diferentes modelos, como ARIMA, redes neurais artificiais e suavização exponencial de Holt-Winters, foram utilizados para prever as emissões de CO_2 . Os resultados foram comparados e avaliados utilizando métricas de desempenho, como RMSE, MAE e MAPE, para determinar a precisão das previsões de cada modelo.	A conclusão da pesquisa é que, dos modelos analisados - incluindo ARIMA, redes neurais artificiais e suavização exponencial de Holt-Winters -, a rede neural artificial apresentou o melhor desempenho na previsão das emissões de CO_2 nos países do Golfo. Isso sugere que a abordagem de aprendizado de máquina pode ser mais eficaz do que os métodos estatísticos tradicionais para esse tipo de previsão. Essa conclusão pode ter implicações importantes para o desenvolvimento de estratégias de mitigação e planejamento ambiental nesses países.
Javanmard Ghaderi 2022	36	Criar um método preciso de previsão das emissões de GEE no Irã até 2028, combinando ML e programação matemática, e compará-lo com métodos convencionais de ML.	Os GEE do setor de energia no Irã, abrangendo o período de 1990 a 2018.	Aplicou nove algoritmos de aprendizado de máquina para prever essas emissões e, em seguida, modelos matemáticos otimizados para estender as previsões até 2028. A precisão dessas previsões foi comparada com métodos tradicionais, e as relações entre os dados foram analisadas com regressão stepwise.	A conclusão do estudo indicou que a abordagem híbrida de aprendizado de máquina e programação matemática, otimizada por algoritmos metaheurísticos, melhorou significativamente a precisão das previsões das emissões de gases de efeito estufa no Irã. Comparado aos métodos tradicionais de aprendizado de máquina, o modelo híbrido apresentou uma melhoria na precisão das previsões, projetando que as emissões de gases de efeito estufa no Irã excederiam 1096 Mt/ano em 2028.

ARIMA - Média Móvel Integrada Autorregressiva, EIA - Energia dos Estados Unidos, RMSE - Raiz do Erro Médio Quadrático, MAE - Erro Médio Absoluto, MAPE - Erro Percentual Médio Absoluto, GEE - Gases de Efeito Estufa, ML - *Machine Learning*.

Fonte: Elaboração própria a partir dos estudos selecionados na revisão de literatura.

Os pesquisadores Alam e Alarjani (2021) adotam ARIMA, rede neural artificial (RNA) e suavização exponencial Holt-Wintersy para antecipar anualmente as emissões de CO_2 nos países do Golfo. Os resultados da pesquisa indicam que o pesquisador deve se concentrar nos modelos RNA, ARIMA (1,1,1), suavização exponencial de Holt-Winters, ARIMA (1,1,2) e ARIMA (2,1,2) para prever dados de séries temporais sobre emissões de CO_2 nestes países.

O estudo de Javanmard e Ghaderi (2022) propõe uma abordagem híbrida de aprendizado de máquina e programação matemática para prever com alta precisão as emissões de GEE com base em dados limitados. Estes são relacionados ao setor de energia e às taxas de emissões de gases de efeito estufa no Irã de 1990 a 2018. A previsão das emissões foram realizadas por nove algoritmos: ANN, AR, ARIMA, SARIMA, SARIMAX, RF, SVR, KNN e LSTM. A precisão das previsões foi avaliada por meio de cinco indicadores. Em seguida, os resultados dos algoritmos de aprendizado de máquina foram utilizados como entrada para um modelo matemático, otimizado pelos algoritmos PSO e GWO para prever as emissões de gases de efeito estufa até 2028. A precisão das previsões foi comparada com os resultados dos algoritmos de aprendizado de máquina, e o modelo de otimização demonstrou melhorias significativas, com pelo menos 31,7% de aumento na precisão usando o algoritmo PSO e 12,8% usando o algoritmo GWO. Com base nas previsões altamente precisas, estima-se que as emissões de gases de efeito estufa no Irã excedam 1096 Mt/ano em 2028.

Os estudos relacionados enfatizam a importância dos métodos de previsão na análise e tomada de decisões relacionadas a emissões de GEE. Diversos modelos, dentre eles ARIMA e suavização exponencial, forneceram informações valiosas sobre as tendências atuais e futuras desses indicadores. Os resultados destacam a necessidade de ações e estratégias para enfrentar os desafios ambientais, considerando diferentes cenários, como o crescimento do setor termelétrico a carvão na China. A previsão da tendência das emissões de GEE possibilita a formulação de políticas ambientais adequadas, com foco em soluções sustentáveis. Em resumo, os métodos de previsão são cruciais para compreender as tendências e tomar decisões em prol de uma política ambiental melhor, visando a redução das emissões de GEE e a promoção da sustentabilidade energética.

3 Metodologia

Este capítulo visa fornecer uma descrição concisa e clara das etapas, processos e metodologia empregadas na elaboração desta pesquisa. O processo iniciou-se com uma revisão da literatura, seguida por um estudo de caso para aplicar os conceitos e analisar os resultados obtidos. A metodologia científica adotada para o desenvolvimento desta dissertação fundamenta-se na revisão de artigos científicos e livros, combinada com a abordagem de estudo de caso.

3.1 Caracterização da Área de Estudo

O escopo deste estudo é restrito ao estado do Amazonas, localizado na Região Norte do Brasil, o qual se destaca como o maior em extensão territorial, abrangendo 18.3% (1.559 mil km²) do território brasileiro (IBGE, 2022). Conforme os dados coletados no Censo 2022, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), possui população de 3.941.175 pessoas, o que representa cerca de 1,9% do total da população do Brasil. Essas informações constituem elementos fundamentais para o enquadramento deste estudo em um contexto científico.

Figura 3.1 – Mapa do Brasil com destaque para os limites do estado do Amazonas



Fonte: Adaptado da (Wikipedia contributors, 2024)

3.2 Dados e sua análise

A análise foi realizada a partir de observações anuais das emissões de CO₂e no estado do Amazonas, no período de 1990 a 2021. Nesta análise foram consideradas as emissões oriundas da produção industrial, tratamento de resíduos, produção de energia, agropecuária, mudança do uso da terra e floresta (MUTF) e da série total, somatória de todas as emissões do estado. Os dados estão disponíveis no SEEG (SEEG, 2021).

Visando compreender o comportamento das emissões de CO_2e , procedeu-se uma análise exploratória dos dados referentes tanto ao estado do Amazonas quanto ao Brasil, realizando uma comparação dos achados. Essa análise exploratória visou extrair informações iniciais dos dados, compreender suas características e detectar padrões, tendências e possíveis correlações entre as variáveis pertinentes. Essa fase preliminar precede a aplicação de técnicas mais complexas de modelagem e inferência estatística.

Para realizar a análise das séries, cada uma das séries foram divididas em duas partes, uma para o ajuste do modelo (1990-2015, 26 observações) e outra para o teste (2016-2021, 6 observações). A partir do intervalo de ajuste, cada série temporal foi analisada utilizando-se os seguintes métodos de Suavização Exponencial: o método com tendência aditiva e erro aditivo (AA), o método com tendência aditiva e erro multiplicativo (MA), o método com tendência multiplicativa e erro multiplicativo (MM), o método com tendência aditiva amortecida e erro aditivo (AA_d), o método com tendência aditiva amortecida e erro multiplicativo (MA_d) e o método com tendência multiplicativa amortecida e erro multiplicativo (MM_d).

Para cada uma das previsões geradas pelos métodos de suavização exponencial, foram comparadas com os valores destinados ao teste a fim de medir a acurácia do modelo. A partir da acurácia dos modelos foram selecionados os 2 (dois) modelos mais parcimoniosos. Também foi feita a análise dos resíduos para verificar a autocorrelação, a heterocedasticidade e se a série possuía uma distribuição normal, através dos testes de Ljung-Box (Ljung; Box, 1978), Jarque-Bera (Jarque; Bera, 1987) e ARCH (Engle, 1982), respectivamente. Por fim, desses modelos, foi feita a seleção do modelo que apresentou maior precisão. As métricas RMSE, MAE e MAPE foram utilizadas para avaliar o desempenho dos modelos em termos de precisão das previsões (Perone, 2022).

O procedimento de modelagem para ajustar um modelo ARIMA a um conjunto de dados de séries temporais não sazonais destinadas ao ajuste, apresentado na Figura 3.2, iniciou-se pela plotagem dos dados para identificar observações incomuns, como tendências, padrões ou dados discrepantes que possam afetar a análise. Em seguida, caso necessário, os dados podem ser transformados para estabilizar a variância.

Após isso, verificou-se a estacionariedade dos dados, uma vez que os modelos ARIMA exigem que os dados sejam estacionários. Se não forem, aplicam-se diferenciações de primeira ordem até que os dados atinjam a estacionariedade, ou seja, até que não apresentem mais tendências.

A análise da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF) foi então realizada para determinar a ordem dos termos autoregressivos (AR) e de médias móveis (MA) do modelo ARIMA (Morettin; Toloi, 2018). Durante a etapa de estimação do modelo ARIMA, empregaram-se critérios de informação, como AIC, AICc e BIC (Akhter; Hassan; Abbas, 2020), a fim de avaliar e selecionar o modelo mais adequado com base na complexidade e ajuste aos dados.

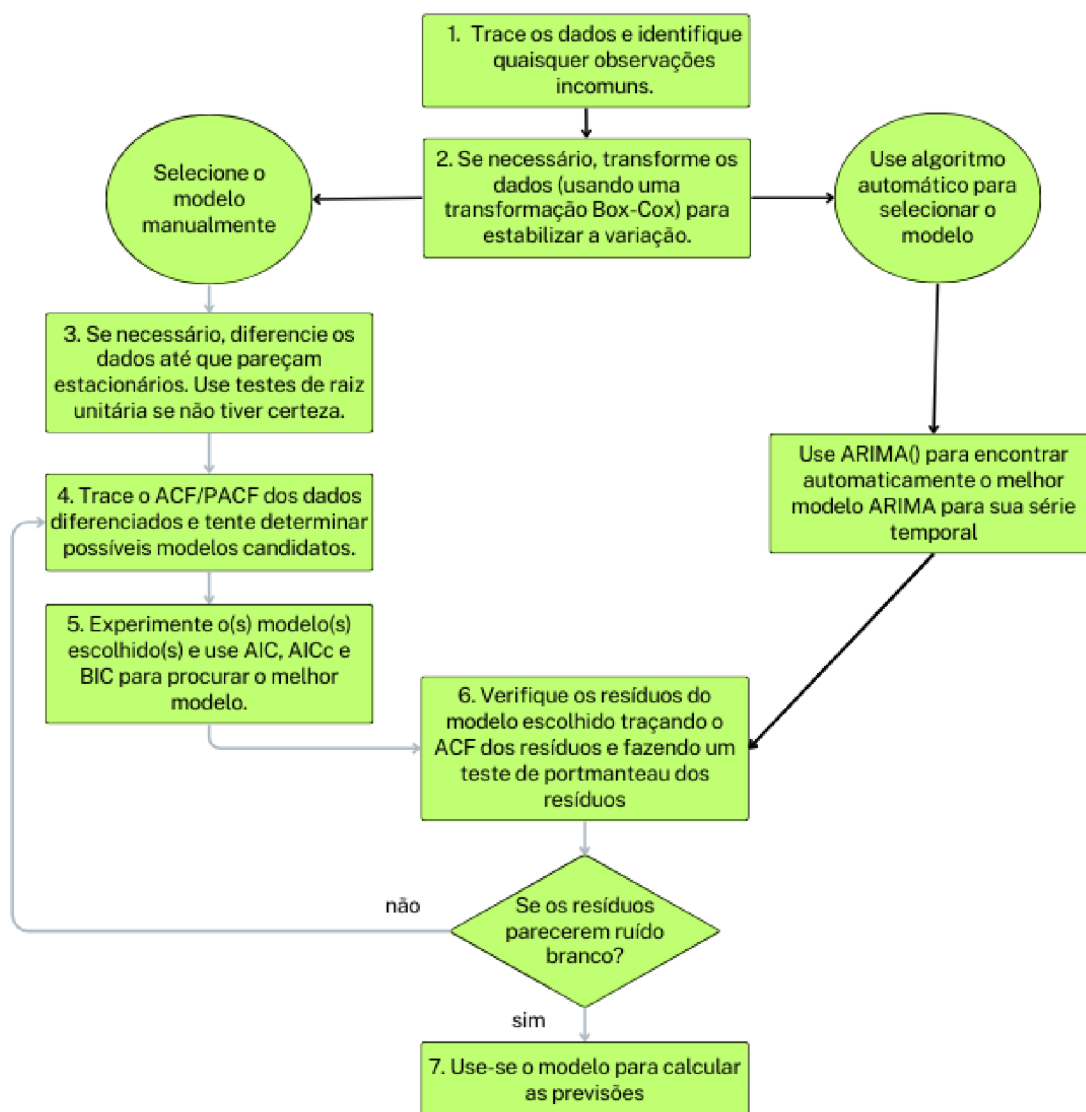
Após ajustar o modelo ARIMA escolhido, os resíduos do modelo são verificados por meio da plotagem da ACF dos resíduos. Se os resíduos não se assemelharem a ruído branco, podem ser necessárias modificações no modelo, como a alteração da ordem dos termos ARIMA ou a inclusão de outros preditores. Complementada pela aplicação dos testes de raízes unitárias ADF (Lin *et al.*, 2024), KPSS (Hassan *et al.*, 2023) e PP (Escobari; Garcia; Mellado, 2017). Posteriormente, ajustes foram realizados nas séries para assegurar a estacionariedade desejada.

Por fim, a validação do modelo ARIMA foi conduzida por meio de uma análise detalhada dos resíduos, garantindo a confiabilidade e precisão da modelagem. As métricas RMSE, MAE e MAPE foram utilizadas para avaliar o desempenho dos modelos em termos de precisão das previsões (Zhao; Li; Qu, 2022).

Em resumo, o procedimento de modelagem ARIMA delineado no fluxograma apresentado na Figura 3.2 oferece uma abordagem sistemática para ajustar modelos ARIMA a dados de séries temporais. Ao seguir essas etapas, é possível identificar e modelar eficazmente a estrutura subjacente dos dados, resultando em previsões precisas. No geral, o procedimento de modelagem ARIMA fornece uma estrutura abrangente para a

previsão de séries temporais, o que possibilita a produção de previsões confiáveis.

Figura 3.2 – Processo geral para previsão usando um modelo ARIMA



Fonte: Adaptado do (Hyndman; Athanasopoulos, 2023)

Os procedimentos foram realizados no software R (R Core Team, 2023), predominantemente utilizando funções dos pacotes *forecast* (Hyndman; Athanasopoulos, 2023) e *urca* (Pfaff, 2011). Para as análises de inferência, adotaram-se níveis de significância de 0,01 e 0,05, enquanto as estimativas por intervalo foram calculadas com um nível de confiança de 0,95.

4 Resultados

Este capítulo apresenta uma análise abrangente dos métodos de previsão aplicados às séries temporais. Inicialmente, realiza-se a análise exploratória dos dados, identificando padrões e tendências relevantes. Em seguida, os resultados das duas abordagens utilizadas, o modelo de suavização exponencial e a metodologia de Box-Jenkins (também conhecida como modelo ARIMA), são apresentados e avaliados quanto à sua capacidade de capturar a estrutura dos dados e fazer previsões precisas. Por fim, uma discussão comparativa entre os métodos é realizada, destacando suas vantagens, limitações e aplicabilidade para a série temporal em questão.

4.1 Análise exploratória dos dados

Nesta seção apresenta-se a análise exploratória dos dados, onde a série de emissão total de CO_2e no estado do Amazonas, no período analisado, é desmembrada em 5 (cinco) categorias, conforme o setor responsável por sua emissão e uma série com o somatório das 5 categorias, o mesmo acontecendo com a série relativa ao Brasil. As categorias consideradas são: processos industriais, tratamento de resíduos, produção de energia, agropecuária e mudança do uso da terra e floresta (MUTF).

A Tabela 4.1 apresenta o total das emissões de CO_2e acumuladas entre 1990 e 2021, tanto no estado do Amazonas quanto no território nacional brasileiro. É notável que a categoria MUTF é a principal responsável pela emissão de CO_2e , tanto no Amazonas quanto no Brasil. Chama a atenção o fato de que, no Amazonas, a soma das emissões das demais categorias se restringe a menos de 20% do total de emissões da categoria MUTF. Por sua vez, no âmbito do nacional, a MUTF representa cerca de 56% das emissões, enquanto o somatório de todas as outras categorias corresponde a 44%.

Tabela 4.1 – Distribuição do total das emissões de CO_2e por categoria, 1990-2021

Categorias	Emissões de CO_2e (t)		Percentual (%)	
	Amazonas	Brasil	Amazonas	Brasil
Processos industriais	6.722.129	2.621.947.214	0,32	3,67
Tratamento de resíduos	41.105.647	1.952.989.378	1,95	2,74
Produção de energia	197.950.441	10.497.943.205	9,38	14,71
Agropecuária	85.495.161	15.796.695.543	4,05	22,13
MUTF	1.780.134.357	40.515.229.878	84,31	56,76
Total	2.111.407.735	71.384.805.218	100,0	100,0

CO_2e - dióxido de carbono equivalente, t - toneladas, MUTF - mudança do uso da terra e floresta

Fonte: Elaboração própria com dados da plataforma SEEG

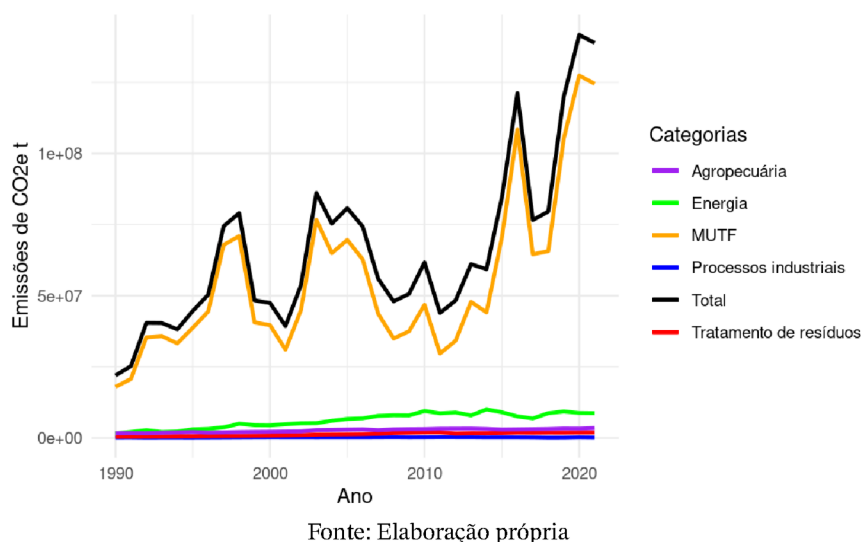
Os gases produzidos pelas atividades industriais respondem por 0,32% das emissões totais dos últimos 31 anos no estado do Amazonas, sendo a categoria com a menor quantidade de emissão de CO_2e . Já no Brasil, a série da categoria tratamento de resíduos tem a menor quantidade de emissão de CO_2e . Também é possível observar que o segundo emissor de GEE no Amazonas é o setor de energia, enquanto no Brasil, o segundo maior é a agricultura, representando quase um quarto das emissões no Brasil.

As emissões de GEE aumentaram significativamente nos últimos 31 anos. No entanto, enormes disparidades são encontradas entre as fontes produtoras de CO_2e individualmente, tanto em termos de valores absolutos de emissões de GEE quanto em termos de seu ritmo de crescimento no período 1990-2021. Em termos absolutos a MUTF tem a maior emissão. Sozinha, ela representa 84,31% das emissões totais neste período. Todas as outras categorias representam 15,69% das emissões.

A Figura 4.1 exibe o histórico das emissões de CO_2e ao longo do período compreendido entre os anos de

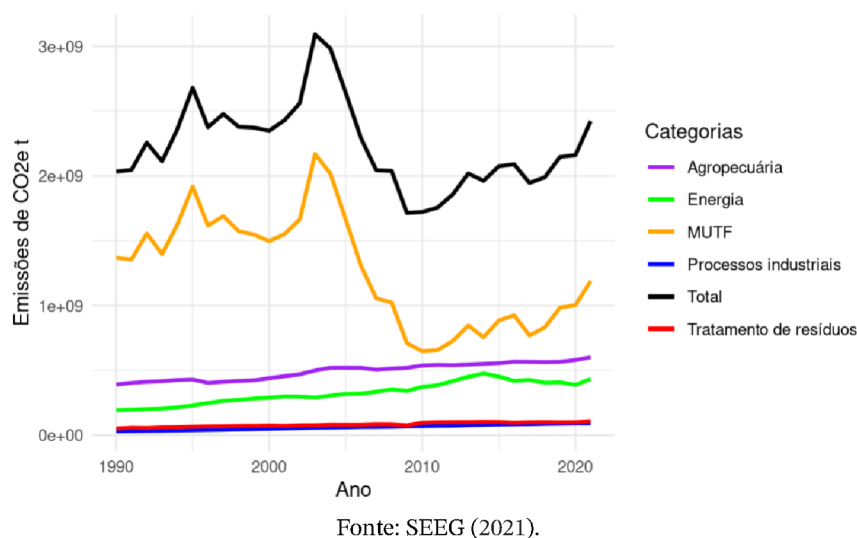
1990 e 2021, relaciona as cinco categorias distintas, bem como à série total, no contexto do estado do Amazonas. Nota-se que as emissões originadas pela MUTF se destacam de maneira notável em relação às outras formas de emissões de GEE, comportamento que se repete em relação a essa mesma categoria de emissões em âmbito nacional, representando as emissões do Brasil.

Figura 4.1 – Emissões de CO_2e distribuído por categoria de origem no estado do Amazonas, 1990-2021



No estado do Amazonas, as emissões de CO_2e associadas à categoria MUTF apresentam uma tendência ascendente desde o ano de 2011. Em contrapartida, no Brasil, as emissões da categoria MUTF revelam uma redução do ano de 2003 até 2011, a partir do qual apresentam uma tendência crescente, conforme ilustrado na Figura 4.2.

Figura 4.2 – Emissões de CO_2e distribuído por categoria de origem no Brasil, 1990-2021



As cinco séries de emissão de CO_2e relativas ao Brasil estão representadas no gráfico da Figura 4.2, observa-se que as séries da indústria, energia e agropecuária possuem crescimento ao longo dos anos, com algumas variações. A categoria tratamento de resíduos tem um crescimento constante sem variação. A categoria

MUTF possui grandes variações, além de possuir um comportamento semelhante ao comportamento do total de emissão de CO_2e .

A Figura 4.3 apresenta as mesmas cinco séries e a série total do estado do Amazonas e do Brasil, mas em uma escala diferente para permitir uma análise mais minuciosa em cada uma das categorias. Começando pela indústria, vê-se uma tendência de crescimento da série até 2011, depois há uma queda acentuada até o ano de 2018, ano em que foi aprovado o decreto n.º 9.394, de 30 de maio de 2018. A medida altera a tabela de incidência de IPI (Imposto sobre Produtos Industrializados) dos concentrados de bebidas, o maior segmento de exportação do Amazonas, reduzindo de 20% a 4% a alíquota de crédito do imposto para quem compra concentrado das empresas do Polo Industrial de Manaus. A partir deste ano, as emissões de CO_2e mostram tendência de crescimento. No Brasil, a série indústria possui crescimento ao longo dos anos, com algumas variações.

Observa-se uma tendência de crescimento ao longo do tempo para a série de tratamento de resíduos, com uma curva ascendente. Nota-se uma queda no ano de 2011, voltando a crescer em 2012, com tendência de alta, enquanto no Brasil a série tem tendência crescente e constante, sem picos ou vales.

A categoria energia tem tendência de alta constante do início da série até 2010, pois cerca de 65% da energia elétrica de Manaus é gerada a partir da queima de combustíveis fósseis (Medeiros; Batista; Calderaro, 2017). Analisando o gráfico na Figura 4.3, confirma-se alta emissão de CO_2e a partir de 2010. Nos anos em 2014 e 2019 houve as maiores crises energéticas da década no Brasil.

Pode-se notar que a série que representa as emissões da agropecuária apresenta uma tendência crescente, no geral, com algumas oscilações ao longo do tempo. Há alguns picos, tanto no Amazonas como no Brasil, que podem ser indicativos de eventos ou mudanças significativas que influenciaram os dados.

É possível observar que as flutuações de picos e vales não são tão regulares na série MUTF no estado do Amazonas. Esta série, devido a sua semelhança com as emissões totais, apresenta uma grande semelhança nas flutuações que as emissões totais brutas, com picos nos anos de 1998, 2003, 2016 e 2020. As maiores quedas nas emissões totais de CO_2e acontecem entre 1998 a 2001, enquanto na MUTF acontece entre 2003 a 2011.

O gráfico da MUTF praticamente se reflete ao gráfico das emissões totais de CO_2e da série histórica tanto no Amazonas quanto no Brasil. A série MUTF tem um pico em 2003. A partir deste ano, o comportamento é de queda até o ano de 2010. Depois, há uma tendência de crescimento até o ano de 2021.

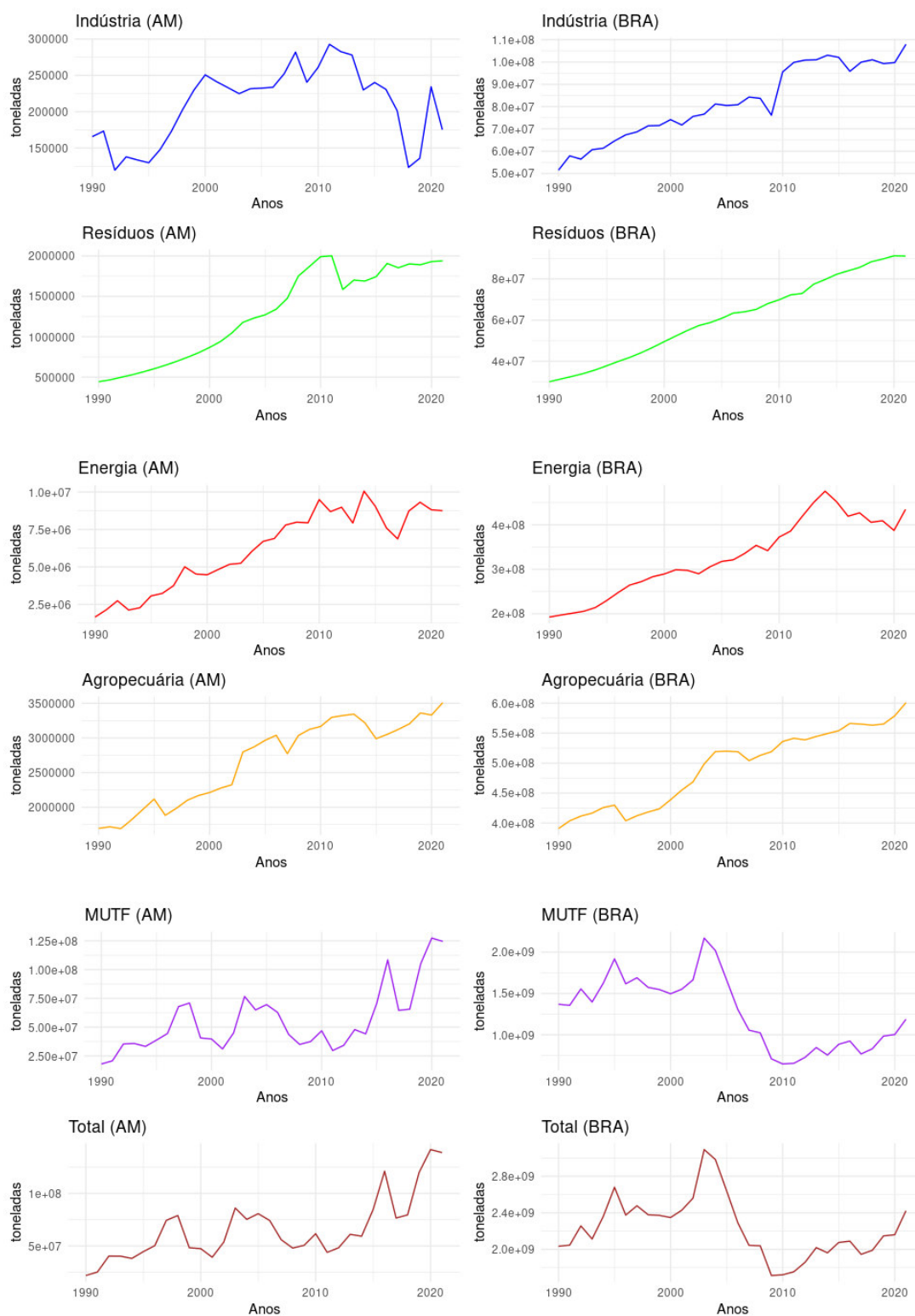
Todas as séries apresentam uma ampla variação ao longo do período de 1990 a 2021. Isso sugere que há fatores variáveis que influenciam as emissões em cada setor, levando a aumentos e diminuições em diferentes anos. Ao comparar os diferentes setores, é possível identificar setor de MUTF e o setor de Energia têm as maiores emissões em termos absolutos no Amazonas.

Esta análise evidencia um comportamento diferente entre a emissão deste gás no Amazonas e no Brasil. No entanto, é importante notar que a interpretação do gráfico deve ser feita em conjunto com outras análises estatísticas, como a decomposição da série temporal em componentes de tendência e efeito aleatório, bem como com a realização de testes de hipóteses e verificação da estacionariedade dos dados.

Com base nas estatísticas descritivas dos dados para o estado do Amazonas, conforme apresentado nas medidas descritivas dos dados, apresentadas na Tabela 4.2, na categoria de processo industriais observa-se que há uma dispersão moderada nas emissões de CO_2e , indicada por um desvio padrão relativamente baixo em relação à média das emissões. Isso sugere uma variação consistente em torno da média, evidenciando uma concentração das emissões nesse valor médio.

Por outro lado, no setor de tratamento de resíduos, a dispersão é maior em comparação ao setor industrial, já que o desvio padrão é substancialmente superior à média das emissões. Esse resultado sugere uma maior variabilidade e uma menor concentração em torno da média no setor de tratamento de resíduos.

Figura 4.3 – Séries temporais das emissões de CO_2e no Amazonas e no Brasil



Fonte: Elaboração própria

A dispersão das emissões no setor de energia e agropecuária são consideráveis, uma vez que o desvio padrão é bastante alto em relação à média das emissões. Isso sugere que as emissões podem variar significativamente em relação à média, indicando uma maior dispersão e uma menor concentração em torno da média.

Tabela 4.2 – Medidas descritivas das emissões de CO_2e do estado do Amazonas, 1990-2021

Medidas	Indústria(t)	Trat. resíduos(t)	Energia(t)	Agr.(t)	MUTF(t)	Total(t)
Mínimo	119602	443991	1658841	1691897	17997441	21961271
1º Quartil	171104	733014	4295020	2113878	35670035	46759049
Mediana	230390	1306333	6788602	2918307	44671311	57659086
Média	210067	1284552	6185951	2671724	55629199	65981492
3º Quartil	240733	1857206	8701040	3174926	68246084	79158021
Máximo	292761	2001427	10053720	3508736	127373987	141685548
Desvio Padrão	51861,96	560435,40	2595071,00	602957,10	28135920,00	30102248,00
Assimetria	-0,341034	-0,156828	-0,268586	-0,331116	1,105897	1,019846
Curtose	-1,186419	-1,641646	-1,386721	-1,534121	0,429676	0,374203

MUTF - mudança do uso da terra e floresta

Fonte: Elaboração própria

As séries MUTF e total apresentam uma dispersão muito alta, com desvio padrão extremamente alto em relação à média das emissões. Ambas têm uma grande variação em relação à média, indicando uma dispersão muito ampla e uma concentração relativamente baixa em torno da média.

Em resumo, os setores de produção de energia, agropecuária, MUTF e as emissões totais possuem dispersão muito alta. Isso sugere uma menor concentração das emissões em torno da média. Por outro lado, os setores de processos industriais e tratamento de resíduos têm dispersões menores, indicando uma maior concentração das emissões em torno das médias.

Além disso, observa-se que o setor de energia também apresenta discrepâncias entre a média e a mediana das emissões. As séries indústria (-0,341034), tratamento de resíduos (-0,156828), energia (-0,268586) e agropecuária (-0,331116) apresentam assimetria negativa. Isso significa que todas essas séries têm uma distribuição com cauda mais longa à esquerda, indicando que a maioria dos valores de emissões está concentrada em níveis relativamente altos, com maior concentração dos valores mais altos.

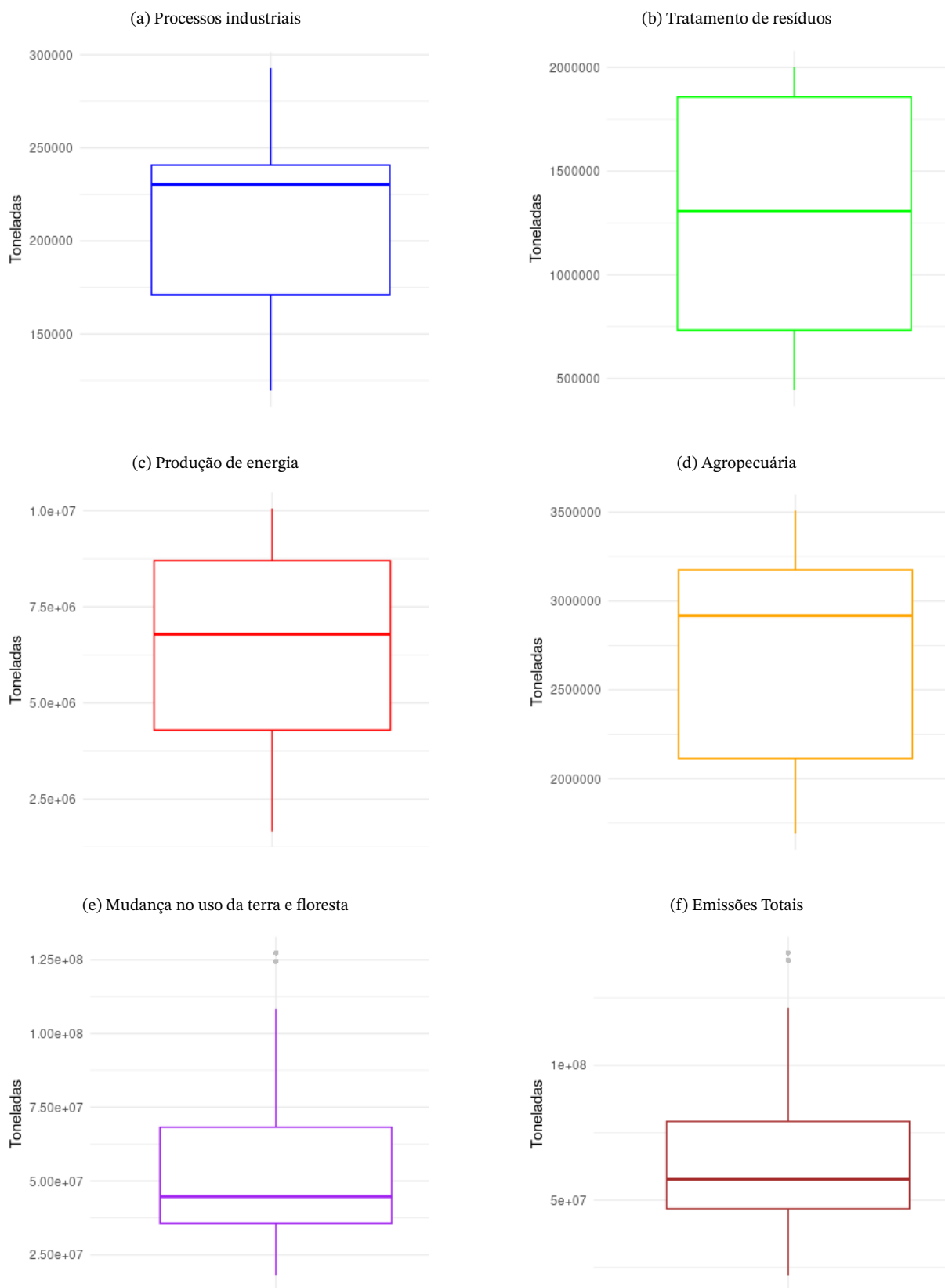
Já as séries de MUTF (1,105897) e total (1,019846) apresentam assimetria positiva. Isso indica que ambas as séries possuem uma distribuição com cauda mais longa à direita, com a maioria dos valores concentrados em níveis relativamente baixos, mas existem alguns valores extremamente altos que influenciam a assimetria, com maior concentração dos valores mais altos.

Quanto a curtose, as séries indústria (-1,186419), tratamento de resíduos (-1,641646), energia (-1,386721) e agropecuária (-1,534121) tem uma distribuição relativamente achatada em relação à normal (curtose negativa), o que sugere que os dados têm menos valores extremos e são menos dispersos em relação à média, com caudas mais leves. Isso indica que os valores de emissões nessas séries estão mais concentrados em torno da média, com menos valores discrepantes.

Por outro lado, as séries de MUTF (0,429676) e total (0,374203) têm uma distribuição mais concentrada nos extremos (curtose positiva), o que sugere que os dados têm mais valores extremos e são mais dispersos em relação à média apresentando caudas mais pesadas. Isso indica que os valores de emissões nessas séries estão mais espalhados e com um número maior de valores discrepantes.

Complementando a análise realizada foram traçados *boxplot* que estão na Figura 4.11. Observa-se que o gráfico relativo às emissões dos processos industriais têm uma mediana próxima ao terceiro quartil indicando distribuição assimétrica, enquanto a série temporal de tratamento de resíduos têm uma mediana próxima à metade do intervalo entre o primeiro e o terceiro quartil, indicando uma distribuição simétrica.

Figura 4.4 – Emissões de dióxido de carbono equivalente no estado do Amazonas



Fonte: Elaboração própria

Os dados da série agropecuária são assimétricos nos 50% centrais, o mesmo não acontecendo nas emissões de tratamento de resíduos e produção de energia, nestes os que parecem ter uma assimetria mais leve. Estas conclusões concordam com as medidas descritivas, aprestadas na Tabela 4.2.

Observa-se que a série indústria apresenta uma distribuição simétrica e concentrada em torno do 2º quartil (mediana), que está próxima de 230.000 t de CO_2e . O intervalo interquartil (IQR), que abrange a metade das observações, varia entre aproximadamente 171.000 e 241.000 toneladas métricas de CO_2e .

As medidas descritivas das séries MUTF e total mostram que a média e a mediana das emissões são próximas, indicando uma tendência geral de estabilidade nesse setor. Além disso, os valores dos quartis e a proximidade entre o mínimo e o máximo sugerem uma distribuição homogênea das emissões no MUTF, sem ocorrências de anos com valores extremos.

As emissões provenientes do tratamento de resíduos apresenta uma distribuição relativamente simétrica e concentrada. A mediana das emissões está próxima de 1.310.000 t de CO_2e , e o IQR abrange valores entre aproximadamente 733.000 t e 1.857.000 t de CO_2e .

As emissões provenientes do tratamento de resíduos apresentados na Figura 4.4b, apresenta uma distribuição relativamente simétrica e concentrada. A mediana das emissões está próxima de 1.310.000 t de CO_2e , e o IQR abrange valores entre aproximadamente 733.000 t e 1.857.000 t de CO_2e .

Já as emissões de CO_2e relacionadas à produção de energia exibem uma distribuição assimétrica à direita. A mediana das emissões está próxima de 6.790.000 t de CO_2e , e o IQR varia entre cerca de 4.300.000 t e 8.700.000 t de CO_2e . Observações acima do limite superior do *boxplot* podem ser consideradas possíveis outliers.

As emissões de CO_2e da agropecuária no estado do Amazonas indicam uma distribuição relativamente simétrica e concentrada. A mediana das emissões está em torno de 2.920.000 t de CO_2e , e o IQR varia entre cerca de 2.114 Mt e 3.175 Mt de CO_2e . Alguns valores abaixo do limite inferior do *boxplot* podem ser considerados possíveis outliers.

As emissões de CO_2e relacionadas à mudança do uso da terra no estado do Amazonas exibem uma distribuição assimétrica à direita. A mediana das emissões está próxima de 44.670 Mt toneladas métricas de CO_2e , e o IQR varia entre cerca de 31.635 Mt e 68.250 Mt toneladas métricas de CO_2e . Observações acima do limite superior do *boxplot* podem ser consideradas possíveis outliers.

O total de emissões de CO_2e no estado do Amazonas mostra uma distribuição assimétrica à direita. A mediana das emissões está próxima de 57.660 Mt de CO_2e , e o IQR varia entre cerca de 46.760 Mt e 79.160 Mt de CO_2e . Observações acima do limite superior do *boxplot* podem ser consideradas possíveis outliers.

A amplitude da série MUTF (diferença entre o maior e o menor valor) é de aproximadamente 109.376 Mt de CO_2e . A Figura 4.4e mostra que a distribuição dos valores é assimétrica, com uma cauda longa no lado direito, indicando haver alguns valores muito maiores que a maioria dos valores observados. Há alguns valores considerados discrepantes (*outliers*, relativo aos anos de 2020 e 2021. A mediana está mais próxima do primeiro quartil, sugerindo uma distribuição com assimetria positiva.

Também é importante considerar como os resultados da análise exploratória podem ser usados para responder às perguntas de pesquisa ou hipóteses iniciais. Por exemplo, se a hipótese inicial era de que a série temporal mostraria uma tendência de crescimento, os resultados da análise exploratória devem ser usados para confirmar ou refutar essa hipótese.

Por fim, é importante lembrar que a análise exploratória é apenas o primeiro passo em um processo mais amplo de análise de dados. Os resultados da análise exploratória devem ser usados como base para a realização de análises mais aprofundadas e para tomar decisões mais precisas.

A análise da tendência nas séries é apresentada na Tabela 4.3, os testes de Cox-Stuart demonstram

que as séries de tratamento de resíduos, produção de energia e agropecuária exibem uma tendência crescente com um p-valor igual a $1,5259\text{e-}05$, indicando uma tendência positiva. Por outro lado, a série de processos industriais, com um p-valor igual a $1,0506\text{e-}01$, sugerem não haver evidência estatística para uma tendência. O p-valor relativamente alto nesses casos indica que a tendência observada nas séries pode estar sujeita a variações aleatórias. Quanto à série total, seu p-valor é $3,8406\text{e-}02$, indicando uma tendência crescente.

Tabela 4.3 – Resultados dos testes de tendência e quebra estrutural

Testes	Indústria	Trat. resíduos	Energia	Agropecuária	MUTF	Total
Cox-Stuar	1,0506e-01	1,5259e-05	1,5259e-05	1,5259e-05	1,0506e-01	3,8406e-02
Chow	2,9097e-108	2,6295e-140	6,103e-161	2,7245e-141	4,9782e-193	6,132e-194

MUTF - mudança do uso da terra e floresta

Fonte: Elaboração própria

Com base nos resultados obtidos através do teste de Chow e análise dos p-valor para cada categoria apresentados Tabela 4.3, é possível tirar conclusões sobre a presença de quebra de estrutura em todas as séries. Na de processos industriais, o p-valor é extremamente baixo ($2,9097\text{e-}108$) indica uma forte evidência estatística de mudança estrutural. Esta tendência é corroborada pela série de tratamento de resíduos, onde também observamos um p-valor muito baixo ($2,6295\text{e-}140$), apontando para a ocorrência significativa de uma quebra de estrutura. Similarmente, as séries energia, agropecuária, MUTF e total apresentam os valores p ($6,103\text{e-}161$, $2,7245\text{e-}141$ e $4,9782\text{e-}193$, $6,132\text{e-}194$, respectivamente), reforçando a hipótese de quebra estrutural. No Apêndice A apresenta-se um gráfico com cada uma das categorias e suas respectivas quebras de estrutura.

4.2 Modelo de suavização exponencial

A Tabela 4.4 apresenta todos os coeficientes dos modelos obtidos pelo método de suavização exponencial. Os modelos de suavização exponencial aplicados à categoria indústria, exceto o MM e MM_d , os parâmetros de suavização para o nível α são relativamente altos, indicando que os modelos dão mais peso aos dados mais recentes na estimativa do nível da série temporal. Os parâmetros de suavização para a tendência β são muito pequenos ($1,00 \times 10^{-4}$) em todos os métodos, indicando que as tendências são suavizadas com extrema cautela e que os modelos tendem a ser mais sensíveis às mudanças nas tendências.

Nos métodos amortecidos (AA_d , MA_d , MM_d), os parâmetros de amortecimento da tendência φ variam, indicando diferentes graus de amortecimento da tendência ao longo do tempo. As estimativas iniciais de nível $\ell(0)$ e de tendência $b(0)$ variam entre os métodos, refletindo diferentes interpretações e ajustes iniciais da série temporal. Não há indicação imediata de que algum dos modelos não seja adequado para realizar previsões.

Os coeficientes na categoria de tratamento de resíduos revelam uma uniformidade notável no parâmetro de suavização α , o qual controla a suavização do nível da série. Para todos os modelos, esse parâmetro é aproximadamente igual a 0,9999, indicando dependência dos valores recentes para as previsões futuras. Este comportamento é típico em métodos de suavização exponencial, onde os dados mais recentes tendem a ter um peso significativamente maior.

A análise dos parâmetros β e φ , responsáveis pela suavização da tendência e pelo amortecimento da tendência, respectivamente, ressalta a variedade de comportamentos capturados pelos diferentes modelos. Por exemplo, o modelo MM_d exibe um valor de φ de 0,9633, sugerindo um amortecimento considerável da tendência ao longo do tempo.

Os valores iniciais estimados para o nível $\ell(0)$ e a tendência $b(0)$ também são reveladores. Esses valores fornecem uma base inicial para as previsões da série temporal. Por exemplo, o modelo AA tem uma estimativa inicial de nível de 381.048,16 e uma estimativa inicial de tendência de 61.187,90.

Para a categoria produção de energia, a análise dos coeficientes revela que os modelos variam significa-

tivamente em termos de parâmetros de suavização (α , β , φ) e estados iniciais $\ell(0)$, $b(0)$. O modelo AA possui um valor muito baixo de α (0,0001), indicando uma suavização mínima do nível, enquanto o modelo MM possui um valor elevado de α (0,6057), sugerindo uma suavização mais agressiva.

Tabela 4.4 – Coeficientes dos modelos estimados pelo Método de Suavização Exponencial, 1990-2015

Categorias	Modelos	α	β	φ	$\ell(0)$	$b(0)$
Processos industriais	AA	0,9759	1,00E-04		132.610,86	6.515,09
Processos industriais	AA_d	0,9409	1,00E-04	0,9473	125.511,47	8.284,31
Processos industriais	MA	0,9666	1,00E-04		135.316,84	6.815,90
Processos industriais	MA_d	0,9615	1,00E-04	0,9717	132.688,05	8.139,24
Processos industriais	MM	0,9545	1,00E-04		138.398,85	1,03
Processos industriais	MM_d	0,9613	1,00E-04	0,9130	136.854,04	1,09
Tratamento de resíduos	AA	0,9999	1,00E-04		381.048,16	61.187,90
Tratamento de resíduos	AA_d	0,9999	1,00E-04	0,9800	381.048,84	71.149,87
Tratamento de resíduos	MA	0,9999	8,03E-02		381.049,74	46.783,05
Tratamento de resíduos	MA_d	0,9999	1,37E-01	0,9800	381.048,49	51.988,83
Tratamento de resíduos	MM	0,9999	1,10E-01		402.747,05	1,07
Tratamento de resíduos	MM_d	0,9997	1,00E-04	0,9633	384.479,34	1,10
Energia	AA	0,0001	1,00E-04		1.219.450,58	333.738,89
Energia	AA_d	0,2848	1,00E-04	0,9800	1.219.450,39	398.487,96
Energia	MA	0,0925	1,00E-04		1.341.028,80	333.738,87
Energia	MA_d	0,3405	1,00E-04	0,9800	1.284.233,72	366.485,38
Energia	MM	0,6057	1,33E-01		1.406.325,64	1,15
Energia	MM_d	0,0001	1,00E-04	0,9325	1.406.325,94	1,18
Agropecuária	AA	0,9424	1,00E-04		1.624.241,20	72.759,59
Agropecuária	AA_d	0,9167	1,00E-04	0,9643	1.624.241,46	86.685,55
Agropecuária	MA	0,8429	1,00E-04		1.624.241,65	67.792,29
Agropecuária	MA_d	0,8464	1,00E-04	0,9800	1.624.241,69	78.190,92
Agropecuária	MM	0,8745	1,00E-04		1.625.866,89	1,03
Agropecuária	MM_d	0,8307	1,00E-04	0,9593	1.625.866,35	1,04
MUTF	AA	0,9999	1,01E-04		17.172.557,63	4.536.902,44
MUTF	AA_d	0,9764	1,00E-04	0,9171	17.172.557,89	4.536.903,49
MUTF	MA	0,9999	1,00E-04		15.611.416,87	4.536.903,58
MUTF	MA_d	0,9973	1,00E-04	0,9800	15.611.415,75	4.536.904,01
MUTF	MM	0,9999	1,00E-04		18.089.780,57	1,10
MUTF	MM_d	0,9999	1,00E-04	0,9800	18.089.780,43	1,13
Total	AA	0,9999	1,02E-04		20.857.836,27	4.969.931,05
Total	AA_d	0,9845	1,00E-04	0,9333	20.857.835,98	4.969.930,93
Total	MA	0,9999	1,00E-04		18.961.669,58	4.969.931,16
Total	MA_d	0,9999	1,18E-04	0,9800	18.961.669,79	4.969.930,24
Total	MM	0,9999	1,00E-04		21.797.933,76	1,08
Total	MM_d	0,9999	1,00E-04	0,9440	21.797.932,82	1,17

α - parâmetro de suavização para o nível, β parâmetro de suavização para a tendência, φ parâmetros de amortecimento da tendência, ℓ - estimativa inicial de nível e b - estimativa inicial de tendência, AA - tendência aditiva e erro aditivo, MA - tendência aditiva e erro multiplicativo, MM - tendência multiplicativa e erro multiplicativo, AA_d - tendência aditiva amortecida e erro aditivo, MA_d - tendência aditiva amortecida e erro multiplicativo, MM_d - tendência multiplicativa amortecida e erro multiplicativo.

Fonte: Elaboração própria

Além disso, os modelos diferem em termos de capacidade de capturar a tendência da série temporal, com valores distintos de β e φ . Por exemplo, o modelo MM_d apresenta um valor alto de β (0,133) indicando uma tendência significativa, enquanto o modelo AA_d tem um valor de φ (0,9800) que indica um amortecimento

considerável da tendência ao longo do tempo.

Os coeficientes da categoria agropecuária mostram que o parâmetro de suavização α , responsável pela suavização do nível da série, varia entre os modelos, com valores como 0,9424, 0,9167, 0,8429, 0,8464, 0,8745 e 0,8307. Isso sugere que diferentes modelos atribuem pesos diferentes aos valores recentes da série para fazer previsões futuras.

Além disso, os parâmetros β e φ , que controlam a suavização da tendência e o amortecimento da tendência, também exibem variação entre os modelos. Por exemplo, os valores de β são todos iguais a 0,0001, enquanto os valores de φ variam entre 0,9593 e 0,9800, indicando diferentes graus de amortecimento da tendência ao longo do tempo.

Na categoria MUTF ao observar os valores dos parâmetros, nota-se que os parâmetros de suavização para o nível α e para a tendência β , variando entre 0,9764 e 0,9999. Isso sugere uma ênfase considerável nos dados mais recentes para previsões futuras. Quanto ao parâmetro de amortecimento da tendência φ , ele varia entre 0,9171 e 0,9800, indicando diferentes graus de amortecimento da tendência ao longo do tempo.

As estimativas iniciais de nível e tendência $\ell(0)$, $b(0)$ exibem considerável variação entre os modelos, o que pode influenciar as previsões futuras da série temporal. Não há um modelo claramente superior em todos os aspectos.

Já a categoria total observa-se variações nos parâmetros de suavização da tendência β e nos parâmetros de amortecimento da tendência φ entre os diferentes modelos. O modelo AA_d possui um valor de φ de 0,9333, enquanto o modelo MM_d tem um valor de φ de 0,9440, indicando diferentes níveis de amortecimento da tendência ao longo do tempo.

Os valores iniciais estimados para o nível $\ell(0)$ e a tendência $b(0)$ também variam entre os modelos. O modelo AA tem uma estimativa inicial de nível de 20.857.836,27 e uma estimativa inicial de tendência de 4.969.931,05. Todos os modelos são adequados para realizar previsões, mas a escolha do modelo mais apropriado dependerá da análise mais aprofundada do comportamento dos dados e da avaliação do desempenho de cada modelo em dados de teste.

A Tabela 4.5 apresenta os resultados dos critérios de informação AIC, AICc e BIC e os teste de diagnóstico para cada um dos métodos aplicados a todas as séries de emissão de CO_2e no estado do Amazonas. Estes dados permitem selecionar o modelo mais parcimonioso, além de verificar a autocorrelação, a heterocedasticidade e a normalização da série. A partir destes dados selecionou-se os dois modelos que apresentaram o melhor ajuste a partir do critério de informação AIC, devido à pequena quantidade de observações da série.

Na categoria indústria, observa-se que os modelos AA e MA têm os menores valores de AIC, AICc e BIC, o que sugere que esses modelos são os mais parcimoniosos e apresentam o melhor equilíbrio entre ajuste e complexidade. Os valores dos testes indicam que os resíduos dos modelos têm autocorrelação e heterocedasticidade, conforme refletido nos valores elevados de p-valor para os testes de Ljung-Box e ARCH, respectivamente. Além disso, o teste de Jarque-Bera indica que os resíduos não apresentam desvios da normalidade.

Para a série tratamento de resíduos os modelos MA e MM apresentam os valores mais baixos para os três critérios (AIC, AICc e BIC), tornando-os os dois modelos mais parcimoniosos. O modelo MA tem um AIC de 669,65, AICc de 669,65 e BIC de 675,94, enquanto o modelo MM tem um AIC de 669,71, AICc de 669,71 e BIC de 676,01. Os resultados dos testes Ljung-Box, ARCH e Jarque-Bera sugerem que os modelos MA e MM exibem características desejáveis, tais como autocorrelação insignificante, homocedasticidade e distribuição normal nos resíduos.

Os dois modelos mais parcimoniosos para a categoria de produção de energia são: o modelo MM_d com valores de 780,58 para AIC, 780,58 para AICc e 788,13 para BIC, enquanto o modelo AA com 780,35 para AIC, 780,35 para AICc e 786,64 para BIC.

Na categoria agropecuária, observa-se que o modelo *MA* apresenta o menor valor de AICc, com 711,26, e o menor valor de BIC, com 717,55. Além disso, o modelo *MM* também se destaca, com um AICc de 712,53 e um BIC de 718,82. Os modelos apresentam bons resultados em relação à autocorrelação, heterocedasticidade e normalidade dos resíduos.

Tabela 4.5 – Resultados dos critérios de informação e dos testes de autocorrelação, heterocedasticidade e normalidade dos resíduos dos modelos, 1990-2015

Categorias	Modelos	AIC	AICc	BIC	Ljung-Box	ARCH	JB
Processos industriais	<i>AA</i>	617,97	617,97	624,26	0,9432	0,1257	0,7969
Processos industriais	<i>AA_d</i>	620,06	620,06	627,61	0,9579	0,1486	0,8080
Processos industriais	<i>MA</i>	617,66	617,66	623,95	0,9374	0,0601	0,8056
Processos industriais	<i>MA_d</i>	619,87	619,87	627,42	0,9056	0,0470	0,7923
Processos industriais	<i>MM</i>	618,40	618,40	624,69	0,8704	0,1112	0,6754
Processos industriais	<i>MM_d</i>	619,58	619,58	627,13	0,8676	0,0083	0,8193
Tratamento de resíduos	<i>AA</i>	698,00	698,00	704,29	0,6224	<1,0x10-15	0,0045
Tratamento de resíduos	<i>AA_d</i>	699,81	699,81	707,36	0,6276	<1,0x10-15	0,0057
Tratamento de resíduos	<i>MA</i>	669,65	669,65	675,94	0,4734	4,33E-15	0,0121
Tratamento de resíduos	<i>MA_d</i>	672,25	672,25	679,80	0,5932	7,88E-15	0,0120
Tratamento de resíduos	<i>MM</i>	669,71	669,71	676,01	0,4956	<1,0x10-15	0,0050
Tratamento de resíduos	<i>MM_d</i>	670,41	670,41	677,95	0,5003	<1,0x10-15	0,0084
Energia	<i>AA</i>	780,35	780,35	786,64	0,8572	0,9192	0,2046
Energia	<i>AA_d</i>	787,03	787,03	794,58	0,8861	0,9780	0,2276
Energia	<i>MA</i>	781,30	781,30	787,59	0,3195	0,9843	0,2406
Energia	<i>MA_d</i>	784,94	784,94	792,48	0,5808	0,7227	0,3564
Energia	<i>MM</i>	791,89	791,89	798,18	0,9081	0,8215	0,0875
Energia	<i>MM_d</i>	780,58	780,58	788,13	0,4839	0,3480	0,1571
Agropecuária	<i>AA</i>	714,00	714,00	720,29	0,9771	0,2093	0,3488
Agropecuária	<i>AA_d</i>	714,96	714,96	722,51	0,9699	0,1801	0,4888
Agropecuária	<i>MA</i>	711,26	711,26	717,55	0,8734	0,0791	0,3980
Agropecuária	<i>MA_d</i>	713,20	713,20	720,75	0,8641	0,0562	0,4891
Agropecuária	<i>MM</i>	712,53	712,53	718,82	0,8881	0,1547	0,4341
Agropecuária	<i>MM_d</i>	713,51	713,51	721,06	0,8446	0,0598	0,4849
MUTF	<i>AA</i>	950,10	950,10	956,39	0,9940	0,9844	0,2999
MUTF	<i>AA_d</i>	951,20	951,20	958,75	0,9800	0,9463	0,3458
MUTF	<i>MA</i>	941,86	941,86	948,15	0,9866	0,5310	0,3646
MUTF	<i>MA_d</i>	944,25	944,25	951,80	0,9624	0,5163	0,4962
MUTF	<i>MM</i>	944,41	944,41	950,70	0,9063	0,5553	0,5043
MUTF	<i>MM_d</i>	946,38	946,38	953,93	0,9619	0,5320	0,4655
Total	<i>AA</i>	950,11	950,11	956,40	0,9900	0,9583	0,3327
Total	<i>AA_d</i>	951,21	951,21	958,76	0,9859	0,9433	0,3539
Total	<i>MA</i>	942,11	942,11	948,40	0,9638	0,6128	0,4463
Total	<i>MA_d</i>	944,26	944,26	951,81	0,9909	0,6109	0,5805
Total	<i>MM</i>	944,73	944,73	951,02	0,8174	0,6374	0,6299
Total	<i>MM_d</i>	946,34	946,34	953,89	0,9624	0,6037	0,5699

AIC - Critério de informação de Akaike, AICc - Critério de informação de Akaike corrigido e BIC - Critério de informação Bayesiano, *AA* - tendência aditiva e erro aditivo, *MA* - tendência aditiva e erro multiplicativo, *MM* - tendência multiplicativa e erro multiplicativo,

AA_d - tendência aditiva amortecida e erro aditivo, *MA_d* - tendência aditiva amortecida e erro multiplicativo,

MM_d - tendência multiplicativa amortecida e erro multiplicativo, *p-valor para cada um dos testes. JB - Jarque-Bera

Fonte: Elaboração própria

Já para a categoria MUTF, observa-se que o modelo *MA* apresenta consistentemente os valores mais baixos em todos os critérios de informação. Especificamente, tanto o AIC quanto o AICc registram 941,86 para

o modelo *MA*, enquanto o BIC é ligeiramente mais alto, com 948,15. O segundo modelo mais parcimonioso é o modelo *MA_d*, com valores de AIC, AICc e BIC de 944,25, 944,25 e 951,80, respectivamente.

Na categoria total, observa-se que o modelo *MA* apresenta os menores valores de AIC, AICc e BIC, indicando sua parcimônia em relação aos demais modelos considerados. O modelo *MA_d* tem os próximos menores valores de AICc (944,26) e BIC (951,81). Estes são os dois modelos mais parcimoniosos para esta categoria.

Quanto ao teste de Ljung-Box todos os modelos possuem valores maiores que 0,05, indica que os resíduos não possuem autocorrelação significativa. Já para o teste de ARCH mostra que os modelos são ARCH possuem heteroscedasticidade e por fim o teste de Jaque Bera a categoria tratamento de resíduos.

A Tabela 4.6 apresenta as métricas de erro com os modelos com maior capacidade preditiva. Em muitos casos, os modelos com tendência aditiva e erro multiplicativo (*MA*) demonstram uma capacidade preditiva mais consistente e superior em comparação com os modelos alternativos. Na categoria indústria, observamos que tanto o modelo *AA* quanto o *MA* têm desempenho muito próximo em termos de RMSE, MAE e MAPE. Ambos os modelos apresentam valores semelhantes nessas métricas, indicando uma capacidade preditiva equivalente.

Tabela 4.6 – Capacidade preditiva dos modelos de suavização exponencial para emissões de CO_2e

Categoria	Modelo	RSME (t)	MAE (t)	MAPE (%)
Processos industriais	<i>AA</i>	91.908,44	79.360,58	52,37
Processos industriais	<i>MA</i>	92.927,43	80.414,55	52,99
Tratamento de resíduos	<i>MA</i>	59.989,31	48.882,34	2,56
Tratamento de resíduos	<i>MM</i>	150.790,80	128.876,00	6,72
Energia	<i>AA</i>	2.778.360,74	2.714.883,91	33,45
Energia	<i>MM_d</i>	2.167.636,45	2.077.571,70	25,89
Agropecuária	<i>MA</i>	52.004,35	51.577,42	1,58
Agropecuária	<i>MM</i>	63.123,08	53.685,36	1,67
MUTF	<i>MA</i>	25.257.310,22	24.036.477,68	24,38
MUTF	<i>MA_d</i>	30.542.172,14	27.778.763,65	26,39
Total	<i>MA</i>	24.469.309,98	23.595.828,21	21,29
Total	<i>MA_d</i>	29.957.411,48	27.673.078,82	23,42

MAE - Erro Médio Absoluto, MAPE - Erro Percentual Médio Absoluto, RMSE - Raiz do Erro Médio Quadrático, *AA* - tendência aditiva e erro aditivo, *MA* - tendência aditiva e erro multiplicativo, *MM* - tendência multiplicativa e erro multiplicativo, *MM_d* - tendência multiplicativa amortecida e erro multiplicativo

Fonte: Elaboração própria

Para a categoria tratamento de resíduos, o modelo *MA* demonstra ser superior em comparação com o *MM*, conforme evidenciado por seus valores substancialmente menores de RMSE, MAE e MAPE. Isso sugere que o modelo *MA* é mais capaz de fazer previsões precisas para as emissões de CO_2e provenientes do tratamento de resíduos.

Na categoria energia, o modelo *MM_d* mostra um desempenho significativamente melhor do que o *AA*, com valores consideravelmente menores de RMSE, MAE e MAPE. Isso indica que o modelo *MM_d* é mais eficaz na previsão das emissões de CO_2e relacionadas à energia.

Para a categoria agropecuária, o modelo *MA* parece ter uma leve vantagem sobre o *MM*, com valores ligeiramente menores de RMSE, MAE e MAPE, sugerindo uma capacidade preditiva ligeiramente superior.

Na categoria MUTF e total, os modelos *MA* apresentam um desempenho melhor do que os modelos

correspondentes com tendência amortecida MA_d . Os modelos MA têm valores mais baixos de RMSE, MAE e MAPE, indicando uma capacidade preditiva mais robusta.

Com base na Tabela 4.7, os modelos mais indicados para previsão da tendência de cada uma das séries consideradas, observa-se diferentes recomendações para os modelos mais parcimoniosos e aqueles com maior capacidade preditiva.

Tabela 4.7 – Modelos mais indicados para previsão da tendência de cada uma das séries

Categorias	Mais parcimonioso		Maior capacidade preditiva		Modelo indicado
	1º	2º	1º	2º	
Processos industriais	MA	AA	AA	MA	AA
Tratamento de resíduos	MA	MM	MA	MM	MA
Energia	AA	MM_d	MM_d	AA	MM_d
Agropecuária	MA	MM	MA	MM	MA
MUTF	MA	MA_d	MA	MA_d	MA
Total	MA	MA_d	MA	MA_d	MA

AA - Tendência aditiva e erro aditivo, MA - tendência aditiva e erro multiplicativo, MM - tendência multiplicativa e erro multiplicativo, MA_d - tendência aditiva amortecida e erro multiplicativo, MM_d - tendência multiplicativa amortecida e erro multiplicativo

Fonte: Elaboração própria

Para a categoria processos industriais, o modelo mais parcimonioso é o MA , enquanto o modelo com maior capacidade preditiva é o AA . No entanto, o modelo indicado é o AA , pois possui o menor erro.

Para a categoria tratamento de resíduos, tanto o modelo mais parcimonioso quanto o com maior capacidade preditiva é o MA , tornando-o também o modelo indicado devido ao menor erro.

Na categoria produção de energia, o modelo mais parcimonioso é o AA , enquanto o modelo com maior capacidade preditiva é o MM_d . No entanto, o modelo indicado é o MM_d , pois possui o menor erro.

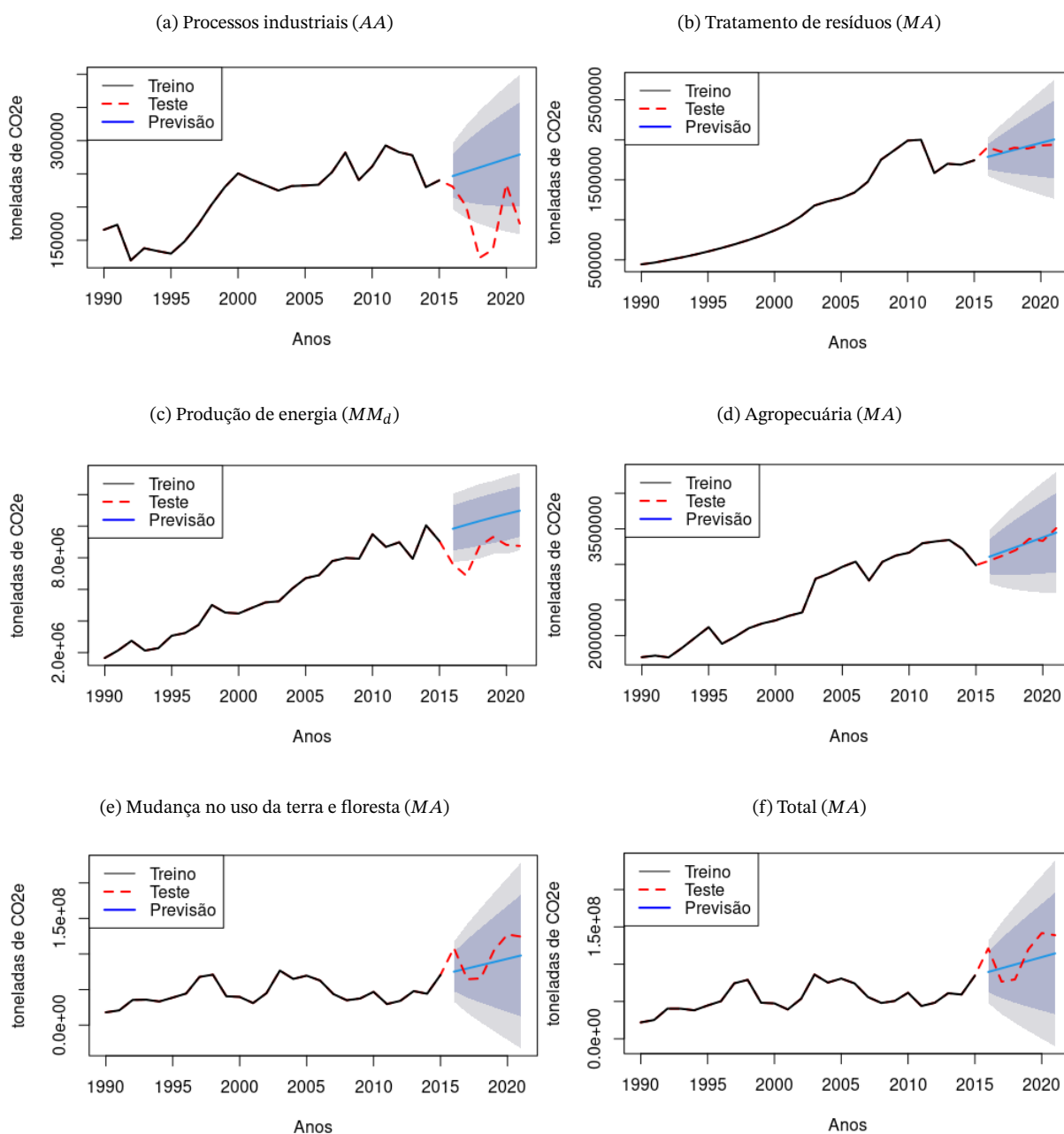
Para a categoria agropecuária, tanto o modelo mais parcimonioso quanto o com maior capacidade preditiva é o MA , sendo também o modelo indicado devido ao menor erro.

Na categoria MUTF, o modelo mais parcimonioso e o de maior capacidade preditiva são ambos o MA , tornando-o também o modelo indicado devido ao menor erro. Os modelos mais adequados para previsão da tendência nessa categoria são determinados com base na combinação de parcimônia e capacidade preditiva, com foco na minimização do erro. Os modelos escolhidos variam de acordo com as características específicas de cada série e a importância atribuída à precisão das previsões.

Na categoria total, observamos uma conclusão semelhante, onde o modelo mais parcimonioso e com maior capacidade preditiva é o MA . Portanto, ele é indicado como o modelo preferencial devido ao menor erro. A escolha dos modelos para previsão da tendência nessa categoria também é baseada na combinação de parcimônia e capacidade preditiva, considerando as características individuais de cada série temporal e a relevância da precisão das previsões.

E, finalmente, a Figura 4.5 apresenta os resultados em gráfico dos modelos ETS selecionados. Os modelos variam em sua precisão conforme a categoria de emissões. Enquanto alguns modelos, como MA para tratamento de resíduos e agropecuária, mostram desempenho satisfatório, outros, como AA para energia e MA para MUTF, apresentam previsões menos precisas. Como acontece com a categoria total, a mudança de comportamento das séries torna os modelos ETS inadequados para a realização de previsões futuras. Confirmando a importância de considerar diferenças entre os modelos para adequá-los aos diferentes tipos de séries temporais.

Figura 4.5 – Os modelos mais indicados para realizar previsões com Suavização Exponencial



AA - Tendência aditiva e erro aditivo, MA - tendência aditiva e erro multiplicativo, MM_d - tendência multiplicativa amortecida e erro multiplicativo
 Fonte: Elaboração própria

4.3 Metodologia de Box-Jenkins - ARIMA

Nesta seção, os resultados do modelo ARIMA aplicado às séries temporais são apresentados. A Tabela 4.8 mostra quantas diferenciações foram necessárias para tornar as séries estacionárias, usando testes como ADF, KPSS e PP, considerando tanto tendência quanto nível. Todos os testes foram realizados com um nível de significância de 0,05.

Quanto à análise de estacionariedade das séries, observa-se que a série indústria requer diferentes números de diferenças, dependendo do teste aplicado. Para torná-la estacionária, é necessária uma diferença no teste ADF com tendência, uma diferença no teste KPSS com tendência e uma diferença no teste PP com

tendência. Enquanto isso, a série resíduos exige duas diferenças no teste ADF com tendência e uma no teste PP com tendência, sem necessidade de diferenças no teste KPSS com tendência.

Tabela 4.8 – Número de diferenças necessárias para uma série estacionária

Séries	ADF		KPSS		PP	
	trend	level	trend	level	trend	level
Processos industriais	1	1	1	1	1	1
Tratamento de resíduos	2	1	0	1	1	1
Energia	1	1	0	1	1	1
Agropecuária	1	1	1	1	1	1
MUTF	2	1	0	0	2	1
Total	2	1	0	1	2	1

ADF - Teste Augmented Dickey-Fuller, KPSS - teste de Kwiatkowski, Phillips, Schmidt e Shin,

PP - teste de Phillips-Perron, *trend* - tendência e *level* - nível

Fonte: Elaboração própria

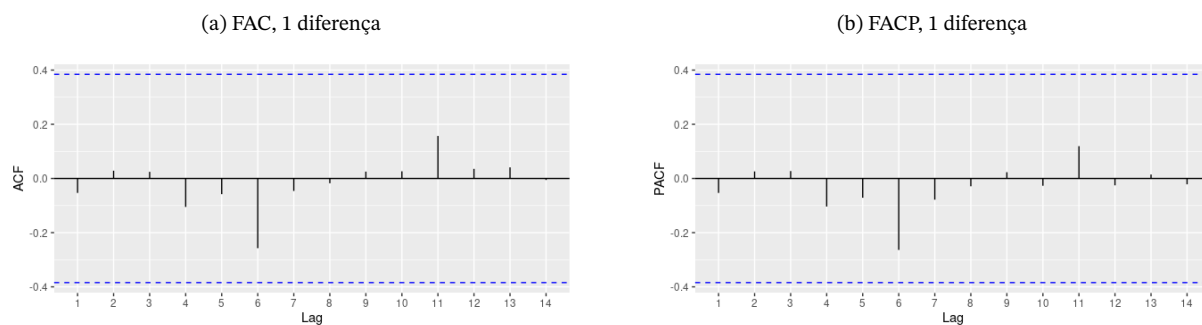
Já a série energia também pede uma diferença no teste ADF com tendência e no teste PP com tendência, com nenhuma diferença necessária no teste KPSS com tendência. Enquanto, a série agropecuária apresenta um padrão consistente, exigindo uma única diferença para se tornar estacionária nos três testes com tendência.

No entanto, a série MUTF revela uma maior variação, com duas diferenças necessárias no teste ADF com tendência e no teste PP com tendência, enquanto nenhuma diferença é requerida no teste KPSS com tendência. Já, ao avaliar a série total, observa-se que são necessárias duas diferenças no teste ADF com tendência, uma diferença no teste KPSS com tendência e uma diferença no teste PP com tendência para alcançar a estacionariedade. Assim o modelo ARIMA foi aplicado as séries indústria, energia e agropecuária com uma diferença e as séries resíduos, MUTF e total com a segunda diferença.

A análise da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) foram usadas na identificação de um modelo ARIMA ao lidar com séries temporais. A FAC mostra como as observações em diferentes momentos estão correlacionadas. No momento atual ($k=0$), a autocorrelação é sempre 1, já que reflete a correlação da série com ela mesma. Por outro lado, a FACP revela padrões de correlação entre os momentos, uma vez que elimina a influência dos lags intermediários.

Ao analisar a série temporal referente aos processos industriais, a Figura 4.6a, que exhibe o gráfico da FAC, revela uma acentuada redução na primeira defasagem. De maneira correspondente, o gráfico da Figura 4.6b, representando a FACP, demonstra uma diminuição gradual. Esses padrões sugerem que um modelo com um termo AR de ordem 1 ($p = 1$) e um termo de MA de ordem 1 ($q = 1$) pode ser apropriado para capturar a dinâmica subjacente dessa série temporal.

Figura 4.6 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para processos industriais

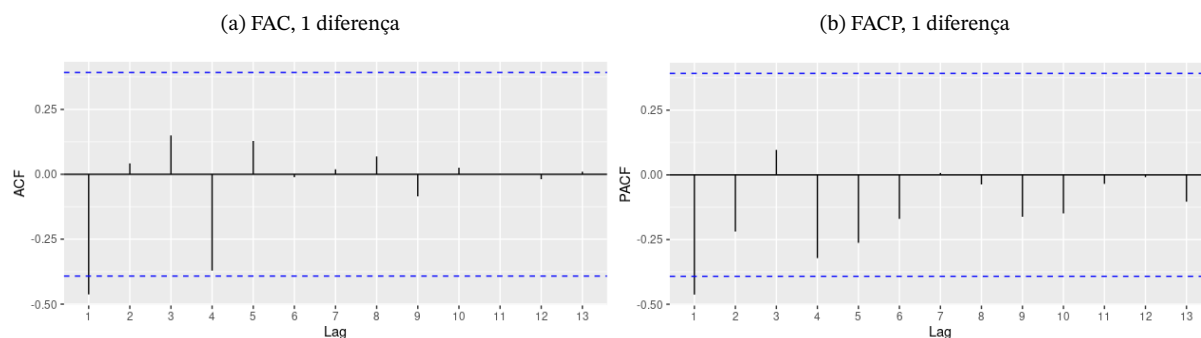


Fonte: Elaboração própria

Para refinar a análise, foram exploradas combinações de valores para p e q no intervalo de 0 a 1. O processo de ajuste do modelo considerou diversas combinações para determinar a configuração ótima que melhor se adequa aos padrões de autocorrelação observados. Essa abordagem permitiu uma seleção mais precisa dos parâmetros do modelo ARIMA.

Na análise da série temporal associada ao tratamento de resíduos, conforme retratado na Figura 4.7a, a FAC revela uma autocorrelação negativa para o lag 1, sugerindo uma correlação positiva entre a observação atual e a anterior, com uma defasagem de dois intervalos. No entanto, ao examinar lags subsequentes, nota-se uma autocorrelação variável, com alguns lags apresentando valores negativos, indicando a possível presença de uma tendência de reversão ao longo do tempo. Na FACP, destaca-se uma correlação notavelmente positiva para o lag 1, atingindo 0.549, persistindo ao longo do primeiro lag. No entanto, ao analisar lags subsequentes, percebem-se oscilações na correlação parcial, sugerindo a possível existência de um padrão autorregressivo (AR) alternado na série temporal, conforme ilustrado na Figura 4.7b.

Figura 4.7 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para tratamento de resíduos

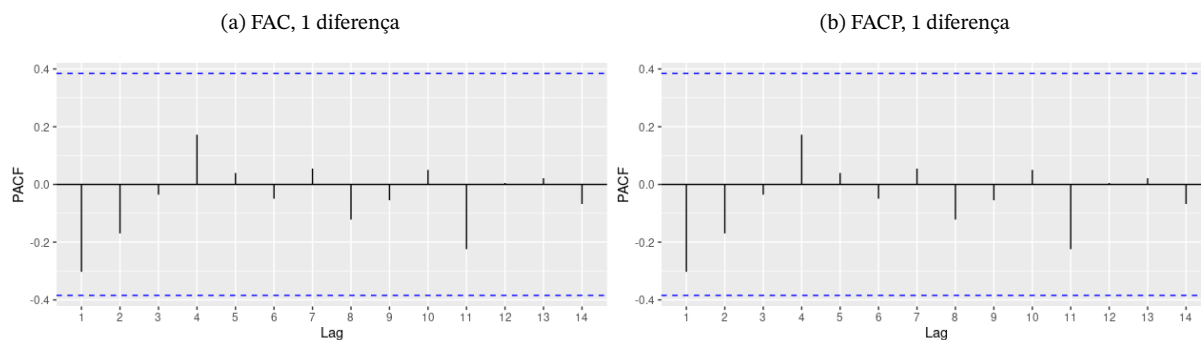


Fonte: Elaboração própria

Os resultados da análise indicam a presença de um padrão autorregressivo (AR), especialmente em lags alternados. Diante disso, sugere-se explorar modelos ARIMA, com a inclusão de um componente AR de ordem 1.

Ao examinar a interpretação dos resultados para a categoria de produção de energia, conforme ilustrado na Figura 4.8a, nota-se uma rápida diminuição das autocorrelações após o primeiro lag na FAC. Essa tendência sugere que a aplicação da diferenciação de ordem 1 foi eficaz em mitigar a autocorrelação serial na série temporal relacionada à produção de energia.

Figura 4.8 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para produção de energia

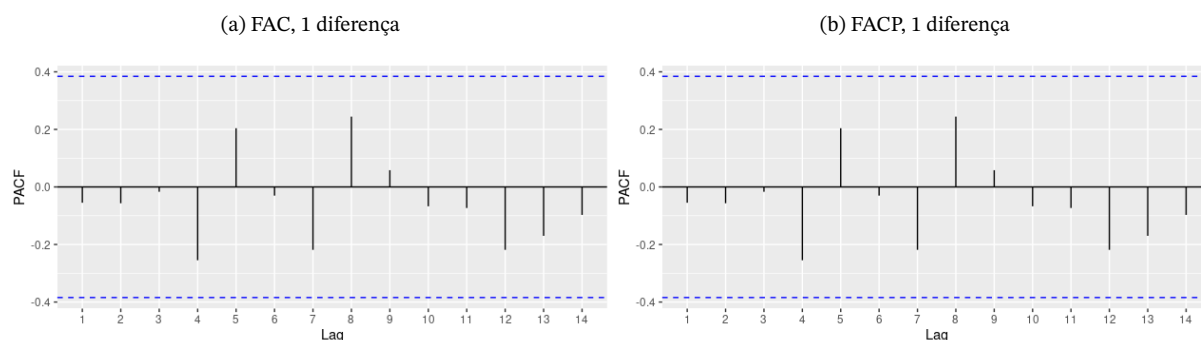


Fonte: Elaboração própria

Já na FACP, destaca-se que a correlação parcial significativa é observada exclusivamente no lag 1. Esse padrão específico sugere que um modelo autorregressivo AR(1) pode ser uma escolha apropriada para modelar a série temporal. Esses resultados são consistentes com a abordagem adotada na análise da série de processos industriais. Para esta série também foram realizadas as combinações de p e q no intervalo de 0 a 1.

Ao examinar a série temporal relacionada à categoria agropecuária, observa-se, na FAC apresentada na Figura 4.9a, uma redução nas autocorrelações logo após o primeiro lag. Essa rápida diminuição sugere a presença de um padrão indicativo de dependência temporal imediata entre as observações. De maneira análoga, a Figura 4.9b, representando a FACP, revela uma diminuição gradual das correlações parciais, esses padrões sugerem que um modelo ARIMA com um termo AR de ordem 1 ($p = 1$) e um termo MA de ordem 1 ($q = 1$) pode ser apropriado para modelar a série temporal agropecuária. A análise foi refinada ao explorar diversas combinações de valores para p e q , no intervalo de 0 a 1, visando uma seleção mais precisa dos parâmetros do modelo.

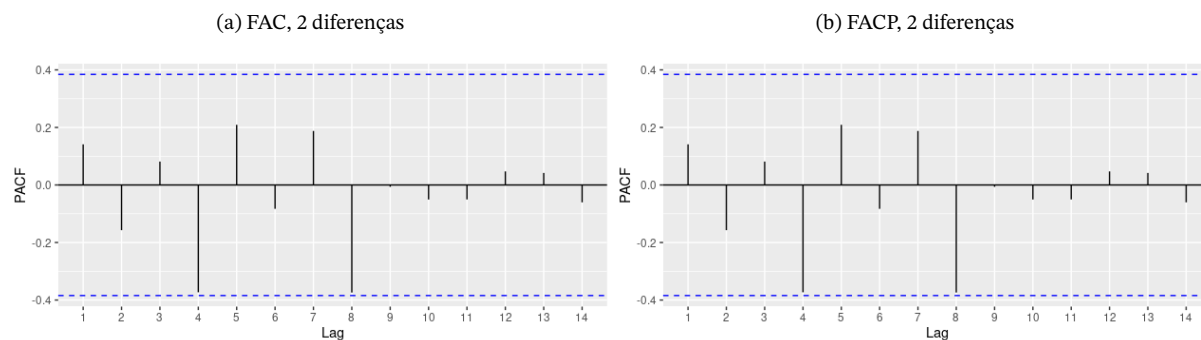
Figura 4.9 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para agropecuária



Fonte: Elaboração própria

Ao examinar a categoria relacionada a MUTF, a análise da FAC, conforme ilustrada na Figura 4.10a, destaca uma autocorrelação notavelmente significativa e positiva no lag 1. A autocorrelação no lag 2 revela significância, mas com uma direção negativa. Ao longo dos lags subsequentes, observa-se uma clara tendência de diminuição gradual dessas autocorrelações. Paralelamente, a análise da FACP, apresentada na Figura 4.10b, revela que a autocorrelação parcial no lag 1 é significativa e positiva, enquanto no lag 2 é significativa e negativa. Assim como na FAC, as autocorrelações parciais nos lags subsequentes demonstram uma progressiva diminuição.

Figura 4.10 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para MUTF

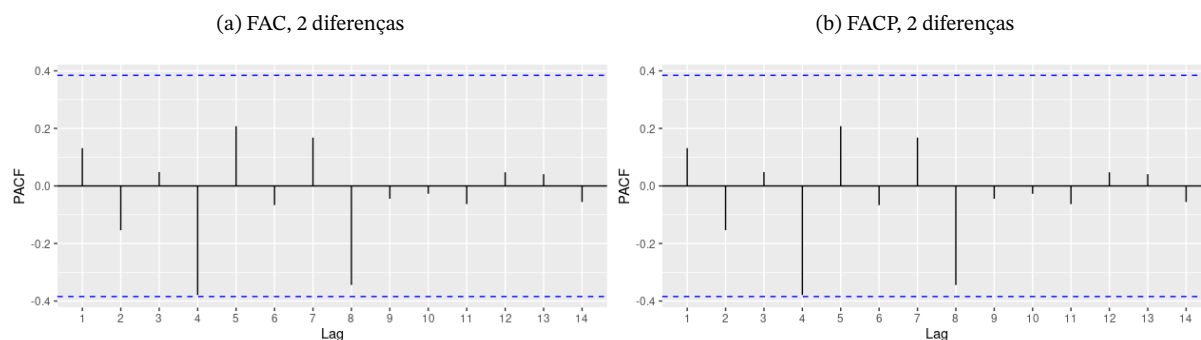


Fonte: Elaboração própria

Com base nessas análises, os parâmetros sugeridos para um modelo ARIMA seriam $p=2$ (em virtude da autocorrelação significativa no lag 2), $d=2$ (considerando a diferenciação da série duas vezes), e $q=2$ (devido à autocorrelação negativa significativa no lag 2 na FACP). Entretanto, é relevante ressaltar que ajustes e validações adicionais são necessários antes da aplicação final do modelo aos dados de teste, garantindo uma modelagem precisa da série temporal MUTF. Essa abordagem reflete a consistência metodológica utilizada nas análises de outras categorias.

A análise da série total revela padrões notavelmente similares aos da série MUTF. Na FAC apresentada na Figura 4.11a, destaca-se que a autocorrelação no lag 1 é positiva e significativa, enquanto no lag 2 é negativa e significativa. Essas autocorrelações subsequentes exibem uma clara tendência de diminuição gradual. De maneira correspondente, a análise da FACP, visualizada na Figura 4.11b, indica que a autocorrelação parcial no lag 1 é positiva e significativa, e no lag 2 é negativa e significativa. Esses padrões são seguidos por diminuições graduais nas autocorrelações parciais nos lags subsequentes.

Figura 4.11 – Autocorrelação (ACF) e a autocorrelação parcial (PACF) para a série total



Fonte: Elaboração própria

Para o modelo ARIMA, corroborando a semelhança com a série MUTF, os parâmetros sugeridos são $p=3$ (devido à autocorrelação significativa no lag 4), $d=3$ (diferenciação da série duas vezes), e $q=2$ (em função da autocorrelação negativa significativa no lag 4 na FACP). É importante destacar que ajustes e validações adicionais são necessários antes da aplicação final do modelo aos dados de teste. Essa abordagem reflete a consistência metodológica utilizada em análises anteriores.

Após a análise da estacionariedade das séries temporais apresentadas até este ponto e a definição dos parâmetros do modelo ARIMA(p,d,q), juntamente com a otimização dos parâmetros e a configuração da tendência, constante e média, buscou-se alcançar o ajuste mais parcimonioso e preciso. Posteriormente, a atenção voltou-se para a avaliação dos resíduos do modelo ajustado por meio de testes estatísticos, incluindo o teste de Ljung-Box para autocorrelação, o teste de Jarque-Bera para normalidade e o teste ARCH para heterocedasticidade condicional.

A partir deste ponto, apresenta-se os modelos mais parcimoniosos para cada uma das séries temporais, cujas informações são detalhadas na Tabela 4.9. Somente após essas verificações minuciosas, avançou-se para a etapa de previsões, projetando os próximos 6 períodos com base no modelo ajustado. A acurácia dessas previsões é então criteriosamente avaliada em comparação com os dados de teste. Para uma visualização clara e eficaz, finalizamos a análise com a geração de um gráfico que destaca as previsões em relação aos dados observados.

Conforme os resultados apresentados na Tabela 4.9, relacionados a aplicação do modelo ARIMA aplicados a série das emissões de CO_2 e originários dos processos industriais, os modelos ARIMA(0,1,0) e ARIMA(0,1,1) destacam-se ao apresentar valores relativamente baixos para esses critérios para o AIC. O modelo

(0,1,0) destaca-se por apresentar o menor conjunto de valores nessas métricas, com AIC de 597,7285, AICc de 599,8054 e BIC de 598,9866. Ele é o mais parcimonioso entre os modelos considerados.

Tabela 4.9 – Modelos mais parcimoniosos da categoria processos industrial, 1990-2015

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(0,1,0)	575,63	577,71	576,85	0,8293	0,1063	0,8282
(1,1,0)	577,61	581,94	580,05	0,8204	0,1206	0,7976
(0,1,1)	577,61	581,94	580,05	0,8204	0,1198	0,8000
(0,1,2)	579,59	586,37	583,25	0,8334	0,1306	0,7491
(2,1,0)	579,60	586,38	583,25	0,8308	0,1271	0,7690
(1,1,1)	579,61	586,39	583,26	0,8208	0,1211	0,7960
(1,1,2)	581,58	591,03	586,45	0,8343	0,1276	0,7367
(2,1,1)	581,59	591,04	586,47	0,8316	0,1265	0,7604
(2,1,2)	581,94	594,32	588,03	0,8146	0,0053	0,7952

AIC - Critério de informação Akaike, AICc - Critério de informação Akaike corrigido, BIC - Critério de informação Bayesiano

*p-valor para cada um dos testes

Fonte: Elaboração própria

Adicionalmente, ao analisar outros aspectos essenciais, verifica-se que o teste Ljung-Box para autocorrelação dos resíduos no ARIMA(0,1,0) possui um valor não significativo (0,7923), indicando a ausência de autocorrelação significativa nos resíduos. O teste Jarque-Bera, que avalia a normalidade dos resíduos, também resulta em um valor não significativo (0,1234), sugerindo que os resíduos se aproximam de uma distribuição normal. Quanto à heterocedasticidade condicional, o teste ARCH revela um valor de 0,8386, indicando a ausência significativa de heterocedasticidade.

A análise dos modelos ARIMA para a categoria tratamento de resíduos, apresentados na Tabela 4.10, mostram o modelo ARIMA(0,2,1) com os menores valores para AIC (633,96), AICc (638,30) e BIC (636,32), seguido do modelo ARIMA(1,2,1). Dado que os modelos apresentam valores muito baixos (ou zero) para os testes de Ljung-Box e Jarque-Bera e valores muito pequenos para o teste ARCH, isso sugere que os resíduos não possuem autocorrelação, seguem uma distribuição normal e não são heterocedásticos. Portanto, ambos os modelos parecem atender a esses critérios.

Tabela 4.10 – Modelos mais parcimoniosos da categoria tratamento de resíduos, 1990-2015

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(0,2,1)	633,96	638,30	636,32	0,6663	<1,0x10-16	4,48E-03
(1,2,1)	635,54	642,32	639,08	0,7272	<1,0x10-16	4,10E-03
(0,2,2)	635,61	642,40	639,15	0,7248	<1,0x10-16	4,18E-03
(0,2,3)	636,99	646,44	641,70	0,7387	<1,0x10-16	4,05E-03
(2,2,3)	637,21	652,81	644,28	0,5702	<1,0x10-16	6,57E-03
(2,2,1)	637,33	646,78	642,04	0,7342	<1,0x10-16	3,95E-03
(1,2,2)	637,46	646,92	642,18	0,7268	<1,0x10-16	4,01E-03
(1,2,0)	638,81	643,14	641,16	0,4858	4,56E-12	2,92E-03
(1,2,3)	638,97	651,35	644,86	0,7244	<1,0x10-16	4,02E-03
(3,2,1)	639,24	651,62	645,13	0,7802	<1,0x10-16	4,07E-03
(2,2,2)	639,31	651,70	645,21	0,7431	<1,0x10-16	3,97E-03
(2,2,0)	639,76	646,54	643,30	0,3787	6,66E-16	2,94E-03
(3,2,2)	640,52	656,12	647,59	0,4737	<1,0x10-16	3,90E-03
(3,2,3)	641,10	660,26	649,34	0,4885	<1,0x10-16	3,93E-03
(3,2,0)	641,72	651,17	646,43	0,4300	<1,0x10-16	2,95E-03
(0,2,0)	643,02	645,10	644,20	0,0940	1,31E-11	2,95E-03

AIC - Critério de informação Akaike, AICc - Critério de informação Akaike corrigido, BIC - Critério de informação Bayesiano

*p-valor para cada um dos testes

Fonte: Elaboração própria

Com base nos resultados apresentados na Tabela 4.11 para diferentes modelos ARIMA, o Modelo (0,1,1) demonstra os valores mais baixos em todas essas métricas, indicando uma escolha parcimoniosa e eficaz para a modelagem da série temporal. No entanto, o Modelo (0,1,2) também apresenta métricas competitivas, indicando uma alternativa. Quanto aos resultados dos testes Ljung-Box, Jarque-Bera e ARCH para os modelos apresentados sugerem que os resíduos são adequados, sem autocorrelação significativa, sem desvios de normalidade e sem evidências de heterocedasticidade condicional. Portanto, esses resultados reforçam a qualidade e a parcimoniosidade dos modelos ARIMA propostos.

Ao analisar os modelos ARIMA propostos para a categoria de produção de energia, apresentados na Tabela 4.11. Dentre as diferentes configurações testadas, o ARIMA(0,1,1) se destaca como o mais parcimonioso, apresentando os valores mais baixos em métricas como AIC (781,6951), AICc (786,0151) e BIC (785,4694).

Tabela 4.11 – Modelos mais parcimoniosos da categoria produção de energia

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(2,1,1)	748,51	757,97	753,39	0,6154	0,8556	0,9274
(1,1,2)	748,85	758,30	753,73	0,5675	0,4518	0,4173
(2,1,2)	750,21	762,59	756,31	0,7585	0,6717	0,7891
(1,1,0)	751,48	755,81	753,91	0,3331	0,8413	0,6634
(0,1,2)	751,55	758,33	755,21	0,5443	0,9520	0,5652
(0,1,0)	752,04	754,12	753,26	0,0432	0,8073	0,2127
(2,1,0)	752,04	758,82	755,70	0,1550	0,5697	0,5106
(1,1,1)	752,12	758,90	755,77	0,2453	0,5630	0,5404
(0,1,1)	752,38	756,72	754,82	0,1513	0,8576	0,5582

Critério de informação de Akaike (AIC), Akaike corrigido (AICc) e Bayesiano (BIC).

*p-valor para cada um dos testes

Fonte: Elaboração própria

Além disso, ao considerar aspectos relacionados a autocorrelação, heterocedasticidade e normalidade dos resíduos. O teste Ljung-Box não evidencia autocorrelação significativa nos resíduos (p-value = 0,8769), indicando que o modelo captura adequadamente a estrutura temporal da série. Os testes Jarque-Bera (p-value = 0,1175) e ARCH (p-value = 0,0206) não mostram desvios significativos da normalidade e heterocedasticidade condicional, respectivamente, reforçando a qualidade do ajuste do ARIMA(0,1,1).

Os dois modelos mais parcimoniosos para a categoria agropecuária, conforme apresentados na Tabela 4.12, são o ARIMA(0,1,0) e o ARIMA(0,1,1). Ambos os modelos têm valores baixos de AIC, AICc e BIC em comparação com os outros modelos listados na tabela. O modelo ARIMA(0,1,0) possuem AIC (670,97), AICc (673,05) e BIC (673,41). Este modelo é o mais simples, com apenas um termo de diferenciação e nenhum termo de autorregressão ou média móvel.

Os testes adicionais reforçam a adequação do ARIMA(0,1,0), já que o teste Ljung-Box para autocorrelação nos resíduos resulta em um valor não significativo (0,6945), indicando ausência de autocorrelação. O teste Jarque-Bera para normalidade também retorna um valor não significativo (0,2090), sugerindo que os resíduos se aproximam de uma distribuição normal. Quanto à heterocedasticidade, o teste ARCH não fornece evidências significativas (0,4894), indicando que a variabilidade dos resíduos não varia de forma condicional ao longo do tempo.

Com base na análise dos modelos ARIMA para a categoria MUTF apresentados na Tabela 4.13, a parcimoniosidade do ajuste é avaliada através de métricas importantes como AIC, AICc e BIC. Dentre os modelos considerados, o ARIMA(0,2,3) destaca-se por apresentar os valores mais baixos nessas métricas, com AIC de 903,5511, AICc de 912,9424 e BIC de 908,4266. Esses resultados sugerem que o Modelo (0,2,3) é mais parcimonioso, indicando uma escolha eficiente e equilibrada.

Quanto a autocorrelação dos resíduos, a presença de heterocedasticidade e a normalidade. Todos os

Tabela 4.12 – Modelos mais parcimoniosos da categoria agropecuária, 1990-2015

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(0,1,0)	670,97	673,05	673,41	0,6945	0,2090	0,4894
(0,1,1)	672,89	677,23	676,55	0,7460	0,2079	0,4884
(0,1,2)	674,62	681,41	679,50	0,8222	0,2351	0,5687
(1,1,0)	672,90	677,24	676,56	0,7390	0,2082	0,4871
(1,1,1)	673,37	680,16	678,25	0,8059	0,3835	0,5881
(1,1,2)	676,63	686,08	682,72	0,8281	0,2331	0,5660
(2,1,0)	674,80	681,58	679,68	0,7687	0,2042	0,5089
(2,1,1)	675,26	684,72	681,36	0,7810	0,4293	0,5968
(2,1,2)	676,70	689,08	684,01	0,9399	0,2213	0,6770

AIC - Critério de Informação Akaike, AICc - Critério de Informação Akaike Corrigido, BIC - Critério de Informação Bayesiano

*p-valor para cada um dos testes

Fonte: Elaboração própria

Tabela 4.13 – Modelos mais parcimoniosos da categoria MUTF

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(0,2,3)	865,09	874,55	869,80	0,1802	0,9406	0,7141
(2,2,3)	865,54	881,14	872,60	0,4977	0,8804	0,8836
(1,2,2)	865,84	875,30	870,55	0,1680	0,9080	0,7214
(0,2,1)	866,05	870,39	868,41	0,2208	0,9396	0,4096
(2,2,2)	866,62	879,00	872,51	0,2218	0,9516	0,7971
(1,2,3)	866,96	879,34	872,85	0,2197	0,9339	0,6914
(3,2,3)	867,17	886,33	875,41	0,5445	0,9005	0,7750
(3,2,2)	867,54	883,14	874,61	0,2101	0,9968	0,7870
(0,2,2)	868,00	874,78	871,53	0,2135	0,9574	0,4491
(1,2,1)	868,01	874,79	871,54	0,2166	0,9526	0,4414
(2,2,1)	869,76	879,21	874,47	0,1964	0,9957	0,3355
(3,2,1)	871,76	884,14	877,65	0,1977	0,9954	0,3315
(2,2,0)	872,48	879,26	876,02	0,0596	0,8657	0,9458
(1,2,0)	873,87	878,20	876,22	0,1320	0,2914	0,7083
(3,2,0)	874,41	883,86	879,12	0,0541	0,8240	0,9452
(0,2,0)	875,99	878,07	877,17	0,0106	0,5155	0,8385

AIC - Critério de Informação Akaike, AICc - Critério de Informação Akaike Corrigido, BIC - Critério de Informação Bayesiano

*p-valor para cada um dos testes

Fonte: Elaboração própria

modelos exibem não significância nos testes de Ljung-Box, indicando a ausência de autocorrelação nos resíduos. Os testes de Jarque-Bera revelam valores não significativos, sugerindo que os resíduos se aproximam de uma distribuição normal. Além disso, os testes ARCH apontam para a ausência de heterocedasticidade condicional nos resíduos.

Como as séries possuem características muito semelhantes ao analisar os resultados apresentados na Tabela 4.14 o modelo ARIMA(2, 2, 2) se destaca como aquele que apresenta os menores valores de AIC, AICc e BIC, indicando um ajuste mais eficiente em comparação com os demais modelos considerados. No entanto, o modelo ARIMA(1, 2, 2) exibe valores mais baixos de AIC e AICc, embora o BIC seja ligeiramente superior, indicando uma possível precisão adicional com uma complexidade adicional.

Na Tabela 4.14, a categoria total observa-se que o modelo ARIMA(1,2,2) se destaca como a opção mais parcimoniosa, apresentando valores competitivos em todas as métricas de informação, com AIC de 903.8038, AICc de 913.1951 e BIC de 908.6793. Além disso, os resultados dos testes adicionais também são favoráveis. Comparativamente, outros modelos, como o ARIMA(2,2,3), o ARIMA(0,2,3), o ARIMA(3,2,3), o ARIMA(2,2,2) e o ARIMA(1,2,3), também exibem métricas aceitáveis, mas o ARIMA(1,2,2) se destaca como a escolha mais parcimoniosa.

Tabela 4.14 – Modelos mais parcimoniosos da categoria total, 1990-2015

ARIMA	AIC	AICc	BIC	Ljung-Box	Jarque-Bera	ARCH
(0,2,3)	864,62	874,07	869,33	0,1172	0,8695	0,7606
(3,2,3)	865,42	884,58	873,67	0,4805	0,8064	0,7233
(1,2,2)	865,59	875,04	870,30	0,1195	0,8452	0,7962
(0,2,1)	866,04	870,38	868,40	0,1696	0,8857	0,4249
(2,2,2)	866,37	878,75	872,26	0,1495	0,9109	0,8135
(1,2,3)	866,57	878,95	872,46	0,1429	0,8654	0,7280
(3,2,2)	867,40	883,00	874,46	0,1385	0,9791	0,7821
(2,2,3)	867,94	883,54	875,01	0,1516	0,8662	0,6465
(0,2,2)	867,98	874,76	871,51	0,1643	0,8952	0,4682
(1,2,1)	867,99	874,78	871,53	0,1672	0,8904	0,4594
(2,2,1)	869,70	879,15	874,41	0,1475	0,9923	0,3392
(3,2,1)	871,68	884,06	877,57	0,1549	0,9886	0,3092
(2,2,0)	872,83	879,61	876,37	0,0479	0,7922	0,9564
(1,2,0)	873,98	878,31	876,34	0,1111	0,2275	0,7570
(3,2,0)	874,79	884,24	879,50	0,0447	0,7503	0,9573
(0,2,0)	876,08	878,16	877,26	0,0078	0,4524	0,7935

AIC - Critério de Informação Akaike, AICc - Critério de Informação Akaike Corrigido, BIC - Critério de Informação Bayesiano

*p-valor para cada um dos testes

Fonte: Elaboração própria

A Tabela 4.15 apresenta resultados detalhados dos 2 (dois) modelos ARIMA com melhor ajuste, segundo os critérios de informação aplicados às séries, destacando o desempenho através das métricas RMSE, MAE e MAPE. Essas métricas são fundamentais para avaliar a precisão das previsões em relação aos valores reais, nesta pesquisa o critério para selecionado o modelo foi o critério de precisão RMSE.

Iniciando pela série indústria, o modelo ARIMA(0, 1, 0) destaca-se por apresentar os menores valores de RMSE, MAE e MAPE entre as opções disponíveis, indicando ser a escolha mais favorável. O valor desta série possui o MAPE igual a 40,90 por possuir uma quebra de estrutura próximo do final das observações usadas para ajuste, no final desta série há uma alteração nos valores proporcionando uma queda acentuada das emissões.

Tabela 4.15 – Capacidade preditiva dos modelos ARIMA para emissões de CO₂e

Categoria	ARIMA	RMSE (t)	MAE (t)	MAPE (%)
Processos industriais	(0,1,0)	71.083,27	56.637,67	39,00
Processos industriais	(1,1,0)	70.829,54	56.317,49	38,82
Tratamento de resíduos	(0,2,1)	76.630,08	61.612,39	3,21
Tratamento de resíduos	(1,2,1)	76.327,09	61.373,16	3,20
Energia	(2,1,1)	2.290.185,00	2.219.377,00	27,50
Energia	(1,1,2)	2.330.383,00	2.253.281,00	27,91
Agropecuária	(0,1,0)	117.358,42	93.161,25	2,76
Agropecuária	(1,1,0)	100.092,18	74.141,38	2,19
MUTF	(0,2,3)	37.430.516,00	29.774.840,00	25,41
MUTF	(2,2,3)	27.662.772,00	25.460.917,00	24,77
Total	(0,2,3)	38.076.897,00	30.479.774,00	23,25
Total	(3,2,3)	27.278.421,00	25.991.303,00	22,74

RMSE - Raiz do Erro Médio Quadrático, MAE - Erro Médio Absoluto, MAPE - Erro Percentual Médio Absoluto

MUTF - Mudança do Uso da Terra e Floresta

Fonte: Elaboração própria

Para a categoria tratamento de resíduos, a análise do RMSE indica que o modelo (0,2,1) é mais adequado em comparação com o modelo (1,2,1). Além de apresentar os menores valores de MAE e MAPE, sugerindo ser a opção mais adequada.

No setor produção de energia, o modelo ARIMA(0,1,1) se destaca como a escolha preferencial, pois apresenta um RMSE de 1.181.166,60, significativamente inferior ao RMSE de 1.402.307,00 do modelo ARIMA(2,1,0). Isso indica que o modelo tem um desempenho melhor na previsão da série temporal de energia.

Na categoria agropecuária, a melhor escolha recai sobre o modelo ARIMA(0,1,0), uma vez que possui um RMSE de 96.752,23, enquanto o modelo ARIMA(0,1,0) apresenta um RMSE mais elevado de 116.883,20. Portanto, o modelo ARIMA(0,1,1) é mais indicado para previsões nesta categoria.

No que diz respeito à categoria MUTF, foram considerados os modelos ARIMA(0,2,3) e ARIMA(2,2,3). O modelo ARIMA(0,2,3) apresenta um RMSE de 29.376.668,55, enquanto o modelo (2,2,3) exibe um valor ligeiramente superior de 32.045.668,71. Portanto, selecionando o modelo com o menor RMSE, o modelo (0,2,3) é a escolha preferida para a categoria MUTF.

Por fim, na categoria total, o modelo ARIMA(2,2,3) é a melhor escolha, pois apresenta RMSE (36.083.737,20), MAE (27.212.106,63) e MAPE (32,45%) são mais baixos que o modelo ARIMA(1,2,2) que possui um RMSE (36.177.474,00), MAE (27.563.042,20) e MAPE (32,70%).

A Tabela 4.16 apresenta os dois modelos mais parcimoniosos, os com maior capacidade preditiva e o modelo indicado para realizar previsões. Na categoria indústria, o modelo (1,1,0) é o mais parcimonioso, enquanto em termos de capacidade preditiva, o modelo (1,1,0) é o mais indicado por possuir o menor erro.

Tabela 4.16 – Modelos mais indicados para previsão da tendência de cada uma das séries

Categorias	Mais parcimonioso		Maior capacidade preditiva		Modelo indicado
	1º	2º	1º	2º	
Processos industriais	(0,1,0)	(1,1,0)	(1,1,0)	(0,1,0)	(1,1,0)
Tratamento de resíduos	(0,2,1)	(1,2,1)	(1,2,1)	(0,2,1)	(1,2,1)
Energia	(2,1,1)	(1,1,2)	(2,1,1)	(1,1,2)	(2,1,1)
Agropecuária	(0,1,0)	(0,1,1)	(1,1,0)	(0,1,0)	(0,1,0)
MUTF	(0,2,3)	(2,2,3)	(2,2,3)	(0,2,3)	(0,2,3)
Total	(0,2,3)	(3,2,3)	(3,2,3)	(0,2,3)	(0,2,3)

MUTF - Mudança do Uso da Terra e Floresta

Fonte: Elaboração própria

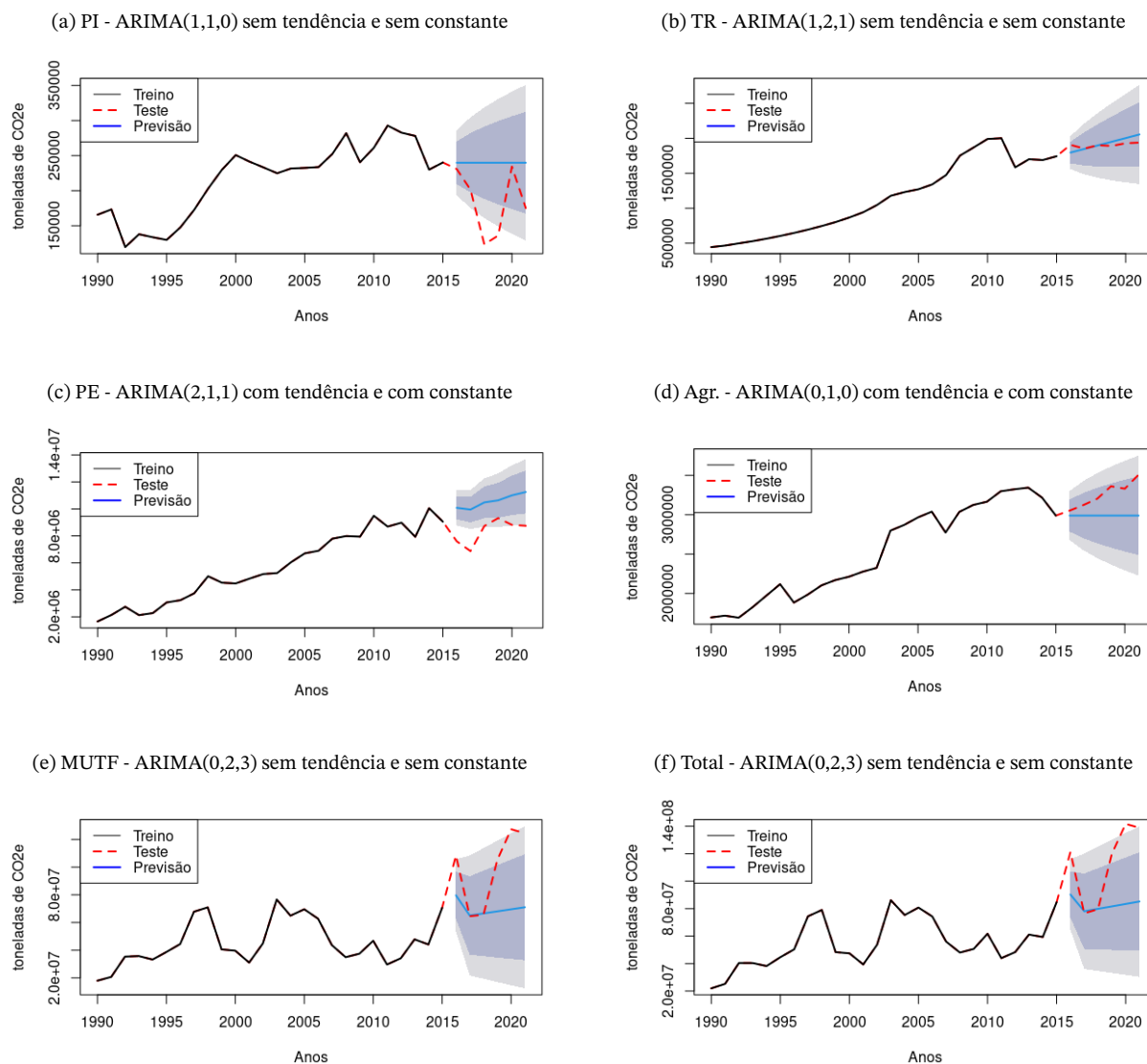
Na categoria resíduos, o modelo (0,2,1) é o mais parcimonioso, enquanto o (1,2,1) é o mais indicado em termos de capacidade preditiva. Com as métricas de erro apresentam valores semelhantes, o modelo (1,2,1) foi selecionado por ser o que apresenta o menor erro.

Na categoria energia, o modelo (2,1,1) é o mais parcimonioso, enquanto o modelo (2,1,1) é o mais indicado em termos de capacidade preditiva. Nesta categoria o modelo (2,1,1) é o mais parcimonioso e possui o menor erro, sendo o modelo selecionado para esta categoria.

Na categoria Agropecuária, o modelo (0,1,0) é o mais parcimonioso, com apenas um componente de diferenciação, o que se mostra suficiente para prever a tendência. Entretanto, em termos de capacidade preditiva, o modelo (1,1,0) é mais indicado. Já na categoria MUTF, o modelo (0,2,3) é o mais parcimonioso, enquanto o modelo (2,2,3) demonstra uma melhor capacidade preditiva, sendo selecionado por sua maior capacidade de ajuste. Na categoria total, os resultados são semelhantes aos da categoria MUTF, com o modelo (0,2,3) sendo o mais parcimonioso e o modelo (3,2,3) sendo o mais indicado em termos de capacidade preditiva. Em ambas as categorias, a escolha dos modelos considera a combinação de parcimônia e capacidade preditiva, para minimizar o erro e fornecer previsões precisas.

O gráfico dos modelos selecionados são apresentados na Figura 4.12, na categoria de processos industriais, as previsões permanecem constantes ao longo dos anos de 2017 a 2020, em torno de 230,826, com os intervalos de confiança de 95%, indicando uma crescente incerteza nas previsões. Essa estabilidade sugere que o modelo adotado assume uma tendência constante, enquanto a amplitude do intervalo de confiança oferece uma medida da incerteza associada às previsões. É importante ressaltar que a série apresenta uma quebra de estrutura próxima ao final das observações usadas para o ajuste, seguida por uma queda acentuada nas emissões de CO_2e nos anos de 2016 a 2018.

Figura 4.12 – Os modelos ARIMA selecionados para realizar previsões



PI - processos industriais, TR - tratamento de resíduos, PE - produção de energia, Agr. - agropecuária
Fonte: Elaboração própria

A Figura 4.12b mostra a categoria de tratamento de resíduos um aumento contínuo ao longo dos anos, com previsões ascendentes de 1,96 milhão em 2017 para 2,24 milhões em 2022. A ampliação gradual dos limites do intervalo de confiança sugere variabilidade nas estimativas, destacando a importância de considerar a incerteza associada às previsões, mesmo diante da perspectiva otimista de crescimento.

Para a produção de energia apresentada na Figura 4.12c, as previsões e intervalos de confiança apontam para um aumento consistente ao longo do período de 2017 a 2022. Os valores previstos aumentam progressivamente, indicando uma trajetória de crescimento positiva. A amplitude dos intervalos de confiança oferece uma

faixa de variação plausível, proporcionando uma perspectiva otimista para o setor e auxiliando em decisões de planejamento.

Na categoria agropecuária, as previsões e intervalos de confiança sugerem um aumento consistente na produção ao longo dos anos. A ampliação dos intervalos de confiança indica maior incerteza nas previsões à medida que avançamos no tempo, conforme a Figura 4.12d.

Os dados para a categoria MUTF indicam uma tendência relativamente constante em torno de 96,75 milhões ao longo dos anos. A Figura 4.12e apresenta a ampliação do intervalo de confiança sugere uma maior incerteza nas previsões futuras. A variabilidade nos dados pode aumentar ao longo do tempo, ressaltando a importância de considerar não apenas a estimativa pontual, mas também a amplitude do intervalo de confiança nas decisões relacionadas a essa categoria.

Na categoria total, as previsões indicam uma tendência de estabilidade em torno de 110,96 milhões ao longo dos anos de 2017 a 2022. No entanto, o aumento da amplitude do intervalo de confiança reflete uma crescente incerteza nas projeções futuras, conforme a Figura 4.12f.

4.4 Discussão entre as previsões realizadas pela Suavização Exponencial e ARIMA

Nesta seção, realiza-se uma discussão entre os modelos ETS e ARIMA, considerando o critério de informação AIC usado para escolher um modelo ETS entre os modelos candidatos ETS ou um modelo ARIMA entre os modelos candidatos ARIMA. Entretanto, não pode ser aplicado para comparar diretamente modelos ETS e ARIMA, uma vez que essas categorias não pertencem à mesma classe de modelos e a probabilidade é calculada de maneiras distintas.

As Tabelas 4.17 e 4.18 mostram os modelos escolhidos e suas métricas de desempenho, como RMSE, MAE e MAPE. Em seguida, analisa-se cada categoria detalhadamente, discutindo os resultados e características dos modelos.

Tabela 4.17 – Modelos selecionados dos métodos de Suavização Exponencial (ETS)

Categorias	ETS	RMSE (t)	MAE (t)	MAPE (%)
Processos industriais	AA	91.908,44	79.360,58	52,37
Tratamento de resíduos	MA	59.989,31	48.882,34	2,56
Energia	MM _d	2.167.636,45	2.077.571,70	25,89
Agropecuária	MA	52.004,35	51.577,42	1,58
MUTF	MA	25.257.310,22	24.036.477,68	24,38
Total	MA	24.469.309,98	23.595.828,21	21,29

RMSE - Raiz do Erro Médio Quadrático, MAE - Erro Médio Absoluto, MAPE - Erro Percentual Médio Absoluto,

AA - Tendência aditiva e erro aditivo, MA - Tendência aditiva e erro multiplicativo,

MM_d - tendência multiplicativa amortecida e erro multiplicativo

Fonte: Elaboração própria

Começando com a categoria de processos industriais, o método ETS(AA) teve melhor desempenho, com um RMSE de 91.908,44 e MAPE de 52,37%. No entanto, o ARIMA parece capturar melhor a tendência constante observada nesta série temporal, com uma diferença no MAPE de 13,55%.

Para a categoria tratamento de resíduos, a ETS(MA) apresentou um RMSE de 59.989,31 e MAPE igual a 2,56%. Por outro lado, o modelo ARIMA(1,2,1) exibiu um desempenho semelhante, com um RMSE de 76.327,09 e MAPE igual a 3,20%. Com base nos critérios de RMSE e MAPE, o modelo ETS(MA) tem um desempenho superior em comparação com o modelo ARIMA(1,2,1). A diferença entre o MAPE dos dois MAPE é de menos de 1%, resultados do MAPE mais próximos entre os métodos de previsão.

Na categoria produção de energia, ambas as técnicas mostraram dificuldades. A ETS(MM_d) resultou em um alto RMSE de 2.167.636,45 e MAPE de 25,89%. O modelo ARIMA(2,1,1) apresenta um RMSE de 2.290.185,00

e MAPE de 27,50%. Portanto, com base nos critérios de avaliação de desempenho, o modelo ETS(MM_d) tem um desempenho melhor na previsão da tendência em comparação com o modelo ARIMA(2,1,1).

Na categoria de agropecuária, a ETS(MA) obteve um RMSE de 52.004,35 e MAPE de 1,58%, enquanto o modelo ARIMA(0,1,0) teve um desempenho ainda melhor, com um RMSE de 117.358,42 e MAPE de 2,76%. O modelo ETS(MA) teve um desempenho superior na previsão da tendência, com uma diferença de 1,18% no MAPE. Para esta série temporal os modelos apresentam previsões satisfatórias.

Tabela 4.18 – Modelos selecionados das técnicas ARIMA

Categoria	ARIMA	RMSE (t)	MAE (t)	MAPE (%)
Indústria	(1,1,0)	70.829,54	56.317,49	38,82
Resíduos	(1,2,1)	76.327,09	61.373,16	3,20
Energia	(2,1,1)	2.290.185,00	2.219.377,00	27,50
Agropecuária	(0,1,0)	117.358,42	93.161,25	2,76
MUTF	(0,2,3)	37.430.516,00	29.774.840,00	25,41
Total	(0,2,3)	38.076.897,00	30.479.774,00	23,25

Erro Médio Absoluto (MAE), Erro Percentual Médio Absoluto (MAPE),

Raiz do Erro Médio Quadrático (RMSE)

Fonte: Elaboração própria

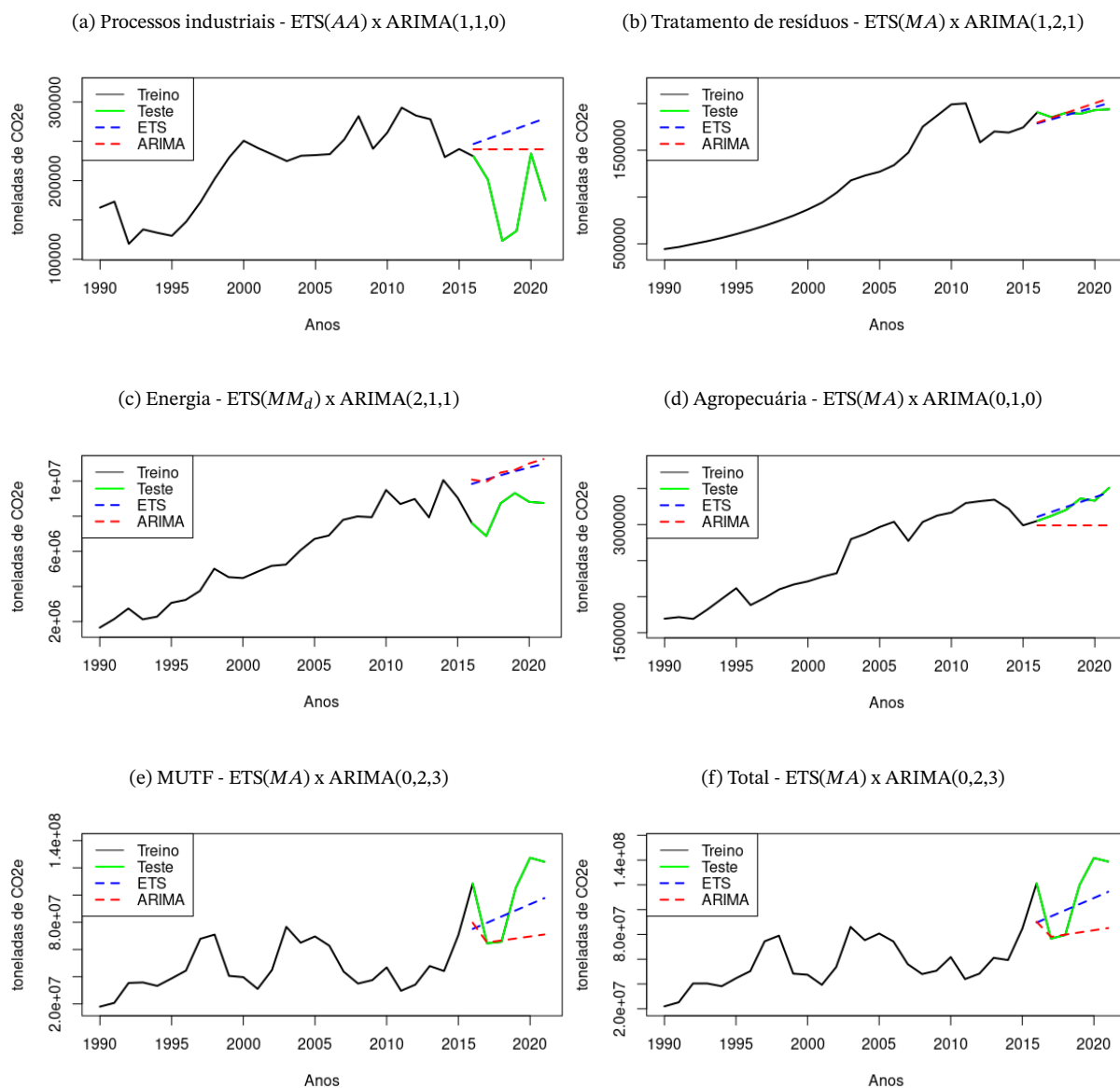
Na categoria de MUTF, a ETS(MA) apresentou um RMSE elevado de 25.257.310,22 e MAPE de 24,38%, enquanto o modelo ARIMA(0,2,3) teve um desempenho ainda pior, com um RMSE de 37.430.516,00 e MAPE de 25,41%. Ambas as técnicas enfrentaram dificuldades nessa categoria.

Para a categoria total, a ETS(MA) teve um RMSE de 24.469.309,98 e MAPE de 21,29%, enquanto o modelo ARIMA(0,2,3) teve um RMSE mais alto de 38.076.897,00 e MAPE de 32,45%. A suavização exponencial parece mais adequada para essa categoria.

A Figura 4.13 mostra os gráficos para cada série, permitindo uma visualização comparativa do desempenho preditivo dos modelos ETS e ARIMA.

Em geral, os modelos ETS e ARIMA apresentam desempenho relativamente equivalente na previsão das séries temporais, com pequenas variações nos valores previstos. As categorias de tratamento de resíduos e agropecuária se destacam por terem previsões mais precisas em ambos os métodos.

Figura 4.13 – Comparação entre os modelos de Suavização Exponencial (ETS) e ARIMA



AA - Tendência aditiva e erro aditivo, MA - Tendência aditiva e erro multiplicativo,
MM_d - tendência multiplicativa amortecida e erro multiplicativo

Fonte: Elaboração própria

5 Conclusão

Inicialmente, esta pesquisa tinha como objetivo identificar o mecanismo gerador dos dados que compõem a série temporal de emissão de CO_2e no estado do Amazonas, fornecendo subsídios para elaboração de políticas públicas preventivas. Os resultados indicam que ambos os métodos produziram resultados semelhantes. A única exceção foi a série de processos industriais, que apresentou previsões divergentes entre os métodos, enquanto as outras séries mostraram previsões bastante semelhantes, com uma diferença de MAPE inferior a 2%.

Quanto aos objetivos específicos, a análise exploratória dos dados revelou que as séries processos industriais, MUTF e totais apresentaram quebras de estruturais, especialmente durante os testes. Essas quebras tornaram as previsões difíceis de serem capturadas pelos métodos utilizados nesta pesquisa, indicando a necessidade de investigar métodos mais adequados para lidar com séries com essas características.

Os resultados sugerem que a escolha entre os modelos de Suavização Exponencial e Box-Jenkins depende das características específicas dos dados e das séries temporais em análise. Ambas as abordagens têm vantagens e limitações, e a seleção do método mais apropriado deve considerar fatores como a presença de tendências e padrões de autocorrelação nos dados.

Com base nesta pesquisa, é possível afirmar que há evidências de um aumento nas emissões de CO_2e ao longo do período analisado em todas as categorias. Esse aumento é corroborado pelo aumento tanto na média quanto na mediana das emissões, juntamente com a presença de uma cauda longa à direita nos dados das categorias.

Compreender as variações nas emissões de CO_2e ao longo do tempo é crucial para o desenvolvimento de políticas públicas eficazes de mitigação e adaptação às mudanças climáticas. Os resultados destacam a importância de considerar a complexidade dos sistemas ambientais e socioeconômicos ao formular estratégias de redução de emissões.

Em estudos futuros, pretende-se aplicar técnicas avançadas, como modelos baseados em aprendizado de máquina, para melhorar ainda mais a capacidade preditiva e a compreensão dos mecanismos subjacentes às emissões de CO_2e . Essas abordagens podem fornecer informações adicionais e contribuir para um entendimento mais preciso do fenômeno.

Em resumo, esta pesquisa contribui para o avanço do conhecimento sobre as emissões de CO_2e no estado do Amazonas e oferece subsídios importantes para o planejamento e implementação de políticas ambientais e climáticas mais eficazes na região. No entanto, são necessários estudos complementares para aprimorar a compreensão dos processos subjacentes às emissões de gases de efeito estufa e desenvolver modelos de previsão e avaliação mais robustos.

Referências

- AGUIAR, L. V.; FORTES, J. D. N.; MARTINS, E. Neutralização compensatória de carbono - estudo de caso: indústria do setor metal mecânico, rio de janeiro (RJ). **Engenharia Sanitaria e Ambiental**, v. 21, p. 197–205, 2016. Citado na p. 13.
- AKHTER, M. F.; HASSAN, D.; ABBAS, S. Predictive ARIMA model for coronal index solar cyclic data. **Astronomy and Computing**, v. 32, p. 100403, 2020. Citado na p. 37.
- ALAM, T.; ALARJANI, A. A comparative study of CO_2 emission forecasting in the gulf countries using autoregressive integrated moving average, artificial neural network, and Holt-Winters exponential smoothing models. **Adv. Meteorol.**, v. 2021, p. 1–9, 2021. Citado nas pp. 15, 34 e 35.
- ALAMRI, S.; KHAN, S. Artificial intelligence based modelling for predicting CO_2 emission for climate change mitigation in saudi arabia. **Int. J. Adv. Comput. Sci. Appl.**, v. 14, n. 4, 2023. Citado nas pp. 31, 33 e 34.
- ALBUQUERQUE, R. C. de. **Modelagem em séries temporais: aplicação em dados de precipitação na região do sertão pernabucano - Brasil**. Tese (Doutorado) — UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO, Recife, PE, jan. 2019. Citado na p. 24.
- ANA. **Rio Negro atinge menor nível em Manaus desde 1902**. 2023. <https://www.gov.br/ana/pt-br/assuntos/noticias-e-eventos/noticias/rio-negro-atinge-menor-nivel-em-manaus-desde-1902>. Accessed: 2023-11-27. Citado na p. 16.
- ANNING, A. K.; OFORI-YEBOAH, A.; BAFFOUR-ATA, F.; OWUSU, G. Climate change manifestations and adaptations in cocoa farms: Perspectives of smallholder farmers in the adansi south district, ghana. **Current Research in Environmental Sustainability**, v. 4, p. 100196, 2022. Citado na p. 13.
- AZEVEDO de; JUNIOR, C. C.; JUNIOR, A. B.; CREMER, M. d. S. Seeg initiative estimates of brazilian greenhouse gas emissions from 1970 to 2015. **Scientific Data**, v. 5, 05 2018. Citado nas pp. 18 e 19.
- BAKAY, M. S.; AĞBULUT Ümit. Electricity production based forecasting of greenhouse gas emissions in turkey with deep learning, support vector machine and artificial neural network algorithms. **Journal of Cleaner Production**, v. 285, p. 125324, 2021. ISSN 0959-6526. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0959652620353695>. Citado na p. 13.
- BOWERSOX, D. J.; CLOSS, D. **Transporte Logístico: Planejamento e Gerenciamento da Cadeia de Suprimentos**. São Paulo: Atlas, 2007. Citado na p. 15.
- BOX, G. E. P.; AL, E. **Time series analysis : forecasting and control**. [S.l.]: John Wiley & Sons, 2015. Citado nas pp. 19, 20, 21, 25 e 26.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis : forecasting and control**. [S.l.]: John Wiley, 2008. Citado na p. 24.
- British Petroleum. **Statistical Review of World Energy**. 2023. <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>. Acessado em : 5 mai. 2023. Citado na p. 14.

- BROWN, J.; CANE, V. R. An analysis of the limiting method. **Br. J. Stat. Psychol.**, v. 12, n. 2, p. 119–126, 1959. Citado na p. 23.
- BROWN, R. G. Exponential smoothing for predicting demand. 1956. Disponível em: <http://legacy.library.ucsf.edu/tid/dae94e00>. Citado nas pp. 21 e 22.
- CALVIN, K.; DASGUPTA, D.; KRINNER, G.; MUKHERJI, A. **IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [core writing team, H. lee and J. romero (eds.)]. IPCC, Geneva, Switzerland. [S.l.], 2023.** Citado na p. 16.
- CANADELL, J. G.; PEDRO, M. S. **Global carbon and other biogeochemical cycles and feedbacks coordinating lead authors.** 2022. Citado na p. 13.
- CASTANHEIRA, N. P. **Métodos quantitativos.** Curitiba - PR: InterSaberes, 2013. Citado na p. 19.
- COPERTINO, M. S.; GARCIA, C. A.; TURRA, A.; CIOTTI, Á. M. Zonas costeiras. In: Carlos A. Nobre e José A. Marengo (Ed.). **Mudanças climáticas em rede: um olhar interdisciplinar.** São José dos Campos, SP: INCT, 2017. p. 307–366. Citado na p. 14.
- COSTA, F. M.; SÁFADI, T. Comparação estatística de duas séries de material particulado (mp10) na cidade de são paulo. **Revista Brasileira de Biometria**, n. 3, 2010. Citado na p. 27.
- CRIPPA, M.; SOLAZZO, E.; GUIZZARDI, D.; TUBIELLO, F. N.; LEIP, A. Climate goals require food systems emission inventories. **Nat. Food**, v. 3, n. 1, p. 1–1, 2022. Citado na p. 14.
- DANKWA, P.; CUDJOE, E.; AMUAH, E. E. Y.; KAZAPOE, R. W.; AGYEMANG, E. P. Analyzing and forecasting rainfall patterns in the manga-bawku area, northeastern ghana: Possible implication of climate change. **Environmental Challenges**, v. 5, p. 100354, 2021. ISSN 2667-0100. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2667010021003280>. Citado nas pp. 13 e 34.
- Defesa Civil-RS, C. M. **Defesa Civil apresenta balanço de ações durante evento meteorológico.** 2023. <https://www.defesacivil.rs.gov.br/defesa-civil-apresenta-balanco-de-acoes-durante-evento-climatico>. Accessed: 2023-11-27. Citado na p. 16.
- EMILIANO, P. C. **Critérios de informação: como eles se comportam em diferentes modelos?** Tese (Doutorado) — Universidade Federal de Lavras, UFL, 2013. Citado na p. 31.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflations. **Econometrica**, v. 50, p. 987–1007, 1982. Citado nas pp. 30 e 37.
- ESCOBARI, D.; GARCIA, S.; MELLADO, C. Identifying bubbles in latin american equity markets: Phillips-Perron-based tests and linkages. **Emerg. Mark. Rev.**, v. 33, p. 90–101, 2017. Citado na p. 37.
- ESTEVO, J. S. China no âmbito da mudança climática. negociações exteriores e políticas domésticas. **Desafios**, v. 32, n. 1, p. 1, 2020. Citado nas pp. 13 e 18.
- FRIEDLINGSTEIN, P.; O’SULLIVAN, M.; JONES. Global carbon budget 2022. **Earth Syst. Sci. Data**, v. 14, n. 11, p. 4811–4900, 2022. Citado na p. 13.
- GARDNER, E. S. Exponential smoothing: The state of the art. **J. Forecast.**, v. 4, n. 1, p. 1–28, 1985. Citado na p. 22.

- GIDDENS, A. **A política das mudanças climáticas**. São Paulo: UNESP, 2010. Citado na p. 13.
- HASSAN, M. K.; KAZAK, H.; ADIGÜZEL, U.; GUNDUZ, M. A.; AKCAN, A. T. Convergence in islamic financial development: Evidence from islamic countries using the fourier panel KPSS stationarity test. **Borsa Istanb. Rev.**, v. 23, n. 6, p. 1289–1302, 2023. Citado na p. 37.
- HATFIELD, J. L.; BOOTE, K. J.; KIMBALL, B. A.; ZISKA, L. H.; IZAURRALDE, R. C.; ORT, D.; THOMSON, A. M.; WOLFE, D. Climate impacts on agriculture: Implications for crop production. **Agronomy Journal**, v. 103, p. 351–370, 03 2011. Citado na p. 13.
- HYNDMAN; ATHANASOPOULOS, R. G. **Forecasting: Principles and Practice**. Monash University, Australia: OTexts, 2023. Citado nas pp. 21, 22, 23, 24, 27, 30, 31 e 38.
- HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D.; GROSE, S. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439–454, 2002. Citado nas pp. 21 e 22.
- IBGE. **Amazonas | Cidades e Estados | IBGE**. 2022. Citado na p. 36.
- ICBE. **Calculating Greenhouse Gases**. 2000. <https://www.icbe.com/emissions/calculate.asp>. Citado na p. 18.
- IPCC. **AR5 Climate Change 2013: The Physical Science Basis — IPCC**. [S.L.]: IPCC, 2013. Citado na p. 18.
- Ippc. **Sixth Assessment Report**. 2021. Citado na p. 14.
- JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. **Int. Stat. Rev.**, v. 55, n. 2, p. 163, 1987. Citado nas pp. 30 e 37.
- JAVANMARD, M. E.; GHADERI, S. F. A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data. **Sustain. Cities Soc.**, v. 82, n. 103886, p. 103886, 2022. Citado nas pp. 31, 33, 34 e 35.
- JUNQUEIRA, H. S.; MEDEIROS, D. L.; COHIM, E. Gerenciamento de resíduos sólidos urbanos de feira de santana: demanda energética e pegada de carbono. **Engenharia Sanitaria e Ambiental**, v. 27, p. 125–139, 02 2022. Citado na p. 14.
- KALE, R. V.; POHEKAR, S. D. Electricity demand and supply scenarios for maharashtra (india) for 2030: An application of long range energy alternatives planning. **Energy Policy**, v. 72, p. 1–13, 2014. Citado nas pp. 32 e 33.
- KHASHEI, M.; BIJARI, M.; ARDALI, G. A. R. Improvement of Auto-Regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). **Neurocomputing**, v. 72, p. 956–967, 2009. Citado nas pp. 21 e 26.
- LIN, L.; LI, M.; HOU, X.; PIPRANI, A. Z. The resource curse in least developed countries: The roles of foreign direct investment, energy efficiency, and electricity access. **Resour. Policy**, v. 89, n. 104564, p. 104564, 2024. Citado na p. 37.
- LINTUNEN, J.; RAUTIAINEN, A. On physical and social-cost-based CO_2 equivalents for transient albedo-induced forcing. **Ecol. Econ.**, v. 190, n. 107204, p. 107204, 2021. Citado na p. 18.

- LIU, L.; ZONG, H.; ZHAO, E.; CHEN, C.; WANG, J. Can china realize its carbon emission reduction goal in 2020: From the perspective of thermal power development. **Appl. Energy**, v. 124, p. 199–212, 2014. Citado nas pp. 32 e 33.
- LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 1978. Citado nas pp. 29 e 37.
- MARTINS, F.; MACHADO, J. A. F.; ESTEVES, P. S. Modelling taylor rule uncertainty: an application to the euro area. **Economic Modelling**, v. 21, n. 3, p. 561–572, 2004. Citado na p. 21.
- MATTOS, V. L. D. de; NAKAMURA, L. R.; KONRATH, A. C.; BORNIA, A. C. Modeling the commercial electricity demand in santa catarina, using the Box-Jenkins methodology. **International Journal of Development Research**, v. 11, p. 48190–48197, 2021. Citado na p. 29.
- MCTI. **Inventário Nacional de GEE subsidia planos do governo na área ambiental. Ministério da Ciência, Tecnologia e Inovação**. 2020. <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2020/10/inventario-nacional-de-gee-subsidia-planos-do-governo-na-area-ambiental>. Accessed: 2024-1-23. Citado na p. 14.
- MEDEIROS, A. S. S.; BATISTA, C. E.; CALDERARO, G. Effects on manaus CO concentrations by power plant fuel switching. In: **Air Quality Conference Brazil 3rd CMAS SOUTH AMERICA 2017, 2017, Vitória. Anais Air Quality Conference Brazil**. [S.l.: s.n.], 2017. Citado na p. 41.
- MORETTIN, P. A.; TOLOI, C. C. **Análise de séries temporais**. São Paulo: Editora Blucher, 2018. Citado nas pp. 19, 21, 25, 26 e 37.
- NAIM, I.; MAHARA, T. Comparative analysis of univariate forecasting techniques for industrial natural gas consumption. **Int. J. Image Graph. Signal Process.**, v. 10, n. 5, p. 33–44, 2018. Citado na p. 15.
- NASCIMENTO, M.; CRUZ, C. D.; PETERNELLI, L. A.; CAMPANA, A. C. M.; PINTO, D. S.; FERREIRA, R. P. Teste dos sinais para tendência: Uma aplicação em melhoramento de plantas. **Revista Brasileira de Biometria**, n. 4, p. 19–30, 2008. Citado na p. 27.
- NOGUEIRA, D. A.; SÁFADI, T. Previsão de preços na bovinocultura de corte. **Organizações Rurais e Agroindustriais/Rural and Agro-Industrial Organizations**, v. 2, p. 1–10, 04 2011. Citado na p. 27.
- NYONI, T.; BONGA, W. G. Prediction of CO_2 emissions in india using arima models. 2019. Citado nas pp. 15 e 21.
- Observatório do Clima, S. **Total Emissions | SEEG - System Gas Emissions Estimation**. 2023. Citado nas pp. 14, 18 e 19.
- Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change [core writing team. In: PACHAURI, R. K.; MEYER, L. A. (Ed.). **Synthesis Report**. Geneva, Switzerland: [s.n.], 2014. Citado na p. 19.
- PEGELS, C. C. Exponential forecasting: some new variations. **Management Science**, v. 15, n. 5, 1969. Citado na p. 22.
- PÉREZ, F. L. **Análise de Dados Categóricos**. 2023. <http://leg.ufpr.br/~lucambio/ADC>. Accessed: 2024-1-3. Citado na p. 31.

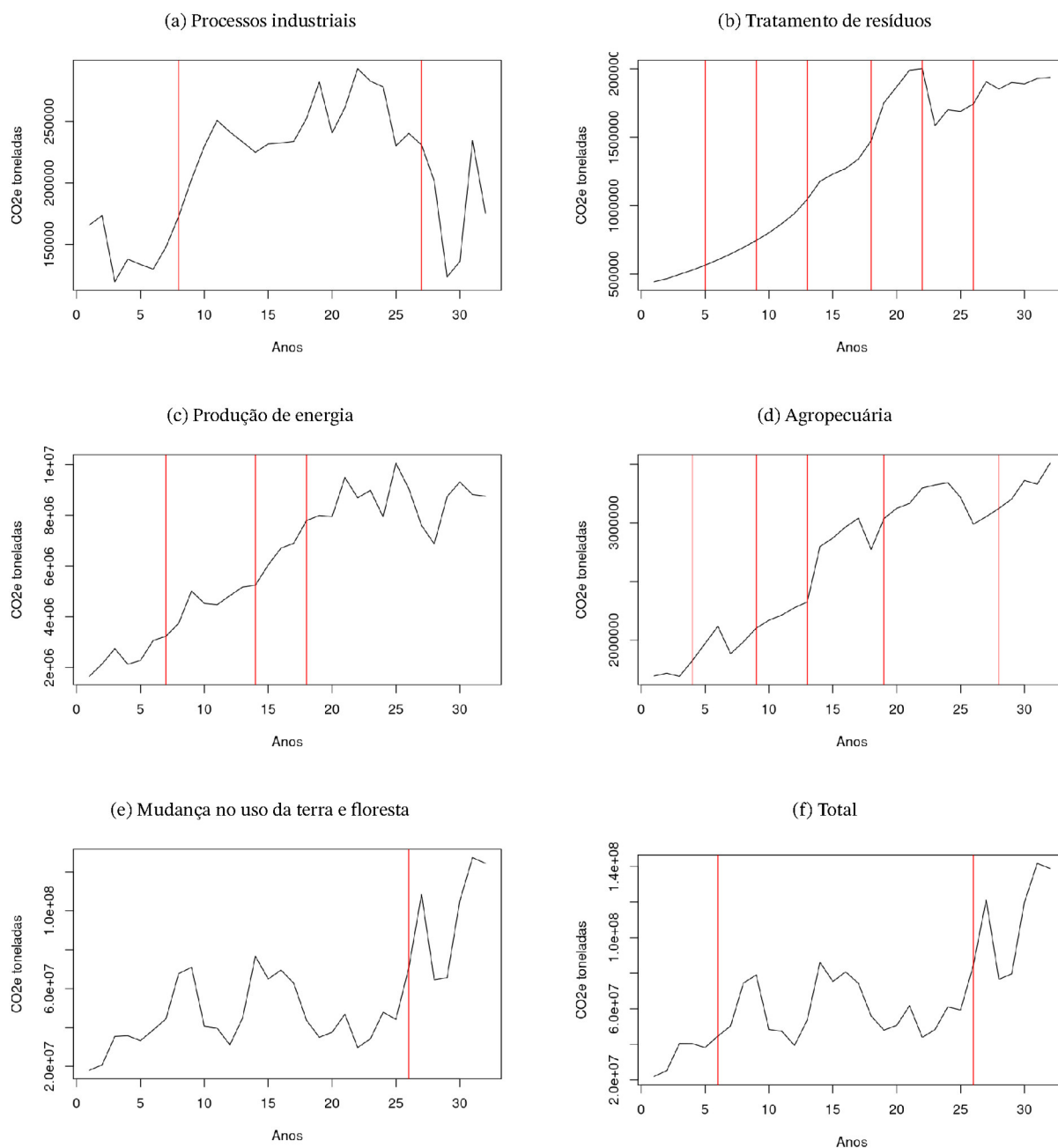
- PERONE, G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. **Eur. J. Health Econ.**, v. 23, n. 6, p. 917–940, 2022. Citado nas pp. 32 e 37.
- PFAFF, B. **Analysis of integrated and cointegrated time series with R**. Nova Iorque, NY, USA: Springer, 2011. Citado na p. 38.
- PINHO, J. A. do N. **Variações Sazonais de Propriedades Óticas de Aerossóis e Identificação de Diferentes Tipos na Amazônia Baseado em Dados AERONET**. Dissertação (Mestrado) — Universidade Federal do Amazonas, Amazonas, Brasil, 2019. Citado na p. 14.
- POTENZA, R. F.; TSAI, D. S.; CARVALHO, K.; SHIMBO, J.; SILVA, C.; SOUZA, E. **Calculadora de emissões por pessoa no Brasil – nota metodológica**. São Paulo, 2021. Citado na p. 14.
- QUINTANA, G.; CARDOSO, A. M.; POTENZA, R. F. **Análise das emissões de gases de efeito estufa e suas implicações para as metas climáticas do Brasil 1970-2021**. [S.l.]: SEEG, 2023. <https://seeg.eco.br/wp-content/uploads/2023/11/SEEG-10-anos.pdf>. Accessed: 2024-1-1. Citado na p. 19.
- R Core Team. **R: The R Project for Statistical Computing**. 2023. Citado na p. 38.
- RAGSDALE, C. **Spreadsheet modeling and decision analysis: A practical introduction to business analytics**. 9. ed. Florence, AL, USA: South-Western College Publishing, 2021. Citado na p. 15.
- ROBERT, H.; DAVID, S. **Stoffer - Time Series Analysis and its Applications with R Examples**. [S.l.]: Springer, 2017. Citado na p. 19.
- SCOLASTRICI, A. dos S. S.; VILLELA, A. A. **3ª Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudanças Climáticas**. [S.l.], 2016. Citado na p. 14.
- SEEG. **Emissions Map | SEEG - System Gas Emissions Estimation**. 2021. Citado nas pp. 19, 36 e 40.
- SEN, P.; ROY, M.; PAL, P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. **Energy (Oxf.)**, v. 116, p. 1031–1038, 2016. Citado nas pp. 32 e 33.
- SVETUNKOV, I.; KOURENTZES, N.; ORD, J. K. Complex exponential smoothing. **Naval Research Logistics (NRL)**, v. 69, p. 1108–1123, 08 2022. Citado na p. 21.
- TANEJA, K.; AHMAD, S.; AHMAD, K.; ATTRI, S. D. Time series analysis of aerosol optical depth over New Delhi using Box–Jenkins ARIMA modeling approach. **Atmos. Pollut. Res.**, v. 7, n. 4, p. 585–596, 2016. Citado na p. 21.
- TAYLOR, J. W. Exponential smoothing with a damped multiplicative trend. **Int. J. Forecast.**, v. 19, n. 4, p. 715–725, 2003. Citado na p. 22.
- TAYRA, F.; REIS, J. A. dos. Impactos dos subsídios aos combustíveis fósseis: Impostos sobre carbono e desdobramentos no Brasil. **Contexto Geográfico**, v. 5, n. 26/11/2020, p. 116–132, dez. 2020. Citado na p. 13.
- THABANI, N.; HAPANYENGWI, H. O.; MUTONGI, C. Modeling and forecasting carbon dioxide emissions in China using autoregressive integrated moving average (ARIMA) models. **EPRA International Journal of Multidisciplinary Research (IJMR)**, v. 5, p. 215–224, abr. 2019. Citado nas pp. 15 e 21.

- TSAY, R. S. **Analysis of Financial Time Series**. [S.l.]: Wiley, 2010. Citado na p. 15.
- TSAY, R. S. **An introduction to analysis of financial data with R**. 1. ed. Hoboken, NJ, USA: Wiley-Blackwell, 2014. Citado na p. 20.
- TSOLKAS, C.; SPILIOTIS, E.; SARMAS, E.; MARINAKIS, V.; DOUKAS, H. Dynamic energy management with thermal comfort forecasting. **Build. Environ.**, v. 237, n. 110341, p. 110341, 2023. Citado na p. 31.
- VERÍSSIMO, A. J.; ALVES, C. C.; HENNING, E.; AMARAL, C. E. D.; CRUZ, A. C. D. Métodos estatísticos de suavização exponencial Holt-Winters para previsão de demanda em uma empresa do setor metal mecânico. **Rev. Gest. Ind.**, v. 8, n. 4, 2013. Citado na p. 15.
- Wikipedia contributors. **Amazonas**. 2024. <https://pt.wikipedia.org/w/index.php?title=Amazonas&oldid=67480904>. Accessed: NA-NA-NA. Citado na p. 36.
- YÖRÜK, G.; BAC, U.; YERLIKAYA-ÖZKURT, F.; ÜNLÜ, K. D. Strategic electricity production planning of turkey via mixed integer programming based on time series forecasting. **Mathematics**, v. 11, n. 8, p. 1865, 2023. Citado nas pp. 33 e 34.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art". **International Journal of Forecasting**, v. 14, n. 1, p. 35–62, 1998. Citado na p. 26.
- ZHAO, L.; LI, Z.; QU, L. Forecasting of beijing PM2.5 with a hybrid ARIMA model based on integrated AIC and improved GS fixed-order methods and seasonal decomposition. **Heliyon**, v. 8, p. e12239, 2022. Citado na p. 37.
- ZIVOT, E.; WANG, J. **Modeling financial time series with S-Plus®**. [S.l.]: Springer Science & Business Media, 2007. v. 191. Citado nas pp. 20 e 21.

Apêndices

Apêndice A – Quebra de estrutura

Figura A.1 – Quebra de estrutura, teste Chow



Fonte: Elaboração própria