

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE
MESQUITA FILHO”

Instituto de Química – Campus de Araraquara

Bacharelado em Química

**MODELO DE SÉRIES TEMPORAIS: COMPARAÇÃO DE MODELOS PREDITIVOS
DA EVOLUÇÃO DA CONCENTRAÇÃO DE CO₂ NA ATMOSFERA**

MARIANA DE SIQUEIRA BARBOSA

Araraquara

2024

MARIANA DE SIQUEIRA BARBOSA

**MODELO DE SÉRIES TEMPORAIS: COMPARAÇÃO DE MODELOS PREDITIVOS
DA EVOLUÇÃO DA CONCENTRAÇÃO DE CO₂ NA ATMOSFERA**

Trabalho de Conclusão de Curso
apresentado à Universidade Estadual
Paulista “Júlio de Mesquita Filho”, sob
a orientação do Prof. Dr. Jorge Manuel
Vieira Capela.

Araraquara, 12 de novembro de 2024.

B238m Barbosa, Mariana de Siqueira

Modelo de séries temporais : comparação de modelos preditivos da evolução da concentração de CO₂ na atmosfera / Mariana de Siqueira Barbosa. -- Araraquara, 2024

42 f. : il.

Trabalho de conclusão de curso (Bacharelado - Química) - Universidade Estadual Paulista (UNESP), Instituto de Química, Araraquara

Orientador: Jorge Manuel Vieira Capela

1. Previsão com Metodologia de Box-Jenkins. 2. Análise de séries temporais. 3. Gases estufa. 4. Modelagem de dados. 5. Suavização (Estatística). I. Título.

MARIANA DE SIQUEIRA BARBOSA

MODELO DE SÉRIES TEMPORAIS: COMPARAÇÃO DE MODELOS PREDITIVOS DA
EVOLUÇÃO DA CONCENTRAÇÃO DE CO₂ NA ATMOSFERA

Tese apresentada ao Instituto de Química, Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para a obtenção do título de Bacharel em Química.

Araraquara, 27 de novembro de 2024

BANCA EXAMINADORA

Prof. Dr. Jorge Manuel Vieira Capela

Instituto de Química – UNESP, Araraquara

Profa. Dra. Érica Regina Filletti Nascimento

Instituto de Química – UNESP, Araraquara

Prof. Dr. Henrique Antonio Mendonça Faria

Instituto de Química – UNESP, Araraquara

RESUMO

Este trabalho tem por objetivo modelar as séries temporais das concentrações de CO₂, utilizando técnicas estatísticas e de aprendizado de máquina, a fim de identificar padrões, tendências e realizar previsões futuras. O efeito estufa, que tem sido intensificado pela crescente concentração de gases como CO₂ na atmosfera, representa um dos maiores desafios ambientais da atualidade. Esses gases desempenham um papel fundamental na alteração do clima global, e sua acumulação tem consequências diretas para o meio ambiente e a saúde humana. A necessidade de enfrentar essa questão se torna cada vez mais premente à medida que os impactos das mudanças climáticas se tornam mais visíveis e abrangentes. A modelagem, que envolveu a utilização dos modelos ARIMA e de suavização exponencial, mostrou a eficácia de abordagens quantitativas na previsão de fenômenos ambientais. O modelo ARIMA destacou-se pela sua capacidade de fornecer previsões mais precisas, enquanto a suavização exponencial se mostrou uma alternativa válida em contextos onde a simplicidade é necessária.

Palavras-chave: Séries Temporais. Emissões de CO₂. Modelo ARIMA. Suavização Exponencial.

ABSTRACT

This study applies statistical and machine learning techniques to model time series of atmospheric CO₂ concentrations. The goal is to identify underlying patterns and trends, and to generate accurate forecasts for the future. As a key driver of climate change, understanding CO₂ levels is essential for shaping environmental policy and decision-making. The research compares the performance of ARIMA and exponential smoothing models in forecasting CO₂ concentrations. The results show that ARIMA models produced more accurate predictions, while exponential smoothing provided a simpler, yet effective alternative. These findings highlight the value of quantitative methods in analyzing environmental time series data and contribute to a deeper understanding of the complex dynamics of climate change.

Keywords: Time Series Analysis, CO₂ Emissions, ARIMA Modeling, Exponential Smoothing.

ÍNDICE DE ABREVIACÕES

ACF – *Autocorrelation Function* (Função de autocorrelação)

AR – *AutoRegressive* (Modelos autorregressivos)

ARIMA – *AutoRegressive Integrated Moving Average* (Modelo Autorregressivo Integrado de Médias Móveis)

ARMA – *AutoRegressive Moving Average* (Modelo Autorregressivo de Médias Móveis)

CO₂ - *Carbon Dioxide* (Gás Carbônico)

ETS – *Error, Trend, Seasonal* (Erro, Tendência, Sazonalidade)

GEE – *Gases do Efeito Estufa*

I - *Differencing* (Diferenciação)

LSTM – *Long Short-Term Memory* (Memória de Longo e Curto Prazo)

MA – *Moving Average* (Médias Móveis)

MAE – *Mean Absolute Error* (Erro Médio Absoluto)

MAPE – *Mean Absolute Percentage Error* (Erro Médio Percentual Absoluto)

ME – *Mean Error* (Erro Médio)

NOAA – *National Oceanic and Atmospheric Administration* (Administração Nacional Oceânica e Atmosférica)

PACF – *Partial Autocorrelation Function* (Função de Autocorrelação Parcial)

SUMÁRIO

1. INTRODUÇÃO.....	9
2. OBJETIVO GERAL.....	10
3. OBJETIVOS ESPECÍFICOS.....	10
4. REVISÃO BIBLIOGRÁFICA	11
4.1 Conceitos e fundamentos de séries temporais.	11
4.2 Modelos de séries temporais univariadas (Suavização exponencial e ARIMA).....	14
4.3 Aplicações de modelagem em séries temporais na área de ciências ambientais, com foco em gases do efeito estufa	16
5. METODOLOGIA	19
6. RESULTADOS E DISCUSSÃO.....	23
6.1 Suavização exponencial.....	23
6.2 Modelo ARIMA.....	26
7. CONCLUSÃO.....	33
8. REFERÊNCIAS	35
9. APÊNDICE.....	37

1. INTRODUÇÃO

O efeito estufa, que tem sido intensificado pela crescente concentração de gases como CO_2 na atmosfera, representa um dos maiores desafios ambientais da atualidade. Esses gases desempenham um papel fundamental na alteração do clima global, e sua acumulação tem consequências diretas para o meio ambiente e a saúde humana. A necessidade de enfrentar essa questão se torna cada vez mais premente à medida que os impactos das mudanças climáticas se tornam mais visíveis e abrangentes.

Para lidar com esse desafio, a compreensão das dinâmicas das concentrações de CO_2 ao longo do tempo é essencial. Essa análise temporal não apenas ajuda a prever cenários futuros, mas também é crucial para a implementação de políticas de mitigação eficazes. Identificar como esses gases se comportam em diferentes períodos permite que tomadores de decisão desenvolvam estratégias informadas para reduzir suas emissões e limitar os efeitos do aquecimento global.

Este trabalho propõe avaliar a modelagem em séries temporais das concentrações de CO_2 , utilizando modelos de previsão em séries temporais com o objetivo de identificar padrões e tendências significativas na variação dessas concentrações. A modelagem de séries temporais é uma ferramenta poderosa para entender a dinâmica subjacente dos dados e suas variações ao longo do tempo.

Para analisar a evolução da concentração de CO_2 ao longo do tempo, será empregada a suavização exponencial. Neste trabalho, será utilizada a forma mais geral e flexível de suavização exponencial, que considera diferentes componentes da série, como o nível, que representa o valor médio da série; a tendência, que se refere à direção e magnitude do crescimento ou declínio ao longo do tempo; e a sazonalidade, que captura padrões sazonais ou cíclicos na série. Também será empregado modelo autorregressivo ARIMA (AutoRegressivo Integrado de Médias Móveis). O ARIMA, incluindo sua versão sazonal (SARIMA), é um modelo adequado para séries temporais com ou sem componentes sazonais, permitindo modelar dependências temporais e identificar padrões de autocorrelação ao longo do tempo, ou seja, como os dados de um momento influenciam os dados do próximo.

Embora este trabalho não explore as redes neurais, é importante ressaltar que essas ferramentas demonstram grande potencial para modelar a complexidade das emissões de CO_2 . Sua capacidade de aprender padrões não lineares e de longo prazo as torna especialmente

adequadas para capturar as interações entre os diversos fatores que influenciam as emissões, como atividades industriais, uso de combustíveis fósseis e práticas agrícolas.

Além de identificar padrões nas concentrações de CO₂, a modelagem de séries temporais poderá avaliar o impacto de variáveis externas, como políticas ambientais, inovações tecnológicas e mudanças no comportamento social. Esses fatores podem influenciar diretamente as emissões de gases de efeito estufa. Compreender essas interações é essencial para prever como as dinâmicas de emissão podem evoluir e identificar medidas que possam ser adotadas para mitigar esses impactos. Essas abordagens representam propostas plausíveis para trabalhos futuros.

Em suma, a análise das concentrações de CO₂ por meio de modelagem em séries temporais não apenas contribui para o conhecimento científico sobre o efeito estufa, mas também oferece uma base sólida para a formulação de políticas públicas. Ao identificar padrões e tendências, este trabalho se propõe a ser uma ferramenta valiosa para pesquisadores e decisores, visando a construção de um futuro mais sustentável e a mitigação dos efeitos das mudanças climáticas.

2. OBJETIVO GERAL

Modelar as séries temporais das concentrações de CO₂ utilizando técnicas estatísticas e de aprendizado de máquina, com o objetivo de identificar padrões, tendências e realizar previsões futuras.

3. OBJETIVOS ESPECÍFICOS

- Coletar e organizar os dados históricos das concentrações de CO₂.
- Realizar uma análise exploratória dos dados, identificando características como estacionariedade, sazonalidade e tendências.
- Selecionar e aplicar modelos de séries temporais adequados para as concentrações de CO₂, considerando suas características específicas.
- Avaliar o desempenho dos modelos utilizando métricas de erro e comparar a eficácia dos diferentes modelos.
- Realizar previsões das concentrações de CO₂ para um horizonte temporal futuro.

4. REVISÃO BIBLIOGRÁFICA

4.1 Conceitos e fundamentos de séries temporais.

Séries temporais são sequências de dados coletadas ou registradas em intervalos de tempo específicos, com o objetivo de analisar e prever o comportamento futuro de variáveis ao longo do tempo. A principal característica que define uma série temporal é sua estrutura cronológica, que atribui uma ordem aos dados, permitindo identificar padrões como tendências, sazonalidades, ciclos e irregularidades. As séries temporais são amplamente utilizadas em áreas como economia, finanças, meteorologia, saúde e indústria, para observar fenômenos e realizar análises preditivas, com base na evolução passada dos dados (Bezerra, 2006).

Um dos conceitos fundamentais em séries temporais é a "tendência", que se refere ao movimento de longo prazo dos dados, indicando se a série está crescendo, decrescendo ou se mantendo estável ao longo do tempo. A tendência pode ser linear ou não-linear e é importante para identificar a direção geral da série. Outro conceito essencial é a "sazonalidade", que representa padrões que se repetem em intervalos regulares, como flutuações mensais nas vendas de produtos sazonais ou variações de temperatura ao longo das estações do ano. A sazonalidade é frequentemente analisada para compreender variações cíclicas que influenciam uma série temporal (Ehlers, 2007).

Além disso, as séries temporais podem exibir características de "ciclos", que são oscilações irregulares e de longo prazo em torno de uma tendência central. Diferente da sazonalidade, que se repete em intervalos previsíveis, os ciclos podem ter durações variadas e geralmente são associados a fenômenos econômicos ou sociais. Outro elemento relevante é a "aleatoriedade" ou "ruído", que representa variações imprevisíveis e sem padrão aparente, muitas vezes resultantes de eventos esporádicos ou choques externos. A análise de séries temporais visa distinguir esses componentes para uma melhor compreensão e modelagem dos dados (Morettin, 2006).

Um modelo aditivo para uma série temporal é dado pela Equação (1):

$$Z_t = T_t + S_t + C_t + \epsilon_t \quad (1)$$

onde Z_t representa o valor da série temporal, T_t denota a tendência, S_t refere-se à sazonalidade, C_t representa o ciclo, e ϵ_t é o erro aleatório para o período.

O modelo aditivo considera a hipótese de que esses quatro componentes são independentes entre si. Quando a variação de um componente tem impacto sobre os outros o modelo mais adequado é o multiplicativo. Nos modelos multiplicativos, os componentes são multiplicados juntos para formar o valor observado da série temporal, e a relação entre eles é de interdependência.

Os métodos de modelagem de séries temporais são variados, sendo alguns dos mais comuns os modelos de médias móveis (MA), modelos autorregressivos (AR) e modelos mistos autorregressivos de médias móveis (ARMA e ARIMA). Cada um desses modelos possui abordagens diferentes para lidar com a autocorrelação entre os dados, isto é, a relação entre valores passados e presentes da série. Por exemplo, os modelos AR se baseiam em relações entre valores passados da própria série, enquanto os modelos MA lidam com médias ponderadas de ruídos passados. Já o ARIMA combina ambas as abordagens, integrando também a diferenciação dos dados para remover tendências. A capacidade de identificar e separar componentes de uma série temporal permite que profissionais de diversas áreas tomem decisões informadas e baseadas em dados históricos.

Com o avanço das técnicas de aprendizado de máquina, novos métodos de análise de séries temporais, como redes neurais recorrentes (RNN) e modelos de alisamento exponencial, têm se tornado ferramentas valiosas para análises mais precisas, levando em conta padrões complexos e não-lineares presentes em dados temporais (Morettin, 2006).

Essas séries não são apenas uma sequência de valores, mas dados estruturados de acordo com uma dimensão temporal, permitindo identificar e analisar padrões subjacentes, ou seja, padrões que estão presentes de forma implícita, mesmo que não sejam imediatamente visíveis sem uma análise cuidadosa. A análise de séries temporais é fundamental para compreender tendências históricas e fazer previsões futuras. Um dos objetivos primários dessa análise é extrair informações relevantes dos dados e modelá-los para que possamos prever seu comportamento com base em padrões identificados (Morettin, 2006).

Os componentes de uma série temporal, como a tendência, a sazonalidade e o ciclo, são a base para sua análise. A tendência refere-se ao comportamento de longo prazo, enquanto a sazonalidade indica variações repetitivas que ocorrem em intervalos regulares. Por outro lado, os ciclos não têm periodicidade fixa, mas estão relacionados a mudanças na atividade econômica, climática ou social, entre outros fatores. Além disso, existe um componente de ruído, representando as flutuações aleatórias que podem interferir na identificação dos padrões. Separar esses componentes é um passo fundamental para obter informações relevantes (Morettin, 2006).

Entre os métodos de modelagem de séries temporais, os modelos de suavização exponencial (ou alisamento exponencial) são populares devido à sua simplicidade e eficácia na captura de padrões de curto prazo. Eles trabalham atribuindo pesos diferentes para os dados históricos, de forma a dar mais importância aos valores mais recentes. Essa abordagem é útil quando as séries temporais apresentam um comportamento que muda rapidamente e é essencial para capturar essas variações de forma responsiva. Diferentes técnicas de alisamento, como o modelo de Holt-Winters, permitem lidar com séries que apresentam tanto tendência quanto sazonalidade (Bezerra, 2006).

Além dos métodos tradicionais, técnicas avançadas de aprendizado de máquina, como as Redes Neurais Recorrentes (RNNs) e os modelos Long Short-Term Memory (LSTM), têm se destacado na modelagem de séries temporais. Esses modelos são especialmente úteis para lidar com dados que apresentam padrões complexos e não lineares, como interações entre diferentes componentes temporais. As RNNs são projetadas para processar sequências de dados e reconhecer dependências temporais de longo alcance, sendo capazes de capturar características sutis que podem ser desafiadoras para métodos lineares (Bezerra, 2006).

Em termos de aplicação prática, a análise de séries temporais é amplamente usada em finanças para prever preços de ações, em meteorologia para estimar padrões climáticos, e na saúde para acompanhar a progressão de doenças. Essas aplicações se beneficiam das previsões baseadas em dados históricos, fornecendo uma base para decisões estratégicas e operacionais. Devido à sua relevância, o estudo de séries temporais continua a evoluir, integrando novas técnicas e refinando métodos existentes para lidar com a crescente complexidade dos dados temporais no contexto moderno (Rodrigues, 2015).

Uma das primeiras etapas para isso é a decomposição da série em seus componentes principais, como tendência, sazonalidade, ciclo e ruído. Essa decomposição facilita a identificação dos padrões recorrentes e o isolamento das flutuações aleatórias. A partir dessa separação, é possível aplicar modelos matemáticos específicos a cada componente, obtendo previsões mais precisas e segmentadas. Essa abordagem é essencial em situações onde cada um desses componentes afeta a variável de interesse de maneira distinta (Rodrigues, 2015).

Outro conceito relevante em séries temporais é a "autocorrelação", que mede o grau de dependência entre valores anteriores e presentes da série. Isso é fundamental porque muitas séries temporais apresentam um comportamento autocorrelacionado, ou seja, os valores em determinado ponto do tempo estão diretamente relacionados aos valores anteriores. Essa característica é a base para a aplicação de modelos autorregressivos (AR), que utilizam os

valores passados para prever futuros. A autocorrelação também é útil para identificar padrões sazonais e cíclicos, especialmente quando se utiliza a função de autocorrelação parcial (PACF) para examinar dependências mais específicas.

Além disso, um desafio importante na modelagem de séries temporais é lidar com dados não estacionários. Uma série temporal estacionária é aquela cujas propriedades estatísticas, como a média e a variância, permanecem constantes ao longo do tempo. No entanto, muitas séries temporais reais apresentam características não estacionárias, como tendência ou sazonalidade, o que exige métodos de diferenciação para transformar a série em estacionária antes da aplicação de certos modelos, como o ARIMA. A identificação da estacionariedade é crucial para a escolha do modelo adequado, sendo um pré-requisito para a aplicação de várias técnicas de análise. No campo da saúde, ela é aplicada para monitorar o surto de doenças e prever sua progressão. No setor de energia, ajuda na previsão de demanda elétrica e na gestão de recursos. Cada aplicação tem suas especificidades, mas o objetivo comum é fornecer um conhecimento mais profundo dos padrões históricos, ajudando na tomada de decisões fundamentadas e estratégicas (Rodrigues, 2015).

4.2 Modelos de séries temporais univariadas (Suavização exponencial e ARIMA)

Entre os principais modelos estão a suavização exponencial e os modelos ARIMA, que são amplamente utilizados devido à sua eficácia na captura de padrões temporais e sua flexibilidade em diferentes contextos. Esses modelos ajudam a entender as variações temporais em dados que apresentam tendências, sazonalidades ou apenas ruído aleatório, fornecendo uma base sólida para a tomada de decisões estratégicas (Camilo, 2021).

Um dos métodos mais simples dessa categoria é a Suavização Exponencial Simples, que é adequada para séries temporais sem tendências ou sazonalidades. Trata-se de uma técnica de previsão que calcula uma média ponderada dos dados históricos, atribuindo maior peso aos valores mais recentes. Essa técnica é usada para suavizar flutuações irregulares nos dados e fornecer uma estimativa do nível da série temporal. O método atribui pesos exponenciais decrescentes aos valores passados, com o objetivo de capturar mudanças recentes na série. Quando há presença de tendência ou sazonalidade, métodos mais complexos, como a Suavização de Holt e a Suavização de Holt-Winters, são aplicados. Enquanto o método de Holt considera tanto o nível quanto a tendência, o de Holt-Winters é capaz de incorporar a sazonalidade, sendo ideal para séries com padrões cíclicos. (Morettin, 2006; Camilo, 2021).

Os modelos ARIMA (AutoRegressive Integrated Moving Average) combinam três componentes principais: autorregressão (AR), médias móveis (MA) e diferenciação (I). A parte autorregressiva utiliza os valores passados da própria série para fazer previsões, enquanto a parte de médias móveis utiliza erros passados (resíduos) para ajustar as estimativas. A diferenciação é empregada para transformar uma série não estacionária em estacionária, removendo tendências e padrões sazonais. A grande vantagem dos modelos ARIMA é sua capacidade de se adaptar a uma ampla gama de séries temporais, ajustando-se de acordo com suas características (Morettin, 2006; Camilo, 2021).

Ao escolher entre a suavização exponencial e os modelos ARIMA, é fundamental entender a natureza dos dados e o objetivo da análise. A suavização exponencial tende a ser mais adequada para séries que apresentam variações de curto prazo e quando se deseja dar maior relevância a observações recentes. Em contrapartida, os modelos ARIMA são mais robustos e permitem uma análise mais detalhada da estrutura da série, especialmente quando há correlações temporais mais complexas ou a presença de não estacionaridade. A escolha do modelo apropriado depende, portanto, do tipo de série temporal e da complexidade dos padrões observados (Camilo, 2021).

Eles oferecem uma abordagem sistemática para entender e prever o comportamento futuro de variáveis, auxiliando na elaboração de estratégias e na gestão de riscos. A habilidade de distinguir entre diferentes padrões temporais e escolher o modelo adequado é essencial para profissionais que lidam com grandes volumes de dados temporais, garantindo previsões mais precisas e decisões melhor fundamentadas.

Um dos modelos mais intuitivos é a suavização exponencial, que ajusta previsões de acordo com a importância dos dados mais recentes. A lógica por trás dessa abordagem é simples: atribuir maior peso a eventos recentes reflete mudanças recentes no comportamento da série, o que é ideal para cenários onde o contexto atual é mais relevante do que o histórico distante. Esse método é aplicado com diferentes complexidades, dependendo se há presença de tendência e sazonalidade (Hyndman, 2008).

Dentro dos modelos de suavização exponencial, a técnica de Holt destaca-se por sua capacidade de lidar com séries que apresentam tendência, realizando previsões que consideram variações lineares ao longo do tempo. O método de Holt-Winters, por outro lado, é uma extensão desse conceito e inclui a capacidade de modelar sazonalidade, além de tendência. Ele se divide em duas abordagens principais: sazonalidade aditiva e multiplicativa. A escolha entre essas duas formas depende de como a sazonalidade afeta a série; se o impacto sazonal é

proporcional ao nível da série, utiliza-se o modelo multiplicativo, caso contrário, aditivo (Hyndman, 2008).

Os modelos ARIMA seguem uma abordagem diferente, concentrando-se na captura de autocorrelações internas e na remoção de características não estacionárias por meio da diferenciação. Um ponto-chave na utilização do ARIMA é a identificação da ordem adequada dos componentes AR, I e MA, o que é geralmente realizado com base na análise de gráficos de autocorrelação e autocorrelação parcial. A aplicação de diferenciação ajuda a remover tendências lineares ou sazonais, permitindo que a série se torne estacionária e viável para análise preditiva. Essa característica confere aos modelos ARIMA uma flexibilidade significativa em relação às séries temporais complexas (Hyndman, 2008).

A modelagem ARIMA é particularmente útil quando as séries temporais exibem comportamentos mais sofisticados, como padrões sazonais irregulares ou correlações temporais complexas. Além disso, variações do ARIMA, como o SARIMA, são desenvolvidas para lidar com sazonalidade explícita, estendendo o modelo para incluir componentes sazonais adicionais. Isso permite uma abordagem mais estruturada para séries que apresentam variações cíclicas específicas em intervalos regulares, como vendas mensais de produtos sazonais ou flutuações climáticas anuais. A escolha de um modelo ARIMA ou sua variante depende da presença de padrões regulares ao longo do tempo. A suavização exponencial é ideal para responder rapidamente a mudanças recentes nos dados, enquanto o ARIMA oferece um tratamento mais rigoroso de dependências e padrões históricos. A aplicação desses modelos é difundida em setores como economia, finanças, controle de qualidade e previsão climática, pois oferecem ferramentas poderosas para antecipar tendências e tomar decisões estratégicas com base em dados históricos (Hyndman, 2008).

4.3 Aplicações de modelagem em séries temporais na área de ciências ambientais, com foco em gases do efeito estufa

Uma das aplicações mais notáveis é a previsão de mudanças climáticas, onde técnicas como ARIMA e modelos de suavização são usadas para projetar variações de temperatura, precipitação e outros parâmetros climáticos ao longo de décadas. Esses modelos ajudam a identificar tendências de aquecimento global, flutuações sazonais e anomalias climáticas, fornecendo uma base sólida para o desenvolvimento de políticas ambientais e estratégias de mitigação de impactos (Albuquerque, 2015).

As séries temporais são utilizadas para identificar padrões históricos que indicam a probabilidade de ocorrência de tais eventos. Modelos SARIMA (ARIMA Sazonal) são particularmente úteis para prever fenômenos que têm uma periodicidade clara, como enchentes sazonais em determinadas bacias hidrográficas. A capacidade de prever eventos extremos permite a tomada de decisões antecipadas para reduzir danos e implementar medidas preventivas em áreas vulneráveis, além de ajudar na gestão eficiente de recursos hídricos (Albuquerque, 2015).

No monitoramento da qualidade do ar, os modelos de séries temporais também são amplamente aplicados. Por meio da análise de dados de poluentes como dióxido de enxofre (SO_2), material particulado (MP_{10}) e dióxido de nitrogênio (NO_2), é possível identificar padrões de poluição em regiões urbanas e industriais. A modelagem permite prever variações nos níveis de poluentes, auxiliando na criação de políticas de controle da qualidade do ar e na implementação de estratégias de mitigação. Além disso, as previsões baseadas em séries temporais são úteis para alertar a população sobre períodos críticos de poluição, protegendo a saúde pública (Albuquerque, 2015).

Os modelos de séries temporais também são usados para prever mudanças nos níveis dos oceanos e no comportamento de correntes marítimas. Ao analisar dados históricos, é possível identificar tendências de aumento do nível do mar e projetar futuros cenários de elevação, o que é crucial para o planejamento de zonas costeiras. Modelos de suavização exponencial, por exemplo, ajudam a suavizar flutuações de curto prazo, fornecendo uma visão mais clara das tendências de longo prazo. Essas análises são essenciais para avaliar o impacto do derretimento das geleiras, aquecimento das águas oceânicas e mudanças nos ecossistemas marinhos (Albuquerque, 2015).

A previsão de vazões em rios, disponibilidade de água em reservatórios e padrões de consumo é feita por meio de modelagens que consideram padrões sazonais e variações cíclicas. Com isso, é possível otimizar a alocação de água, prever períodos de escassez e planejar o uso eficiente dos recursos hídricos em diversas atividades, como irrigação, abastecimento urbano e geração de energia hidrelétrica. Essas aplicações mostram como a modelagem de séries temporais é uma ferramenta valiosa para enfrentar os desafios ambientais de maneira proativa e baseada em dados (Albuquerque, 2015).

O efeito estufa é um fenômeno natural essencial para manter a Terra em uma temperatura adequada para a vida. Ele ocorre devido à presença de gases na atmosfera, como dióxido de carbono (CO_2), metano (CH_4) e óxidos de nitrogênio, que absorvem parte da radiação infravermelha emitida pela superfície terrestre, retendo o calor. Esses gases formam

uma espécie de "cobertor" que impede que o calor se dissipe totalmente para o espaço, mantendo o planeta aquecido. Sem o efeito estufa, a temperatura média da Terra seria muito mais baixa, o que tornaria a vida inviável em várias regiões do globo (Guimarães, 2015).

No entanto, as atividades humanas têm aumentado significativamente a concentração desses gases, intensificando o efeito estufa e levando ao aquecimento global. A queima de combustíveis fósseis, a agricultura intensiva, o desmatamento e outras práticas liberam grandes quantidades de CO₂, CH₄ e outros gases de efeito estufa (GEE) na atmosfera. Esse acúmulo de gases retém mais calor do que o necessário, elevando as temperaturas médias globais e provocando mudanças climáticas. Esse aumento é uma preocupação central, pois o aquecimento global impacta diretamente os sistemas naturais e humanos (Guimarães, 2015).

O aquecimento global, impulsionado pelo efeito estufa intensificado, traz uma série de consequências, como o derretimento das calotas polares, a elevação do nível do mar e eventos climáticos extremos, como secas prolongadas e tempestades intensas. O derretimento das geleiras afeta diretamente as regiões costeiras, ameaçando comunidades e ecossistemas que dependem de um equilíbrio climático. Além disso, as mudanças nas temperaturas e padrões de precipitação afetam a produção agrícola, a disponibilidade de água potável e a biodiversidade, colocando em risco a segurança alimentar e hídrica de muitas regiões do mundo (Pelegri, 2018).

A redução dos gases de efeito estufa na atmosfera é fundamental para mitigar o impacto do aquecimento global. Para isso, diversas políticas e ações têm sido adotadas globalmente, como o Acordo de Paris, que visa limitar o aumento da temperatura média global em até 1,5°C em relação aos níveis pré-industriais. Entre as medidas estão a transição para fontes de energia renováveis, o reflorestamento, a promoção da eficiência energética e o incentivo à adoção de tecnologias limpas. Essas iniciativas visam diminuir as emissões de GEE e promover o desenvolvimento sustentável, preservando os recursos naturais e a qualidade de vida das futuras gerações. A conscientização sobre a necessidade de reduzir a emissão de gases poluentes é crucial para incentivar mudanças de comportamento, como o consumo consciente e o apoio a práticas sustentáveis. Além disso, o investimento em pesquisas para o desenvolvimento de novas tecnologias e métodos de produção mais limpos pode impulsionar a transição para uma economia de baixo carbono. Somente com uma abordagem abrangente será possível frear o aquecimento global e preservar o equilíbrio climático do planeta (Pelegri, 2018).

Um dos principais efeitos do aumento dos GEE é a elevação do nível do mar. Com o aquecimento global, as calotas polares e geleiras estão derretendo em ritmo acelerado, o que adiciona grandes volumes de água aos oceanos. Além disso, o aquecimento das águas provoca

sua expansão térmica, contribuindo ainda mais para o aumento do nível do mar. Regiões costeiras e ilhas baixas são particularmente vulneráveis, enfrentando riscos como a erosão de suas áreas, a salinização de aquíferos e a destruição de habitats. Esses impactos também ameaçam milhões de pessoas que vivem em áreas costeiras, obrigando deslocamentos forçados e afetando a economia local (Albuquerque, 2010).

A intensificação dos eventos climáticos extremos também é um efeito alarmante dos GEE. Tempestades mais frequentes e severas, ondas de calor prolongadas, secas intensas e chuvas torrenciais são consequências diretas do aquecimento global. Esses fenômenos representam riscos à agricultura, prejudicando a produção de alimentos e ameaçando a segurança alimentar em várias regiões. As populações mais vulneráveis são as mais afetadas por esses eventos, pois enfrentam maiores dificuldades para se adaptar e se recuperar dos impactos, agravando desigualdades sociais e econômicas (Albuquerque, 2010).

O excesso de CO₂ se dissolve na água, formando ácido carbônico, o que reduz o pH dos oceanos. Esse processo afeta diretamente a vida marinha, principalmente os organismos que dependem de conchas calcárias, como corais e moluscos. Com a diminuição da biodiversidade e a degradação dos recifes de corais, ecossistemas inteiros são colocados em risco, prejudicando atividades econômicas relacionadas, como a pesca e o turismo.

A exposição a ondas de calor intensas e prolongadas pode causar problemas de saúde, como desidratação, exaustão por calor e agravamento de doenças cardiovasculares e respiratórias. Além disso, a qualidade do ar pode ser prejudicada pelo aumento de poluentes atmosféricos, exacerbando condições como asma e alergias. Esses impactos demandam uma maior atenção das autoridades de saúde pública e políticas de adaptação e mitigação para proteger as populações mais expostas (Albuquerque, 2010).

5. METODOLOGIA

A coleta de dados foi realizada a partir de bases de dados públicas, como o National Oceanic and Atmospheric Administration (NOAA, 2024). Essas fontes forneceram informações relevantes e necessárias para a pesquisa. O acesso a dados confiáveis foi fundamental para garantir a qualidade das análises subsequentes.

Após a coleta, os dados passaram por um processo de pré-processamento, onde foram identificados e tratados *outliers* e valores ausentes. Transformações adequadas foram realizadas quando necessário, visando garantir que os dados estivessem em um formato apropriado para a

análise. Esse passo foi fundamental para melhorar a integridade dos dados e evitar distorções nos resultados.

Em seguida, foi realizada uma análise exploratória dos dados, que permitiu identificar padrões, tendências, sazonalidade e autocorrelação. Durante essa fase, também foram definidos os conjuntos de treino e teste, preparando o caminho para a modelagem. Essa análise inicial foi essencial para compreender melhor a natureza dos dados antes de aplicar técnicas mais complexas.

A validação de modelos preditivos é importante para avaliar seu desempenho e a capacidade de generalização em relação a valores desconhecidos. Neste trabalho, adotou-se uma abordagem adaptada para séries temporais, garantindo que a ordem temporal dos dados fosse respeitada (HYNDMAN; ATHANASOPOULOS, 2021; JAMES et al., 2013; BERGMEIR; HYNDMAN; KOO, 2018). O procedimento consistiu em dividir o conjunto de dados em $k=10$ partes iguais, reservando um dos subconjuntos como conjunto de teste, enquanto os $k-1$ subconjuntos restantes foram utilizados para treinar o modelo. Em geral, $k=5$ e $k=10$ são os valores mais utilizados, mas o valor ideal de k pode depender do tamanho do conjunto de dados e do problema específico. Para conjuntos de dados menores, um valor maior de k pode ser mais apropriado.

A série temporal dos níveis atmosféricos mensais de dióxido de carbono (CO_2) foi considerada desde janeiro de 1959 até setembro de 2024. Os dados temporais entre janeiro de 1959 e dezembro de 2017 foram designados como conjunto de treinamento, e os dados restantes, de janeiro de 2018 a setembro de 2024, como conjunto de teste.

Na etapa de modelagem, foram ajustados dois modelos de séries temporais univariadas: a suavização exponencial e o ARIMA. Essa escolha de modelos teve como objetivo capturar as características específicas das séries temporais, permitindo uma previsão mais precisa.

As principais componentes na suavização exponencial compreendem o erro (E), que representa a discrepância entre os valores observados e previstos, a tendência (T), que indica a direção geral dos dados, e a sazonalidade (S), responsável por identificar padrões regulares (MORETTIN; TOLOI, 2006). Foi empregada a função ETS do pacote *forecast* no R, utilizando a sintaxe ETS(serie, model). O primeiro argumento dessa função é a série temporal, enquanto o segundo consiste em uma sequência de caracteres que especificam os tipos de método para o erro, a tendência e a sazonalidade (HYNDMAN; KHANDAKAR, 2008).

No modelo de suavização exponencial estão envolvidos parâmetros tais como o fator de suavização de nível α , o fator de suavização para a tendência β e o fator de suavização para

a sazonalidade γ . Um valor para α próximo de 1 indica que os valores passados têm pouco peso na previsão atual, enquanto um valor próximo de 0 indica que a previsão é baseada principalmente nos valores passados. Um valor de β próximo de 1 indica uma tendência que muda rapidamente para se ajustar aos dados mais recentes, enquanto um valor próximo de 0 indica que a tendência muda lentamente ao longo do tempo. Um valor de γ próximo de 1 indica que a sazonalidade é ajustada rapidamente para se adaptar a padrões sazonais recentes, enquanto um valor próximo de 0 indica que a sazonalidade muda lentamente ao longo do tempo. (HYNDMAN;ATHANASOPOULOS, 2014).

Para o caso do modelo ARIMA utilizou-se a função *auto.arima()* do R, também disponível no pacote *forecast*, para automatizar o processo de seleção do modelo mais apropriado para a série temporal em questão. Essa função fornece um modelo na notação ARIMA(p,d,q)(P,D,Q), em que, para os dados não sazonais, “p” representa o número de termos autoregressivos (AR), “d” indica o número de diferenciações necessárias para tornar a série estacionária, e “q” denota o número de termos de média móvel (MA). As letras “P”, “D” e “Q” têm significados semelhantes, mas aplicados aos dados sazonais. Os coeficientes do modelo dos termos de média móvel (MA) não sazonal e dos termos de média móvel sazonal (SMA) representam a contribuição das observações passadas dos termos de erro para prever o valor atual da série temporal. Esses coeficientes são importantes para entender como o modelo ARIMA está usando as informações passadas para fazer previsões sobre a série temporal (CRYER;CHAN, 2008; HYNDMAN; KHANDAKAR,2008; MORETTIN;TOLOI,2006).

Por fim, a eficiência dos modelos foi avaliada utilizando métricas de acurácia e análise residual. Essa avaliação foi importante para determinar a eficácia das previsões geradas pelos modelos ajustados. Para isso, foram utilizadas métricas de erro, que permitem comparar os valores previstos pelo modelo com os valores reais. Neste texto, foram consideradas quatro métricas de erro comumente utilizadas: Erro Médio Absoluto (MAE), Erro Médio (ME), Raiz Quadrática Média do Erro (RMSE) e Erro Médio Percentual Absoluto (MAPE).

Os erros de ajuste são definidos por

$$e_t = Z_t - \hat{Z}_t \quad (2)$$

onde Z_t o valor observado na série temporal e $\hat{Z}_t$ o valor estimado pelo modelo.

O MAE mede o desvio médio absoluto das previsões em relação aos valores, sendo definido por:

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (3)$$

O MAE é uma boa opção quando se deseja uma medida simples e intuitiva da magnitude média do erro.

O ME calcula a média aritmética dos erros, ou seja,

$$ME = \frac{1}{n} \sum_{i=1}^n e_i \quad (4)$$

O ME é útil para verificar se o modelo apresenta viés. Um valor positivo indica que, em média, o modelo superestima os valores, enquanto um valor negativo indica que subestima.

O RMSE calcula a raiz quadrada da média dos quadrados dos erros e é definida por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (5)$$

O RMSE tem a mesma unidade de medida dos valores originais, o que facilita a interpretação. Valores menores de RMSE indicam um melhor ajuste do modelo.

O MAPE calcula a média dos erros percentuais absolutos e é definida por:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{Z_i} \times 100\% \quad (6)$$

O MAPE é expresso como uma porcentagem e indica a magnitude média do erro em relação aos valores reais.

Gráficos de dispersão, histogramas, autocorrelação e quantil-quantil (Q-Q) são ferramentas essenciais para avaliar a qualidade do ajuste dos modelos os dados observados, especialmente no que se refere aos resíduos.

O gráfico de autocorrelação exibe a correlação entre os resíduos em diferentes defasagens (lags). Ele é útil para verificar se há padrões não capturados pelo modelo, como tendências ou sazonalidades não modeladas. Se os resíduos apresentam autocorrelação significativa em defasagens baixas, isso sugere que o modelo não conseguiu capturar completamente a dependência temporal, e ajustes adicionais podem ser necessários.

O gráfico Q-Q compara a distribuição dos resíduos com uma distribuição teórica, geralmente a normal. Para um bom ajuste, os resíduos devem seguir uma distribuição normal. Se os pontos do gráfico Q-Q estiverem próximos da linha reta, isso indica que os resíduos são

normalmente distribuídos, o que é uma suposição importante para a validade do modelo de suavização exponencial. No entanto, se houver desvios significativos, como "caudas" mais longas ou curvaturas, isso sugere que o modelo pode não estar capturando corretamente todas as características dos dados.

Em conjunto, a análise desses gráficos ajuda a diagnosticar a adequação do modelo, indicando se os resíduos são aleatórios e se o modelo de suavização exponencial oferece um bom ajuste aos dados de CO₂, sem deixar padrões significativos não explicados (Morettin, 2006; Camilo, 2021; Bezerra, 2006).

6. RESULTADOS E DISCUSSÃO

Os modelos de previsão, suavização exponencial e ARIMA, apresentados na metodologia, foram aplicados ao conjunto de dados dos níveis de CO₂. A avaliação desses modelos foi feita em termos das métricas de acurácia: ME, MAE, RMSE e MAPE, também definidas na metodologia. Além da acurácia, foi realizada uma análise dos resíduos quanto à sua dispersão, normalidade e autocorrelação.

6.1 Suavização exponencial

O modelo de suavização exponencial que melhor se ajustou foi o que engloba Erro, Tendência e Sazonalidade, denotado por ETS(M,A,A), que possui um tipo de erro multiplicativo (M), com tendência e sazonalidade aditivas (A). A Figura 1 mostra a série temporal dos níveis de CO₂ (linha preta), o ajuste do modelo de suavização exponencial à série de treino (linha vermelha) e à série de teste (linha azul), além da região de 95% confiança para a previsão (região azul transparente).

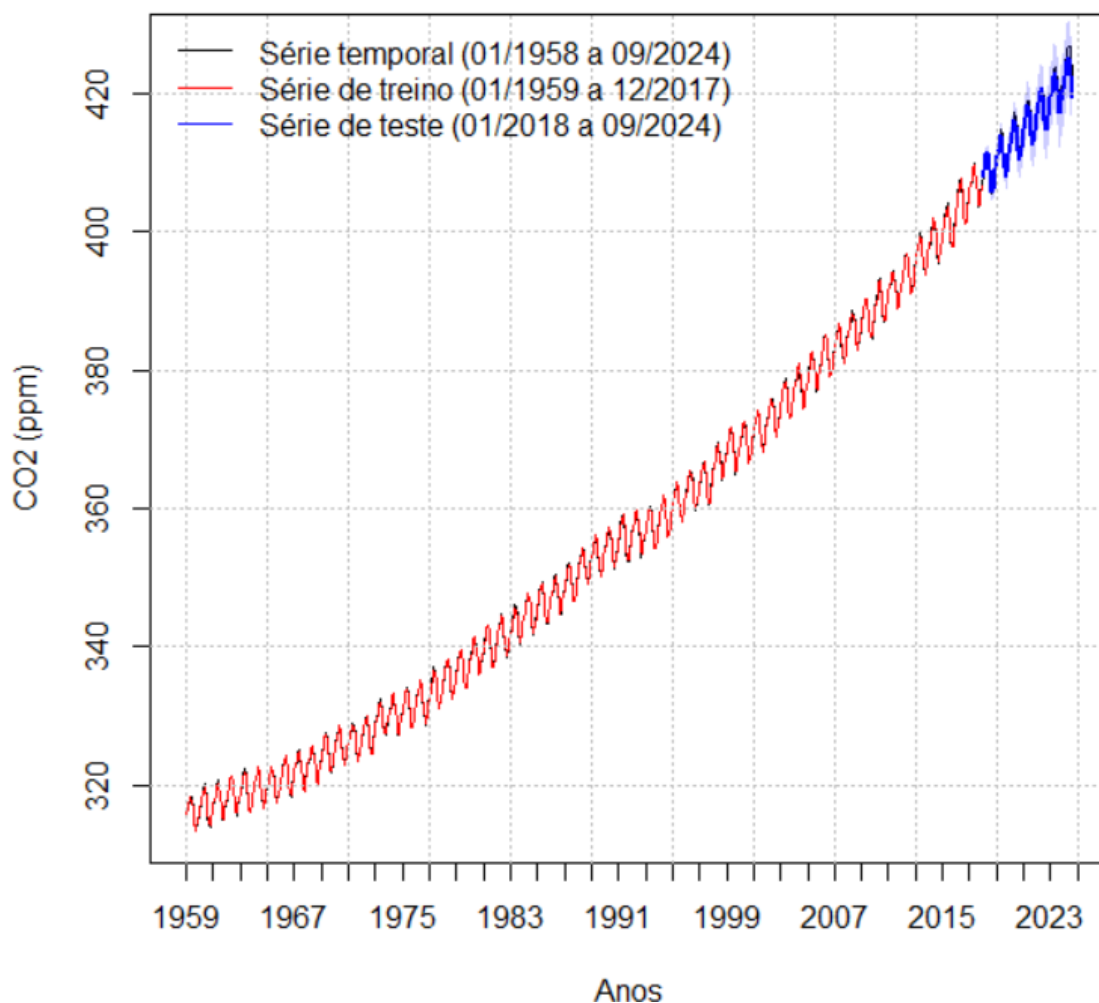
CO₂ médio global

FIGURA 1: SÉRIE TEMPORAL DOS NÍVEIS DE CO₂ (LINHA PRETA), AJUSTE DO MODELO DE SUAVIZAÇÃO EXPONENCIAL À SÉRIE DE TREINO (LINHA VERMELHA) E À SÉRIE DE TESTE (LINHA AZUL), ALÉM DA REGIÃO DE 95% CONFIANÇA PARA A PREVISÃO (REGIÃO AZUL TRANSPARENTE). (FONTE: AUTORIA PRÓPRIA).

O parâmetro de suavização obtido para o nível da série, $\alpha = 0,5723$, indica que cerca de 57% da nova previsão é baseada nos dados mais recentes, enquanto os 43% restantes são derivados das estimativas anteriores. Um valor de alpha mais alto indica que o modelo atribui maior peso às observações recentes. O parâmetro para a tendência, $\beta = 0,0075$, é um valor relativamente baixo. Isso sugere que a tendência não está mudando rapidamente e que o modelo considera a tendência de forma sutil, podendo indicar uma leve tendência de crescimento. Por fim, o parâmetro para a sazonalidade, $\gamma = 0,1045$, indica que o modelo atribui algum peso às flutuações sazonais, mas não de maneira intensa. Esse valor sugere que a sazonalidade tem um impacto moderado nas previsões.

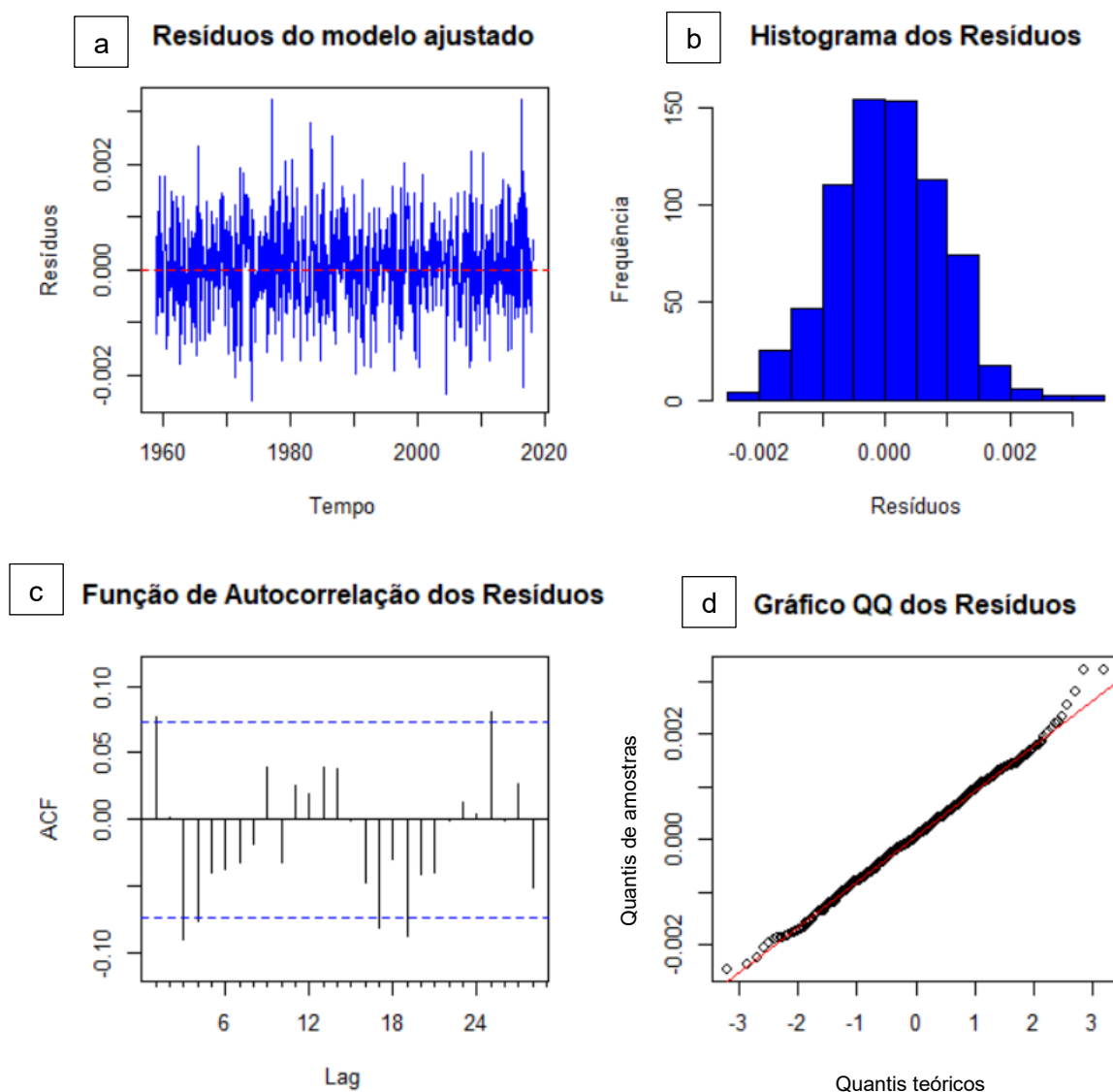


FIGURA 2: AJUSTE DO MODELO DE SUAVIZAÇÃO EXPONENCIAL AOS DADOS DE CO₂: (A) DISPERSÃO, (B) HISTOGRAMA, E (C) GRÁFICOS DE AUTOCORRELAÇÃO E (D) QUANTIL-QUANTIL DOS RESÍDUOS (FONTE: AUTORIA PRÓPRIA).

A Figura 2 ilustra o gráfico de dispersão, o histograma, e os gráficos de autocorrelação e quantil-quantil dos resíduos relativos ao ajuste do modelo de suavização exponencial aos dados de CO₂.

Observa-se na Figura 2(a) que os resíduos estão distribuídos aleatoriamente em torno de zero, e o histograma, Figura 2(b), exibe características de uma distribuição normal, o que também é evidenciado pela linearidade do gráfico quantil-quantil, Figura 4(d). No gráfico da ACF (Função de Autocorrelação), Figura 2(c), observa-se que em poucas defasagens (*lags*)

os coeficientes de autocorrelação excedem as linhas pontilhadas (faixa de 95% de confiança), indicando que a autocorrelação dos dados é limitada, ou seja, a maioria das correlações entre os valores da série temporal em diferentes momentos no tempo não é estatisticamente significativa. Essa falta de autocorrelação sugere que os resíduos são independentes, o que é desejável em modelos de séries temporais, pois indica que o modelo é apropriado e que as previsões feitas não apresentam padrões não capturados.

6.2 Modelo ARIMA

Com o objetivo de selecionar o melhor modelo de ajuste para os dados temporais de CO₂, além da suavização exponencial, também foi testado o modelo ARIMA (Média Móvel Integrada AutoRegressiva), incluindo a possibilidade de uma componente sazonal.

O modelo ARIMA que melhor se ajustou foi o ARIMA(0,1,3)(0,1,1)[12]. Este modelo não possui termos autoregressivos (AR) na parte não sazonal, sendo que a série foi diferenciada uma vez para torná-la estacionária e incluindo também três termos de média móvel (MA). Na parte sazonal, a série não apresenta termos autoregressivos; foi diferenciada uma única vez e possui um termo de média móvel. Além disso a série apresenta sazonalidade anual, com dados mensais que se repetem a cada 12 meses. Devido à presença de sazonalidade este modelo é denominado de SARIMA.

A Figura 3 mostra a série temporal dos níveis de CO₂ (linha preta), o ajuste do modelo ARIMA à série de treino (linha vermelha) e à série de teste (linha azul), além da região de 95% confiança para a previsão (região azul transparente).

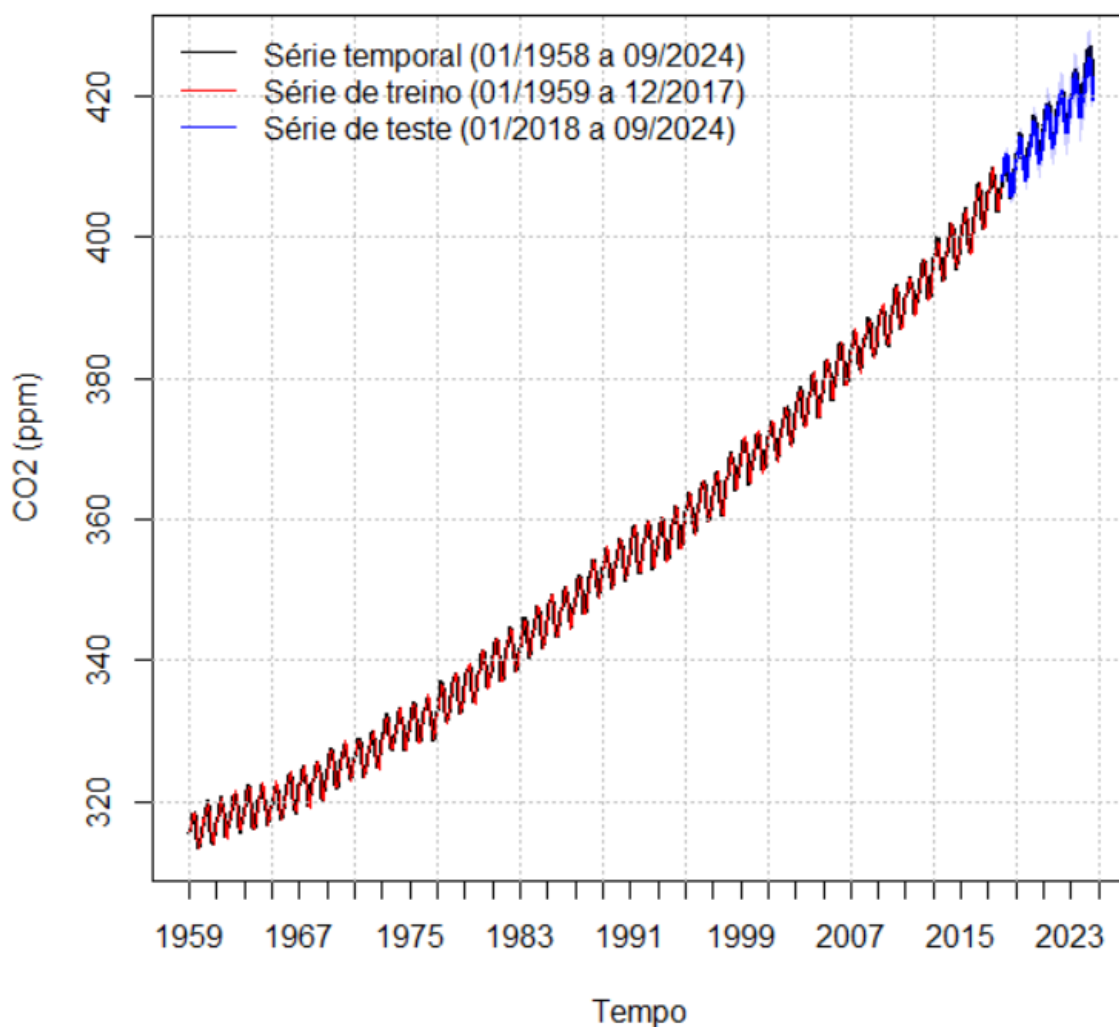
CO₂ médio global

FIGURA 3: SÉRIE TEMPORAL DOS NÍVEIS DE CO₂ (LINHA PRETA), O AJUSTE DO MODELO ARIMA À SÉRIE DE TREINO (LINHA VERMELHA) E À SÉRIE DE TESTE (LINHA AZUL), ALÉM DA REGIÃO DE 95% CONFIANÇA PARA A PREVISÃO (REGIÃO AZUL TRANSPARENTE) (FONTE: AUTORIA PRÓPRIA).

Foi obtido um coeficiente de média móvel não sazonal $ma_1 = -0,33503$ e um coeficiente de média móvel sazonal $sma_1 = -0,8636$. Esses coeficientes indicam que há uma dependência negativa dos erros passados, tanto no contexto do modelo ARIMA não sazonal quanto na parte sazonal, o que sugere que os erros passados tendem a ser ajustados de forma oposta ao seu sinal, com maior intensidade no componente sazonal.

A Figura 4 ilustra o gráfico de dispersão, o histograma, e os gráficos de autocorrelação e quantil-quantil dos resíduos relativos ao ajuste do modelo ARIMA aos dados de CO₂.

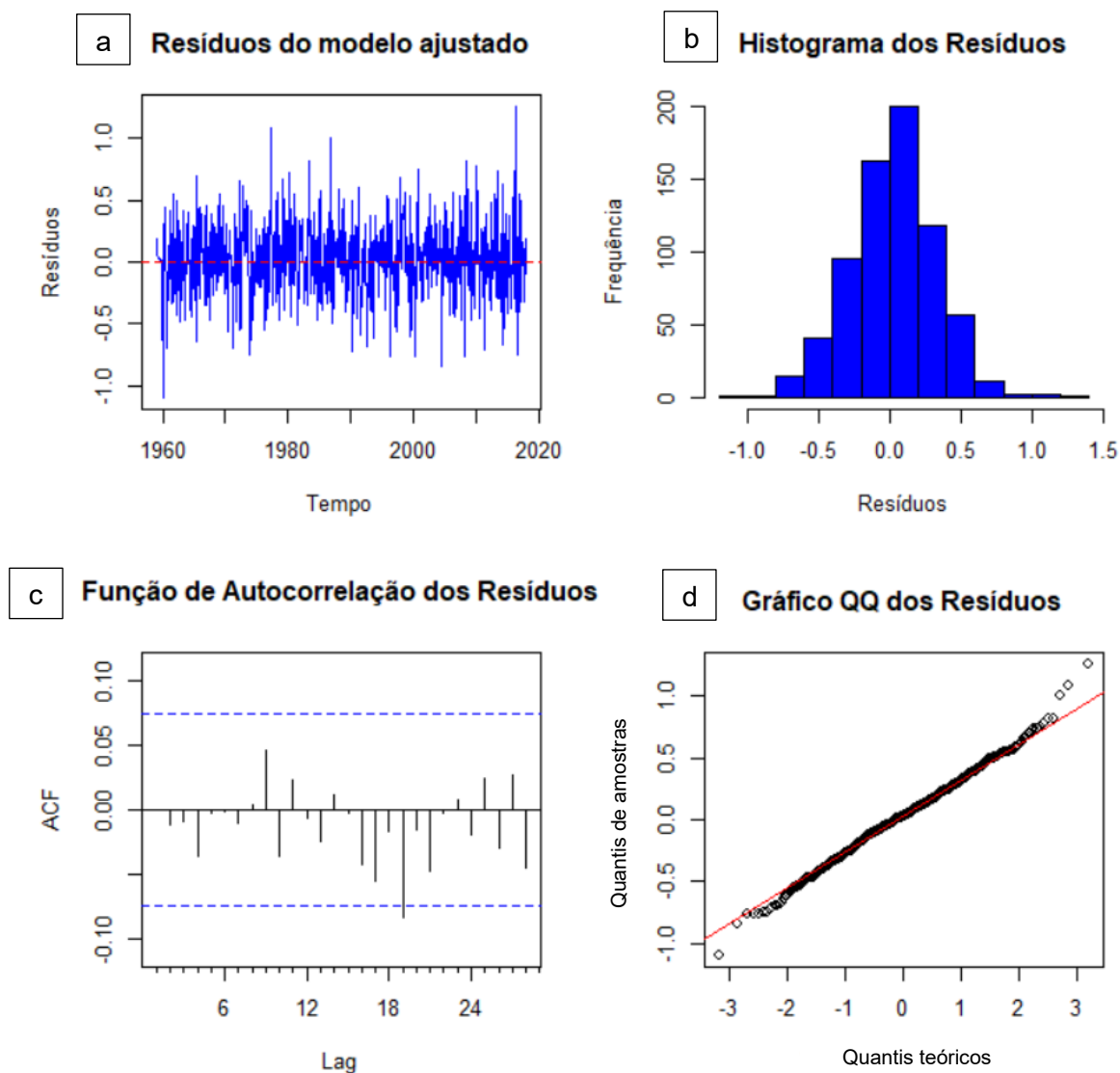


FIGURA 4: AJUSTE DO MODELO ARIMA AOS DADOS DE CO₂: (A) DISPERSÃO, (B) HISTOGRAMA, E (C) GRÁFICOS DE AUTOCORRELAÇÃO E (D) QUANTIL-QUANTIL DOS RESÍDUOS (FONTE: AUTORIA PRÓPRIA).

Observa-se na Figura 4(a) que os resíduos estão distribuídos aleatoriamente em torno de zero; no entanto, o intervalo de variação é mais amplo do que o observado para os resíduos da suavização exponencial. O histograma, Figura 4(b), exibe características de uma distribuição normal, e o gráfico quantil-quantil, Figura 4(d), apresenta um perfil bastante próximo do linear. No gráfico ACF (Função de Autocorrelação), Figura 4(c), observa-se que,

em apenas uma defasagem (*lag*), o coeficiente de autocorrelação excede as linhas pontilhadas, indicando que a maioria das correlações entre os valores da série temporal em diferentes momentos não é estatisticamente significativa. Essa falta de autocorrelação sugere que os resíduos são independentes, o que é desejável em modelos de séries temporais, pois indica que o modelo é apropriado e que as previsões feitas não apresentam padrões não capturados.

O fato da amplitude dos resíduos do ARIMA nas Figuras 2 e 4 ser maior que o da suavização exponencial pode ser uma consequência de sua maior capacidade de modelar complexidades nos dados, levando a previsões que podem ser mais variáveis. Isso não significa que o ARIMA seja um modelo inferior; em vez disso, reflete a diferença na abordagem entre um modelo mais complexo e um mais simples.

As métricas de acurácia ME, MAE, RMSE e MAPE para a previsão com a suavização exponencial e o modelo ARIMA, nos conjuntos de treino e teste, são apresentadas nas Tabelas 1 e 2.

TABELA 1: MÉTRICAS DE ACURÁCIA ME, MAE, RMSE E MAPE DAS PREVISÕES PARA O CONJUNTO TREINO UTILIZANDO A SUAVIZAÇÃO EXPONENCIAL E O MODELO ARIMA

	ME	MAE	RMSE	MAPE	ACF1
Suav. Exp.	0,022	0,243	0,309	0,069	0,078
ARIMA	0,027	0,238	0,306	0,067	0,001

TABELA 2: MÉTRICAS DE ACURÁCIA ME, MAE, RMSE E MAPE DAS PREVISÕES PARA O CONJUNTO TESTE UTILIZANDO A SUAVIZAÇÃO EXPONENCIAL E O MODELO ARIMA

	ME	MAE	RMSE	MAPE	ACF1
Suav. Exp.	0,787	0,840	1,038	0,200	0,761
ARIMA	0,707	0,792	0,983	0,189	0,769

Observa-se na Tabela 1 que o modelo ARIMA se destaca em todas as métricas de acurácia, mostrando um desempenho ligeiramente melhor em termos de MAE, RMSE e MAPE, além de uma autocorrelação dos resíduos muito baixa (ACF1). Isso sugere que o ARIMA é mais eficaz em capturar os padrões nos dados do que a suavização exponencial, principalmente em termos de não deixar autocorrelação nos resíduos, o que é crucial para a validade das previsões. Portanto, se o objetivo é a precisão e a confiabilidade das previsões, no conjunto de treino o modelo ARIMA parece ser a melhor escolha com base nesses resultados.

Na Tabela 2 nota-se que o modelo ARIMA apresenta um desempenho consistentemente melhor em todas as métricas (ME, MAE, RMSE, MAPE) para o conjunto de treino, indicando que ele gera previsões mais precisas em termos de erro absoluto e percentual, assim como em termos de erro quadrático. A diferença entre os valores ACF1 nos dois modelos também não é significativa o suficiente para ser um fator determinante no desempenho geral. Os valores de ACF1 são um pouco elevados, sugerindo que as previsões apresentam autocorrelação significativa e que os erros estão correlacionados com valores anteriores. Isso pode resultar da presença de dependências temporais nos dados, como sazonalidade ou tendências, que não foram totalmente modeladas.

6.3 Previsão

Nesta seção, será apresentada a previsão dos níveis médios de CO₂ global. O modelo ARIMA foi selecionado nas seções anteriores com base na comparação das métricas de acurácia entre os modelos ETS e ARIMA, sendo este último escolhido por apresentar um desempenho superior. Dessa forma, o ARIMA foi utilizado para realizar a projeção dos níveis de CO₂ nos próximos 12 meses.

A previsão será realizada para o período de outubro de 2024 a setembro de 2025, utilizando uma série temporal composta por dados mensais que abrangem o intervalo de janeiro de 1959 a setembro de 2024, totalizando 789 observações. O modelo ARIMA(1,1,1)(0,1,1) foi selecionado por meio da função `auto.arima`, que identificou automaticamente os melhores parâmetros com base nas características da série temporal.

Em seguida, a função `forecast` foi empregada para gerar as projeções, fornecendo não apenas os valores estimados para os próximos 12 meses, mas também os intervalos de confiança de 95%, permitindo avaliar a incerteza inerente às previsões realizadas.

A figura 5 ilustra a previsão dos níveis médios globais de CO₂, revelando uma clara tendência crescente ao longo do tempo. A série temporal original, que vai de janeiro de 1959 a setembro de 2024, é representada pela linha preta contínua, evidenciando o aumento contínuo dos níveis de CO₂ ao longo das décadas. A linha vermelha mostra os valores ajustados pelo modelo ARIMA (1,1,1), refletindo como o modelo se adequa aos dados históricos.

A linha azul, por sua vez, representa as previsões para o período de outubro de 2024 a julho de 2025. Além disso, a área sombreada ao redor da linha azul indica os intervalos de confiança de 95%.

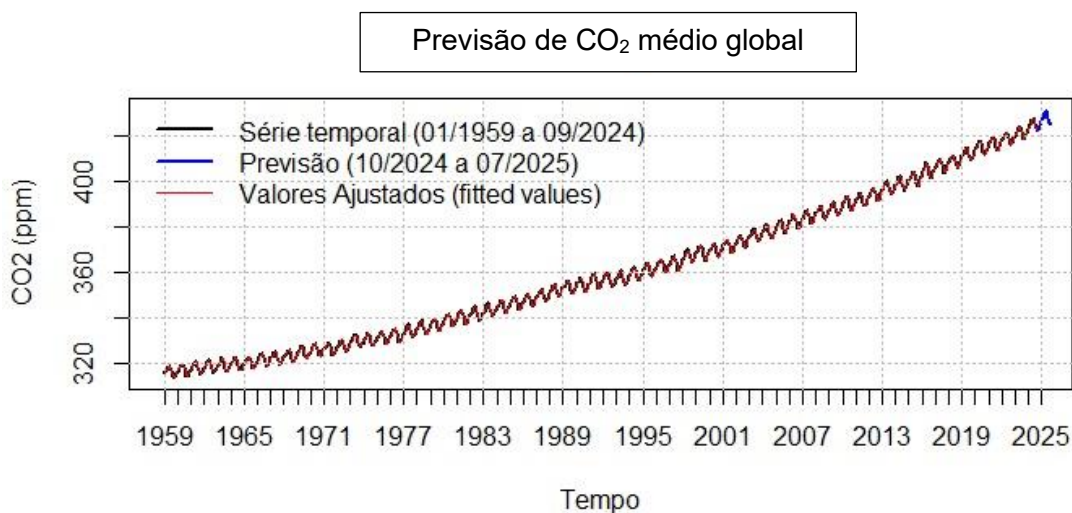


FIGURA 5: PREVISÃO DA EMISSÃO DE CO₂ GLOBAL (FONTE: AUTORIA PRÓPIA)

A Tabela 3 apresenta a previsão dos níveis médios de CO₂ global, expressos em partes por milhão (ppm), para o período de outubro de 2024 a setembro de 2025. Os dados mostram uma continuidade da tendência crescente dos níveis de CO₂, refletindo os padrões observados ao longo das últimas décadas.

Cada linha da tabela indica a previsão mensal, juntamente com os limites inferior e superior de confiança de 95%. O valor previsto para outubro de 2024 é de 422,15 ppm, com um limite inferior de 421,53 ppm e um limite superior de 422,77 ppm. Ao longo dos meses, as previsões mostram um aumento progressivo, atingindo um valor previsto de 430,09 ppm em maio de 2025, antes de apresentar uma leve redução nos meses seguintes.

Os limites de confiança fornecem informações sobre a incerteza associada a cada previsão, permitindo uma melhor interpretação das possíveis variações nos níveis de CO₂. Por exemplo, em fevereiro de 2025, a previsão é de 427,27 ppm, com limites que variam de 426,32 ppm a 428,22 ppm, indicando uma margem de erro que deve ser considerada ao avaliar as projeções futuras.

TABELA 3: PREVISÃO DOS NÍVEIS MÉDIOS DE CO₂ GLOBAL (PPM) PARA O PERÍODO DE OUTUBRO DE 2024 A SETEMBRO DE 2025

Data	Previsão	Limite Inferior 95%	Limite Superior 95%
Outubro 2024	422,1463	421,5267	422,766
Novembro 24	423,6961	422,9582	424,4341
Dezembro 2024	425,131	424,3133	425,9486
Janeiro 2025	426,3089	425,4227	427,195
Fevereiro 2025	427,2716	426,3228	428,2204
Março 25	427,9306	426,9232	428,9381
Abril 2025	429,4367	428,374	430,4995
Mai 2025	430,0926	428,9773	431,208
Junho 2025	429,6259	428,4603	430,7915
Julho 2025	427,8118	426,598	429,0255
Agosto 2025	425,709	424,449	426,9691
Setembro 2025	424,4213	423,117	425,7260

Por meio dos dados disponibilizados e com base na data em que a previsão foi realizada, setembro de 2024, pode-se realizar uma comparação entre a emissão média de CO₂ registrada pelo NOAA e a prevista no estudo para o período de outubro de 2024. O valor disponibilizado pelo NOAA apresentou uma média global de CO₂ (ppm) de 422,38, este valor bate com o previsto considerando o limite de confiança superior. Mostrando de maneira objetiva e com dados que a previsão foi bem ajustada ao modelo.

Com base nas métricas de acurácia obtidas, ficou evidente que o modelo ARIMA se destacou como uma escolha superior para a análise da série temporal. Os resultados indicaram que esse modelo foi capaz de capturar melhor as nuances dos dados, proporcionando previsões mais precisas. Essa superioridade foi evidenciada através de comparações diretas entre os desempenhos dos dois modelos

Apesar do desempenho do ARIMA, a suavização exponencial também se revelou uma abordagem válida. Embora apresentasse erros um pouco maiores, ainda assim forneceu previsões que foram consideradas aceitáveis. Essa flexibilidade sugere que, em determinadas circunstâncias, a suavização exponencial pode ser uma alternativa útil, especialmente em contextos onde a simplicidade do modelo é desejável.

A análise das métricas de acurácia permitiu uma compreensão das capacidades de cada modelo. O ARIMA, com sua complexidade, mostrou-se mais eficaz em capturar padrões mais sutis, enquanto a suavização exponencial apresentou uma abordagem mais direta e menos complexa. Essa comparação entre os dois modelos trouxe à tona a importância de considerar diferentes metodologias na modelagem de séries temporais.

Além disso, a escolha do modelo ideal depende muito do contexto e dos objetivos da análise. Em situações onde a precisão das previsões é crucial, o ARIMA pode ser preferido. Entretanto, para análises exploratórias ou quando um modelo mais simples é suficiente, a suavização exponencial ainda pode ser uma boa opção.

Em síntese, os resultados evidenciam que, embora o modelo ARIMA tenha se destacado como a melhor escolha para a série temporal em questão, a suavização exponencial também deve ser considerada. Essa análise comparativa contribuiu para uma melhor compreensão das técnicas disponíveis e suas respectivas adequações em diferentes cenários analíticos.

7. CONCLUSÃO

A pesquisa possibilitou uma análise das concentrações de CO₂ ao longo do tempo. Através da coleta de dados de fontes confiáveis, como o NOAA, e do subsequente pré-processamento, conseguimos garantir que as informações utilizadas fossem precisas e relevantes. Esse cuidado na etapa inicial foi fundamental para o sucesso da análise.

A análise exploratória realizada revelou padrões significativos nas séries temporais de concentração de CO₂, permitindo uma compreensão mais aprofundada das tendências e sazonalidades presentes nos dados. A definição criteriosa dos conjuntos de treino e teste também contribuiu para a robustez das modelagens. Com esses elementos, estabelecemos uma base sólida para a aplicação dos modelos de previsão.

A modelagem, que envolveu a utilização dos modelos ARIMA e de suavização exponencial, mostrou a eficácia de abordagens quantitativas na previsão de fenômenos ambientais. O modelo ARIMA destacou-se pela sua capacidade de fornecer previsões mais precisas, enquanto a suavização exponencial se mostrou uma alternativa válida em contextos onde a simplicidade é necessária.

A avaliação da performance dos modelos, baseada nas métricas de acurácia, permitiu identificar as *strengths* e limitações de cada abordagem. Essa análise não apenas validou as previsões geradas, mas também orientou decisões futuras sobre qual modelo utilizar em

diferentes cenários. O uso do software R facilitou todo o processo, proporcionando uma plataforma robusta para a implementação de técnicas de análise de séries temporais.

Em síntese, este estudo não apenas contribuiu para o entendimento da evolução da concentração de CO₂ na atmosfera, mas também ressaltou a importância de escolher o modelo de previsão adequado. Os resultados obtidos evidenciam que, para análises complexas como essa, a aplicação de técnicas robustas e a consideração cuidadosa das características dos dados são essenciais.

8. REFERÊNCIAS

ALBUQUERQUE, Rosendo Chagas de. **Modelagem em séries temporais aplicados a dados climatológicos no sertão paraibano.** Disponível em: <https://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/8188/1/PDF%20-%20Rosendo%20Chagas%20de%20Albuquerque.pdf>. Acesso em: 19 out. 2024.

ALBUQUERQUE, Aline Mota. **Mensuração das emissões de gases do efeito estufa e estratégias de mitigação: o caso Coelce.** 2010. 93 f. Dissertação (Mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Fortaleza, 2010.

BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. **A note on the validity of cross-validation for evaluating autoregressive time series prediction.** *Computational Statistics & Data Analysis*, v. 120, p. 70–83, abr. 2018.

BEZERRA, M. I. S. **Análise de séries temporais.** São Paulo: Universidade de São Paulo, 2006.

CAMILO, Gustavo Inocência. **Avaliação da precisão dos métodos ARIMA, suavização exponencial e redes neurais na previsão de séries temporais anuais brasileiras.** Disponível em: <https://repositorio.ufu.br/bitstream/123456789/25914/1/AvaliacaoPrecisaoMetodos.pdf>. Acesso em: 20 out. 2024.

CRYER, J. D.; CHAN, K. S. **Time series analysis: with applications in R.** New York: Springer, 2008.

EHLERS, R. S. **Análise de séries temporais.** Porto Alegre: Universidade Federal do Rio Grande do Sul, 2007.

HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K.; SNYDER, R. D. **Forecasting with exponential smoothing: the state space approach.** Berlin: Springer-Verlag, 2008.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice.** 3. ed. Melbourne: Otexts, 2021.

GUIMARÃES, C. C.; DORN, R. C. **Efeito estufa usando material alternativo.** *Química Nova na Escola*, São Paulo, v. 37, n. 2, p. 153–157, maio 2015.

JAMES, G. et al. **An introduction to statistical learning: with applications in R.** [S.l.]: Springer, 2013.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais.** São Paulo: Edgard Blücher, 2006.

NOAA (National Oceanic and Atmospheric Administration). Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases. **Trends in CO₂.** Disponível em: <https://gml.noaa.gov/ccgg/trends/>. Acesso em: 11 out. 2024.

PELEGRINI, M.; ARAÚJO, W. R. B. **Efeito estufa e camada de ozônio sob a perspectiva da interação radiação-matéria e uma abordagem dos acordos internacionais sobre o clima.** *Química Nova na Escola*, São Paulo, v. 40, n. 2, p. 72–78, maio 2018.

RODRIGUES, Luciene Cleia; SILVA, Paula Pinho Cabral da; LINDEN, Ricardo. **Séries temporais no consumo de energia elétrica no estado do Rio de Janeiro.** Rio de Janeiro: Universidade Estadual do Rio de Janeiro, 2015.

9. APÊNDICE

9.1 Rotinas R para Ajuste do Modelo de Suavização Exponencial

A seguir, são apresentados os scripts utilizados no processo de ajuste do modelo de suavização exponencial com a biblioteca *forecast* no R. O objetivo é prever a série temporal de CO2 médio global com base em dados históricos e avaliar a acurácia das previsões.

9.1.1 Carregamento das bibliotecas e dos dados

Primeiro, carregam-se as bibliotecas necessárias para o processamento e visualização dos dados, seguidas pela leitura dos arquivos contendo as séries temporais:

```
library(forecast)
library(ggplot2)
library(ggfortify)
# Leitura dos dados de entrada
dados <- read.table("Dados.txt", header = F, sep = "\t", dec = ",")
teste <- read.table("Teste.txt", header = F, sep = "\t", dec = ",")
dados.full <- read.table("Dados_full.txt", header = F, sep = "\t", dec = ",")
```

9.1.2 Transformação dos dados em séries temporais

As séries de dados são transformadas em objetos do tipo série temporal (ts), que facilitam o trabalho com modelos de previsão no R. A função *ts()* é usada para especificar o período de início e fim, bem como a frequência (número de observações por ano).

```
# Definição das séries temporais
dados.fullts <- ts(dados.full, start = c(1959, 1), end = c(2024, 9), frequency = 12)
dados.ts <- ts(dados, start = c(1959, 1), end = c(2017, 12), frequency = 12)
teste.ts <- ts(teste, start = c(2018, 1), end = c(2024, 9), frequency = 12)
```

9.1.3 Ajuste do modelo de suavização exponencial

O modelo de suavização exponencial é ajustado à série temporal de CO2 histórico utilizando a função *ets()* da biblioteca *forecast*. O parâmetro "ZZZ" especifica que o modelo selecionará automaticamente os melhores componentes (suavização, tendência e sazonalidade).

```
co2 <- dados.ts
ajuste <- ets(co2, "ZZZ") # Ajuste do modelo de suavização exponencial
```

9.1.4 Previsão dos valores futuros

Após o ajuste do modelo, é gerada a previsão para o período de teste (2018-2024), com intervalos de confiança de 95%.

```
ajuste_prev <- forecast(ajuste, h = length(teste.ts), level = 95)
print(ajuste_prev)
```

9.1.5 Avaliação da acurácia do modelo

A acurácia do modelo é verificada comparando as previsões geradas com os valores reais da série de teste (2018-2024), utilizando a função `accuracy()`.

```
acuracia <- accuracy(ajuste_prev$mean, teste.ts)
print(acuracia)
```

9.1.6 Visualização dos resultados

Por fim, o script gera um gráfico para visualizar a série temporal original, as previsões ajustadas e os intervalos de confiança. O gráfico é ajustado com linhas e legendas para facilitar a interpretação.

```
# Visualização dos resultados
plot(dados.fullts, main = "CO2 médio global", xlab = "Anos", ylab = "CO2 (ppm)", xaxt = "n")
axis(1, at = seq(from = 1959, to = 2025, by = 2), labels = seq(from = 1959, to = 2025, by = 2),
las = 1)
lines(fitted(ajuste), lwd = 1, lty = 1, col = "red") # Série ajustada
lines(ajuste_prev$mean, lwd = 2, lty = 1, col = "blue") # Previsões
abline(v = seq(1959, 2025, 6), h = seq(320, 420, 20), lty = 3, col = "gray") # Linhas de
referência
# Intervalos de confiança
if (!is.null(ajuste_prev$lower) && !is.null(ajuste_prev$upper)) {
  polygon(c(time(ajuste_prev$mean), rev(time(ajuste_prev$mean))),
    c(ajuste_prev$lower[, 1], rev(ajuste_prev$upper[, 1])),
    col = rgb(0, 0, 1, 0.2), border = NA)
}
# Adicionando legenda
legend("topleft", legend = c("Série temporal (01/1958 a 09/2024)",
  "Série de treino (01/1959 a 12/2017)",
  "Série de teste (01/2018 a 09/2024)"),
  lwd = c(1, 1, 1), lty = c(1, 1, 1), col = c("black", "red", "blue"), bty = 'n')
```

9.1.7 Análise de Resíduos do Modelo Ajustado

Após o ajuste e a previsão a análise de resíduos é uma etapa importante para verificar a qualidade do modelo ajustado. Resíduos próximos de zero e sem padrões evidentes indicam que o modelo está ajustado adequadamente. São gerados diversos gráficos para analisar os resíduos:

- **Gráfico dos Resíduos:** Mostra os resíduos ao longo do tempo.
- **Histograma dos Resíduos:** Mostra a distribuição dos resíduos.
- **Função de Autocorrelação (ACF):** Avalia a autocorrelação dos resíduos.
- **Gráfico QQ:** Avalia a normalidade dos resíduos.

```
# Análise de resíduos
residuos <- residuals(ajuste)
# Configurar layout para múltiplos gráficos
par(mfrow = c(2, 2))
# Gráfico dos resíduos
plot(residuos, main = "Resíduos do modelo ajustado", ylab = "Resíduos", xlab = "Tempo", type
= 'l', col = "blue")
abline(h = 0, col = "red", lty = 2)
# Gráfico de histograma dos resíduos
hist(residuos, main = "Histograma dos Resíduos", ylab = "Frequência", xlab = "Resíduos", col
= "blue", border = "black")
# Função de Autocorrelação dos resíduos
Acf(residuos, main = "Função de Autocorrelação dos Resíduos")
# Gráfico de QQ dos resíduos
qqnorm(residuos, main = "Gráfico QQ dos Resíduos")
qqline(residuos, col = "red")
# Resetar o layout gráfico
par(mfrow = c(1, 1))
```

9.2 Rotinas R para Ajuste do Modelo ARIMA

Este apêndice descreve o uso da função `auto.arima()` da biblioteca `forecast` para ajustar um modelo ARIMA automático à série temporal de CO2. O script inclui a realização de previsões, avaliação da acurácia e visualização dos resultados, bem como o cálculo dos intervalos de confiança para as previsões.

9.2.1 Carregamento das Bibliotecas e Dados

Primeiro, são carregadas as bibliotecas necessárias para o processamento de séries temporais (`forecast`, `ggplot2`, `ggfortify`). Em seguida, os dados são lidos a partir de arquivos de texto. Os dados são divididos em três conjuntos: dados históricos (`Dados.txt`), dados de teste (`Teste.txt`) e dados completos (`Dados_full.txt`).

```
library(forecast)
library(ggplot2)
library(ggfortify)
# Carregar os dados
dados <- read.table("Dados.txt", header = F, sep = "\t", dec = ",")
teste <- read.table("Teste.txt", header = F, sep = "\t", dec = ",")
dados.full <- read.table("Dados_full.txt", header = F, sep = "\t", dec = ",")
```

9.2.2 Criação das Séries Temporais

As séries temporais são criadas utilizando a função `ts()` para transformar os dados em objetos que permitem a modelagem e previsão. As séries são definidas com o ano e o mês como ponto de partida e com a frequência mensal (`frequency = 12`).

```
# Criar séries temporais
dados.fullts <- ts(dados.full, start = c(1959, 1), end = c(2024, 9), frequency = 12)
dados.ts <- ts(dados, start = c(1959, 1), end = c(2017, 12), frequency = 12)
teste.ts <- ts(teste, start = c(2018, 1), end = c(2024, 9), frequency = 12)
```

9.2.3 Ajuste do Modelo ARIMA Automático

A função `auto.arima()` é usada para ajustar um modelo ARIMA (Auto-Regressive Integrated Moving Average) à série temporal de CO₂. A função seleciona automaticamente o melhor modelo com base nos critérios de informação. O parâmetro `stepwise = FALSE` desativa o método passo a passo para a busca do modelo, e o `approximation = FALSE` assegura que os cálculos sejam feitos sem aproximações. O parâmetro `trace = TRUE` exibe as etapas do processo de busca do modelo.

```
# Ajuste do modelo ARIMA automático
co2 <- dados.ts
ajuste <- auto.arima(co2, stepwise = FALSE, approximation = FALSE, trace = TRUE)
```

9.2.4 Previsão dos Valores Futuros

Uma vez ajustado o modelo ARIMA, as previsões são feitas utilizando a função `forecast()`. A previsão é realizada para o período de teste (2018-2024), e os intervalos de confiança de 95% são calculados automaticamente.

```
# Prever os valores futuros
ajuste_prev <- forecast(ajuste, h = length(teste.ts), level = 95)
print(ajuste_prev)
```

9.2.5 Cálculo da Acurácia do Modelo

A acurácia do modelo é verificada utilizando a função `accuracy()`, que compara as previsões geradas com os valores reais da série de teste.

```
# Calcular a acurácia
acuracia <- accuracy(ajuste_prev$mean, teste.ts)
```

```
print(acuracia)
```

9.2.6 Visualização dos Resultados

A visualização dos resultados é feita por meio de gráficos que comparam a série temporal original, o ajuste do modelo ARIMA e as previsões realizadas.

```
# Visualização dos resultados
plot(dados.fullts, lwd = 2, main = "CO2 médio global", xlab = "Tempo", ylab = "CO2 (ppm)",
xaxt = "n")
lines(teste.ts, lwd = 2, lty = 2, col = "black")
axis(1, at = seq(from = 1959, to = 2025, by = 2), labels = seq(from = 1959, to = 2025, by = 2),
las = 1)
lines(fitted(ajuste), lwd = 1, lty = 1, col = "red") # Série ajustada
lines(ajuste_prev$mean, lwd = 2, lty = 1, col = "blue") # Previsões
abline(v = seq(1959, 2025, 6), h = seq(320, 420, 20), lty = 3, col = "gray") # Linhas de
referência

# Verificar a estrutura do objeto de previsão para obter os intervalos de confiança
if (!is.null(ajuste_prev$lower) && !is.null(ajuste_prev$upper)) {
  polygon(c(time(ajuste_prev$mean), rev(time(ajuste_prev$mean))),
    c(ajuste_prev$lower[, 1], rev(ajuste_prev$upper[, 1])),
    col = rgb(0, 0, 1, 0.2), border = NA)
}
# Adicionar legenda
legend("topleft", legend = c("Série temporal (01/1958 a 09/2024)",
  "Série de treino (01/1959 a 12/2017)",
  "Série de teste (01/2018 a 09/2024)"),
  lwd = c(1, 1, 1), lty = c(1, 1, 1), col = c("black", "red", "blue"), bty = 'n')
```

9.2.7 Análise de Resíduos do Modelo Ajustado

Após o ajuste e a previsão a análise de resíduos é uma etapa importante para verificar a qualidade do modelo ajustado. Resíduos próximos de zero e sem padrões evidentes indicam que o modelo está ajustado adequadamente. São gerados diversos gráficos para analisar os resíduos:

- **Gráfico dos Resíduos:** Mostra os resíduos ao longo do tempo.
- **Histograma dos Resíduos:** Mostra a distribuição dos resíduos.
- **Função de Autocorrelação (ACF):** Avalia a autocorrelação dos resíduos.
- **Gráfico QQ:** Avalia a normalidade dos resíduos.

```
# Análise de resíduos
residuos <- residuals(ajuste)
# Configurar layout para múltiplos gráficos
par(mfrow = c(2, 2))
# Gráfico dos resíduos
```

```
plot(residuos, main = "Resíduos do modelo ajustado", ylab = "Resíduos", xlab = "Tempo", type  
= 'l', col = "blue")  
abline(h = 0, col = "red", lty = 2)  
# Gráfico de histograma dos resíduos  
hist(residuos, main = "Histograma dos Resíduos", ylab = "Frequência", xlab = "Resíduos", col  
= "blue", border = "black")  
# Função de Autocorrelação dos resíduos  
Acf(residuos, main = "Função de Autocorrelação dos Resíduos")  
# Gráfico de QQ dos resíduos  
qqnorm(residuos, main = "Gráfico QQ dos Resíduos")  
qqline(residuos, col = "red")  
# Resetar o layout gráfico  
par(mfrow = c(1, 1))
```