

ESTUDO DA CONCENTRAÇÃO DE CO₂ E DAS ANOMALIAS DE TEMPERATURA UTILIZANDO O MÉTODO DOS MÍNIMOS QUADRADOS

João Pedro Fernandes de Aquino¹, Ivan Mezzomo²

Resumo: Este estudo visa utilizar conceitos atmosféricos, climáticos e computacionais a fim de analisar dados públicos para obter informações relacionadas a como esses dados se comportam, auxiliando no aprimoramento de modelos estatísticos voltados à análise climática. O objetivo deste trabalho é analisar dados históricos de concentração de CO₂ e anomalias de temperatura, aplicando o Método dos Mínimos Quadrados (MMQ) com diferentes funções de aproximação para determinar o melhor ajuste. A qualidade das funções de aproximação foi avaliada pelo coeficiente de determinação R², o teste t de Student para a significância dos coeficientes e o teste F de ANOVA para a significância das funções, assegurando uma análise robusta dos dados climáticos.

Palavras-chave: Método dos Mínimos Quadrados; Análise Estatística; Python; Concentração de CO₂

1. Introdução

Nos últimos anos os efeitos das mudanças climáticas têm se mostrado cada vez mais evidente a partir dos mais diferentes fenômenos por todo o planeta. Essas rápidas mudanças impulsionadas por ações antropogênicas impactam diretamente a nossa sociedade, resultando em desastres climáticos que ameaçam a vida e a subsistência de milhões de pessoas. Diante desse cenário, torna-se crescente o interesse em entender as causas e os comportamentos dessas mudanças de forma a buscar mitigar suas causas e consequências.

Um dos principais responsáveis pelas mudanças climáticas globais é o aumento na concentração de dióxido de carbono (CO₂). Esse crescimento contínuo nos últimos anos impacta diretamente os padrões climáticos do planeta a partir da intensificação do efeito estufa. Paralelamente a isso, as anomalias de temperatura do planeta são um indicador chave para avaliar os efeitos dessas mudanças, medindo registros de temperaturas que se desviam além do esperado para uma média histórica. A partir disso, a realização de uma análise conjunta dos níveis de CO₂ e das anomalias de temperatura é necessária e fundamental para uma melhor compreensão da dinâmica climática afim de prever e mitigar impactos futuros.

A análise dos dados será realizada com base no Método dos Mínimos Quadrados (MMQ), uma técnica que permite determinar uma função de aproximação $\varphi(x)$ que melhor se ajusta a um conjunto de pontos, minimizando os erros entre os valores observados e os valores de $\varphi(x)$. A minimização dos erros torna a aplicação do MMQ uma abordagem precisa e coesa para a modelagem dos dados e análise dos resultados, sendo essa a principal razão para sua escolha neste trabalho.

O estudo utilizará dados mensais de concentração de CO₂ e de anomalias de temperatura fornecidos pelo Global Monitoring Laboratory e pelo National Centers for Environmental Information, respectivamente, abrangendo o período de 1980 a 2023. A preparação dos dados e os métodos estatísticos serão aplicados computacionalmente utilizando linguagem de programação Python por sua flexibilidade na manipulação dos dados e em modelagens estatísticas.

Neste estudo, vamos investigar os impactos das mudanças climáticas a partir da concentração de CO₂ e das anomalias de temperatura, contribuindo na compreensão de como esses dados se comportam a partir da escolha das funções de aproximação. Além disso, as análises realizadas visam fornecer recomendações para futuras pesquisas dentro desse campo facilitando novas abordagens ao tema.

Neste contexto, o presente trabalho irá utilizar as funções de aproximação do MMQ, visando estudar qual dessas funções de aproximação do MMQ melhor se ajusta a dados relacionados à concentração do CO₂ e às anomalias de temperatura. Para avaliar a qualidade das funções de aproximação resultantes do MMQ, será o coeficiente de determinação (R²), o teste t de Student e o teste F de ANOVA. Esses métodos de avaliação foram escolhidos por sua eficácia em analisar a proporção da variabilidade dos dados e em avaliar a significância dos coeficientes e adequação geral da função de aproximação, possibilitando uma análise completa e precisa da qualidade das funções de aproximação.

2. Fundamentação Teórica

A análise das mudanças atmosféricas e de temperatura tem se mostrado fundamental para compreender as mudanças climáticas e seus impactos no clima global. A partir da utilização do método dos mínimos quadrados para os dados históricos dessas mudanças, é possível ajustar modelos de regressão, que auxiliam na interpretação da relação entre as variáveis estudadas. Segundo [6], os modelos de regressão são técnicas estatísticas utilizadas para estudar a relação entre duas variáveis. Assim, visando garantir a confiabilidade dessas análises e dos modelos

em que elas são fundamentadas, torna-se necessário a realização de testes de hipótese sobre os coeficientes dos modelos, verificando-se a significância estatística das estimativas e a adequação do modelo proposto.

2.1. Mudanças climáticas e efeito estufa

Para [1], o clima é uma característica fundamental do ambiente terrestre, sendo ele a força dinâmica que modela diversas regiões do globo. O sistema climático é composto de atmosfera, superfície terrestre, corpos d'água e seres vivos, sendo ele descrito em termos de variabilidade de média de temperatura, precipitação e vento para um determinado período, sendo, por convenção, 30 anos o período mínimo a ser considerado.

O clima e todo o seu sistema são dinâmicos, sendo influenciado por seu funcionamento interno e fatores externos, como fenômenos naturais e ações antropogênicas, como pode ser visto na Figura 1.

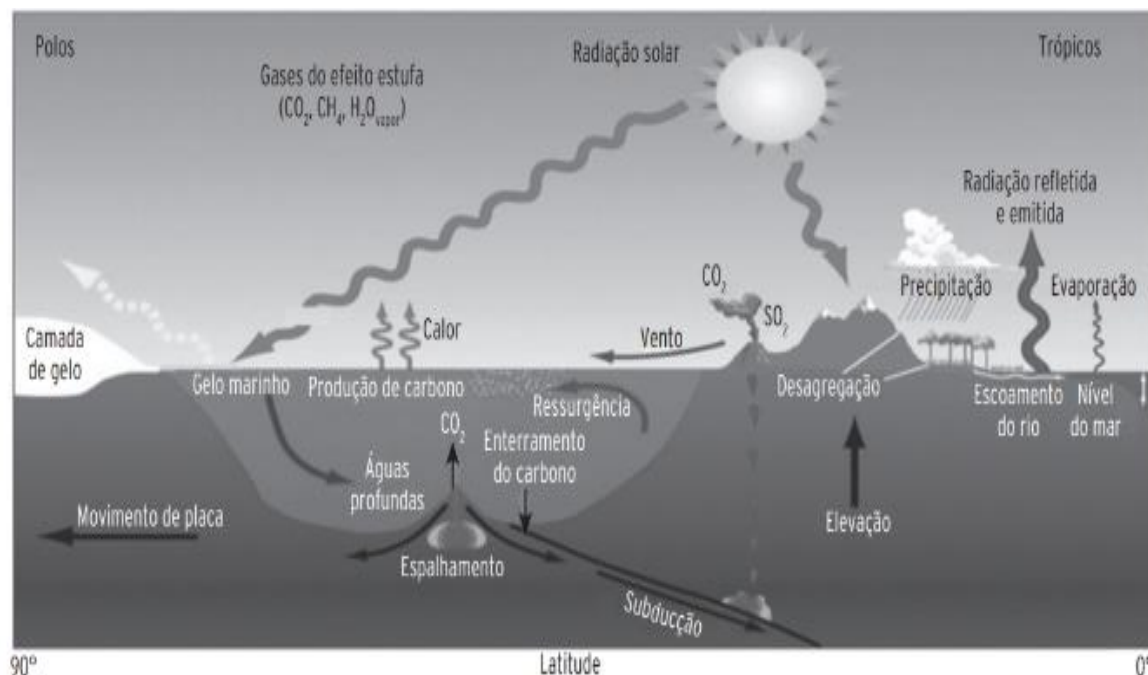


Figura 1. Representação do Sistema Climático. Fonte: adaptado de [1].

Na Figura 1, é possível observar que a radiação solar possui papel fundamental nas interações do sistema climático, aquecendo a superfície terrestre e influenciando padrões climáticos globais. A formação de nuvens desempenha um papel crucial no sistema climático ao refletirem a radiação solar de volta ao espaço (albedo) e ao reterem o calor emitido pela superfície terrestre. A camada de gelo marinho, com seu alto albedo, reflete a luz solar enquanto correntes oceânicas redistribuem calor influenciando o clima global e correntes de ressurgência trazem nutrientes à superfície, impactando ecossistemas marinhos e a absorção de carbono. O movimento das placas tectônicas afeta o clima ao longo de escalas geológicas, causando erupções vulcânicas e alterando a geografia dos continentes, impactando padrões climáticos e correntes oceânicas.

Como visto na imagem acima, esses componentes interagem de maneira dinâmica e interconectada, onde alterações em um dos fatores componentes do sistema pode desencadear mudanças em outros, levando a significativas mudanças climáticas globais. Segundo [1], existem três fatores que podem alterar o balanço da radiação na Terra, sendo eles a alteração na radiação solar incidente, as alterações no albedo de uma região e, por fim, alterações na radiação de onda longa (infravermelha) dissipada para o espaço. Esta última sofre influência direta do efeito estufa, sendo ele responsável por manter parte dessa radiação no planeta.

De acordo com [2], o efeito estufa é fundamental para a vida na Terra, uma vez que sem ele a temperatura terrestre seria aproximadamente 30°C mais baixa. Esse efeito ocorre devido a presença de gases na atmosfera capazes de absorver energia em forma de radiação, sendo a concentração de dióxido de carbono (CO₂) um dos principais contribuintes para esse fenômeno, medida em partes por milhão (ppm). O papel do efeito estufa, entretanto, pode ser nocivo ao planeta, uma vez que intensificado com o excesso de gases, irá reter calor em excesso no planeta aumentando a sua temperatura de forma prejudicial para os seres que nela habitam, caracterizando assim o fenômeno conhecido como aquecimento global. O excesso de gases de efeito estufa como o CO₂ é o resultado da atividade humana, especialmente a partir do século XVIII, com o advento da revolução industrial e o uso de combustíveis fósseis para alimentar as máquinas desenvolvidas.

2.2. Método dos Mínimos Quadrados

De acordo com [5], o Método dos Mínimos Quadrados (MMQ) fornece uma maneira de determinar uma função de aproximação $\varphi(x)$ que melhor se aproxima de uma função $f(x)$ ou para um conjunto de pontos (x_m, y_m) .

Segundo [3], é possível observar que essa aproximação pode ser trabalhada para o caso contínuo, em que $f(x) \in [a, b]$, onde $[a, b]$ é um intervalo fechado e limitado, ou no caso discreto, onde o tabelamento é dado por um conjunto de pares ordenados (x_m, y_m) . Neste trabalho, abordaremos o caso discreto devido ao fato de se tratar de um tabelamento de conjunto de pontos. Segundo [4], para o caso discreto, dados os pontos $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ e n funções $g_1(x), g_2(x), \dots, g_n(x)$ escolhidas de alguma forma, com o número de pontos m , tabelados, maior que o número n de funções, nosso objetivo é encontrar os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_n$, de forma que a função

$$\varphi(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \dots + \alpha_n g_n(x) \quad (1)$$

se aproxime ao máximo do conjunto de pontos dos pares ordenados.

Ainda de acordo com [4], o MMQ consiste em escolher os coeficientes α_j 's que minimizam a soma dos quadrados dos desvios. Portanto, os coeficientes devem minimizar a função

$$F(\alpha_1, \alpha_2, \dots, \alpha_n) = \sum_{k=1}^m [y_k - \varphi(x_k)]^2. \quad (2)$$

Usando o cálculo diferencial, calculando as derivadas parciais $\frac{\partial F}{\partial \alpha_j}$ da Equação (2), para $1 \leq j \leq n$, temos:

$$\frac{\partial F}{\partial \alpha_j} = 2 \sum_{k=1}^m [y_k - \alpha_1 g_1(x_k) - \alpha_2 g_2(x_k) - \dots - \alpha_n g_n(x_k)] [-g_j(x_k)] \quad (3)$$

Para minimizar a função descrita na Equação (2), é necessário encontrar seus pontos críticos aplicando a condição $\frac{\partial F}{\partial \alpha_j} = 0$ e dividindo por 2 a Equação (3), temos:

$$\sum_{k=1}^m [y_k - \alpha_1 g_1(x_k) - \alpha_2 g_2(x_k) - \dots - \alpha_n g_n(x_k)] [-g_j(x_k)] = 0$$

A partir disso é possível montar o seguinte sistema de equações lineares:

$$\begin{cases} \sum_{k=1}^m [y_k - \alpha_1 g_1(x_k) - \alpha_2 g_2(x_k) - \dots - \alpha_n g_n(x_k)] [-g_1(x_k)] = 0 \\ \sum_{k=1}^m [y_k - \alpha_1 g_1(x_k) - \alpha_2 g_2(x_k) - \dots - \alpha_n g_n(x_k)] [-g_2(x_k)] = 0 \\ \vdots \\ \sum_{k=1}^m [y_k - \alpha_1 g_1(x_k) - \alpha_2 g_2(x_k) - \dots - \alpha_n g_n(x_k)] [-g_n(x_k)] = 0 \end{cases} \quad (4)$$

A partir das propriedades do somatório, manipula-se a Equação (4), passando os termos com y para o lado direito da igualdade, resultando no sistema de equações lineares com n equações e n incógnitas:

$$\begin{cases} \sum_{k=1}^m [\alpha_1 (g_1(x_k) \cdot g_1(x_k)) + \alpha_2 (g_1(x_k) \cdot g_2(x_k)) + \dots + \alpha_n (g_1(x_k) \cdot g_n(x_k))] = \sum_{k=1}^m y_k g_1(x_k) \\ \sum_{k=1}^m [\alpha_1 (g_2(x_k) \cdot g_1(x_k)) + \alpha_2 (g_2(x_k) \cdot g_2(x_k)) + \dots + \alpha_n (g_2(x_k) \cdot g_n(x_k))] = \sum_{k=1}^m y_k g_2(x_k) \\ \vdots \\ \sum_{k=1}^m [\alpha_1 (g_n(x_k) \cdot g_1(x_k)) + \alpha_2 (g_n(x_k) \cdot g_2(x_k)) + \dots + \alpha_n (g_n(x_k) \cdot g_n(x_k))] = \sum_{k=1}^m y_k g_n(x_k) \end{cases} \quad (5)$$

As funções $g_i(x)$ que irão compor a função de aproximação devem ser escolhidas analisando o comportamento do tabelamento de pontos. De acordo com [4], a família de funções escolhidas pode não ser linear nos parâmetros, se tornando necessário efetuar uma linearização do problema através de alguma transformação conveniente. Assim, a partir dessa transformação, os valores dos coeficientes são calculados de maneira semelhante a função de aproximação linear. Visando reduzir as demonstrações teóricas ao escopo do trabalho, será demonstrado apenas os casos abordados.

2.2.1 Função Linear

A função linear possui a forma $y = \alpha_1 + \alpha_2 x$. A escolha dessa função para se aproximar de um determinado tabelamento de pontos implica na utilização das funções $g_1(x) = 1$ e $g_2(x) = x$ na Equação (1), sendo a função de aproximação dada por $\varphi(x) = \alpha_1 + \alpha_2 x$.

Substituindo as funções $g_1(x) = 1$ e $g_2(x) = x$ na Equação (5), temos que

$$\begin{cases} \sum_{k=1}^m (\alpha_1 + \alpha_2 x_k) = \sum_{k=1}^m y_k \\ \sum_{k=1}^m (\alpha_1 x_k + \alpha_2 x_k^2) = \sum_{k=1}^m y_k x_k \end{cases} \quad (6)$$

Com a resolução da Equação (6), o valor das incógnitas α_1 e α_2 da função de aproximação podem ser obtidos por:

$$\alpha_1 = \frac{\sum_{k=1}^m y_k x_k \sum_{k=1}^m x_k - \sum_{k=1}^m y_k \sum_{k=1}^m x_k^2}{(\sum_{k=1}^m x_k)^2 - m \sum_{k=1}^m x_k^2} \quad (7)$$

e

$$\alpha_2 = \frac{\sum_{k=1}^m y_k \sum_{k=1}^m x_k^2 - m \sum_{k=1}^m y_k x_k}{(\sum_{k=1}^m x_k)^2 - m \sum_{k=1}^m x_k^2}. \quad (8)$$

2.2.2 Função Quadrática

A função quadrática possui a forma $y = \alpha_1 + \alpha_2 x + \alpha_3 x^2$. A escolha dessa função para se aproximar de um determinado tabelamento de pontos implica na utilização das funções $g_1(x) = 1$, $g_2(x) = x$ e $g_3(x) = x^2$ na Equação (1), sendo a função de aproximação dada por $\varphi(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2$.

Substituindo as funções $g_1(x) = 1$, $g_2(x) = x$ e $g_3(x) = x^2$ na Equação (5), temos que seus coeficientes $\alpha_1, \alpha_2, \alpha_3$ são obtidos a partir da resolução do sistema de equações lineares:

$$\begin{cases} \sum_{k=1}^m (\alpha_1 + \alpha_2 x_k + \alpha_3 x_k^2) = \sum_{k=1}^m y_k \\ \sum_{k=1}^m (\alpha_1 x_k + \alpha_2 x_k^2 + \alpha_3 x_k^3) = \sum_{k=1}^m y_k x_k \\ \sum_{k=1}^m (\alpha_1 x_k^2 + \alpha_2 x_k^3 + \alpha_3 x_k^4) = \sum_{k=1}^m y_k x_k^2 \end{cases} \quad (9)$$

2.2.3 Função Logarítmica

A função logarítmica é uma função da forma $y = \alpha_1 + \alpha_2 \cdot \ln(x)$. A escolha dessa função para se aproximar de um determinado tabelamento de pontos implica na utilização das funções $g_1(x) = 1$ e $g_2(x) = \ln(x)$ na Equação (1), sendo a função de aproximação dada por

$$\varphi(x) = \alpha_1 + \alpha_2 \cdot \ln(x). \quad (10)$$

A partir da aplicação da Equação (10) nas equações (7) e (8), temos por consequência a utilização dos valores do tabelamento de $\ln(x)$ e y para o cálculo dos coeficientes α_1 e α_2 .

2.2.4 Função Exponencial

A função exponencial corresponde a uma função do tipo $y = \beta_1 e^{\beta_2 x}$. Segundo [4], para realizar a linearização desta função é necessário aplicar logaritmo neperiano em ambos os lados da equação exponencial e manipular essa função usando as propriedades de logaritmo. Considerando que $\varphi(x) = \ln(y)$, $\beta_1 = e^{\alpha_1}$ e $\beta_2 = \alpha_2$, temos que a função de aproximação linearizada é representada por

$$\varphi(x) = \alpha_1 + \alpha_2 x. \quad (11)$$

Aplicamos a Equação (11) nas Equações (7) e (8), e utilizando os valores de x e $\ln(y)$ no tabelamento para o cálculo dos coeficientes α_1 e α_2 .

2.2.5 Função Potencial

A função potencial corresponde a uma função do tipo $y = \beta_1 x^{\beta_2}$. A linearização da função é realizada aplicando logaritmo neperiano em ambos os lados da equação potencial, utilizando as propriedades de logaritmo para manipular essa equação. Considerando que $\varphi(x) = \ln(y)$, $\beta_1 = e^{\alpha_1}$ e $\beta_2 = \alpha_2$, temos que a função de aproximação linearizada é representada por

$$\varphi(x) = \alpha_1 + \alpha_2 \cdot \ln(x). \quad (12)$$

Aplicamos a Equação (12) nas Equações (7) e (8), e utilizando os valores de $\ln(x)$ e $\ln(y)$ no tabelamento para o cálculo dos coeficientes α_1 e α_2 .

2.3. Métodos estatísticos

Para avaliar um modelo de regressão existem várias abordagens que visam medir sua eficácia e relevância. Visando reduzir o escopo do trabalho e abordar as diferentes partes que integram a regressão, serão discutidas a análise do coeficiente de determinação e testes de hipótese, considerando o teste t de Student para coeficientes da regressão e do teste F de Análise de Variância (ANOVA).

2.3.1 Coeficiente de Determinação

O coeficiente de determinação R^2 é usado para avaliar a qualidade do ajuste de um modelo de regressão, indicando a proporção da variabilidade da variável dependente que é explicada pelas variáveis independentes no modelo. Para um dado tabelamento de m pontos $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, o coeficiente é calculado pela equação abaixo

$$R^2 = \frac{\sum_{k=1}^m [\varphi(x_k) - \bar{y}]^2}{\sum_{k=1}^m (y_k - \bar{y})^2} \quad (13)$$

onde $\bar{y}$ é a média dos valores de y e y_k é o valor de índice k no tabelamento. Quanto mais próximo de 1 for o valor do coeficiente de determinação, mais a variabilidade de y leva em consideração a variação em x . Entretanto, o coeficiente de determinação deve ser analisado com cautela, podendo apresentar valores próximos a 1 mesmo em casos que as variáveis não possuam relação lineares. Além disso, é sempre possível obter maiores valores de R^2 apenas aumentando a quantidade de termos da função de aproximação, como em funções polinomiais [6].

2.3.2 Teste de Hipótese

De acordo com [7], o teste de hipótese fornece uma metodologia na qual possa ser verificada se uma afirmação é verdadeira ou falsa para determinados dados amostrais de uma população. A ideia principal do procedimento que, dada uma variável X associada a uma população e um determinado parâmetro θ , será feita uma afirmação chamada hipótese nula, do tipo $H_0: \theta = \theta_0$, para colocar à prova. Por consequência, haverá uma hipótese alternativa, que será aceitável no caso da negação da hipótese nula. Para um caso geral: $H_1: \theta \neq \theta_0$.

Em um teste de hipótese a existência de erros pode ser dividida em Erro de tipo I e Erro de tipo II, sendo mais comumente considerado para a aplicação do teste hipótese o erro de tipo I. A probabilidade de cometer esse erro é chamada de valor-p, e seu cálculo depende do grau de liberdade e tipo de teste escolhido para a hipótese. O valor-p deve ser comparado a um valor arbitrário α chamado de nível de significância, que representa a taxa máxima de erros aceitável para o teste. Quanto menor o nível de significância, menor a probabilidade da hipótese H_0 ser verdadeira, uma vez que isso ocorrerá quando α for maior que o valor-p. A decisão de validação da hipótese

também pode ser definida a partir de uma região crítica (RC), que será construída a partir do tipo de teste aplicado. Caso o valor observado da estatística pertença a RC, a hipótese H_0 é rejeitada.

2.3.3 Teste t de Student para Coeficientes da Regressão

De acordo com [6], o teste t de Student na regressão é utilizado para testar a significância dos coeficientes da regressão, ou seja, este procedimento nos permite testar hipóteses sobre os coeficientes de uma regressão. Para testar a hipótese de que um coeficiente qualquer α_1 de uma determinada função de aproximação do MMQ é igual a uma constante α_{10} , as hipóteses são formuladas da seguinte maneira:

Hipótese Nula $H_0: \alpha_1 = \alpha_{10}$

Hipótese Alternativa $H_1: \alpha_1 \neq \alpha_{10}$

Assumimos no teste de hipótese que na nossa amostra de dados os erros ε_i são normalmente distribuídos com média zero e variância populacional desconhecida σ^2 , com os valores y_i seguindo uma distribuição normal. O valor de um coeficiente qualquer α_1 calculado a partir da equação para ele estabelecida pelo MMQ é uma combinação linear normalmente distribuída dos dados do tabelamento, e que apresenta variância $\frac{\sigma^2}{S_{xx}}$, onde S_{xx} representa a soma dos desvios dos valores x_i em relação à sua média $\bar{x}$ e é calculada por $S_{xx} = \sum_{i=1}^m (x_i - \bar{x})^2$.

A partir disso o teste t de Student para o teste de um coeficiente α_1 na regressão pode ser calculado a partir de

$$t_0 = \frac{\alpha_1 - \alpha_{10}}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \quad (14)$$

onde o seu denominador é chamado de erro padrão estimado. Entretanto, como geralmente não se sabe a variância populacional σ^2 do conjunto de dados, é utilizado um estimador não viesado para a variância populacional chamado de média dos quadrados dos resíduos, calculado a partir de

$$MSRes = \frac{SSRes}{m-k} \quad (15)$$

em que o denominador é o grau de liberdade dos resíduos, no qual m é o número de observações ou pontos de um dado tabelamento e k é o número de parâmetros estimados pelo modelo, com o valor $SSRes$ representando a soma dos quadrados dos resíduos, dada por:

$$SSRes = \sum_{i=1}^m [y_i - \varphi(x_i)]^2. \quad (16)$$

O valor desse estimador é necessário para calcular o erro padrão estimado, que é calculado a partir de:

$$se(\alpha_1) = \sqrt{\frac{MSRes}{S_{xx}}}. \quad (17)$$

Consequentemente, a partir da Equação (16) temos que o teste t_0 será obtido a partir de:

$$t_0 = \frac{\alpha_1 - \alpha_{10}}{se(\alpha_1)}. \quad (18)$$

Para tomar uma decisão é comparado o valor observado de t_0 com o quantil superior $\frac{\alpha}{2}$ da distribuição $t_{\frac{\alpha}{2}, m-2}$ ou é utilizado o valor-p.

2.3.4 Teste F de ANOVA para Modelo da Regressão

Segundo [6] o teste F é uma ferramenta estatística usada para avaliar se há uma relação significativa entre a variável independente x e a variável dependente y em um modelo de regressão. Ele permite testar a hipótese nula de que o coeficiente da inclinação da reta α_2 é igual a zero, ou seja, que não há relação linear entre x e y . O objetivo é verificar se a variabilidade explicada pelo modelo é significativamente maior que a variabilidade dos erros ou resíduos, indicando que x tem um impacto significativo sobre y . O teste F compara a variabilidade explicada pelo modelo com a variabilidade não explicada, resultando em uma estatística F.

O teste F na regressão pode ser realizado aplicando a ANOVA, decompondo a variabilidade total dos dados em duas partes. A primeira parte mede o quanto da variação em y é explicado por x , sendo representada pela Equação (17). A segunda parte é a soma dos quadrados explicada pela regressão, que mede o quanto da variação em y é explicado por x e é calculado pela equação:

$$SSR = \sum_{i=1}^m [\varphi(x_i) - \bar{y}]^2. \quad (19)$$

Sendo a identidade fundamental da ANOVA representada por $SST = SSR + SSRes$ em que a soma total dos quadrados, que mede a variabilidade total em y é dada pela equação:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (20)$$

A partir dessa decomposição, teste F avalia as hipóteses:

Hipótese Nula $H_0: \alpha_2 = 0$ (não há relação entre x e y).

Hipótese Alternativa $H_1: \alpha_2 \neq 0$ (há uma relação significativa entre x e y).

Se rejeitarmos H_0 , conclui-se que explica uma parte significativa da variabilidade em y .

A partir dessas hipóteses, a estatística F é calculada a partir da relação entre a variabilidade explicada pelo modelo e a variabilidade dos resíduos. A variabilidade dos resíduos é representada pela Equação (16), enquanto a variabilidade explicada pelo modelo é calculada a partir de:

$$MSR = \frac{SSR}{k-1} \quad (21)$$

em que o denominador é o grau de liberdade da regressão, com k representando o número de parâmetros que estão sendo estimados no modelo.

Utilizando-se das Equações (16) e (21), temos que F é dado por:

$$F = \frac{MSR}{MSRes}. \quad (22)$$

Após o cálculo, o valor de F deve ser comparado com uma tabela de distribuição F . Se o valor for suficientemente grande, ou seu valor-p for menor que o nível de significância estabelecido, a hipótese nula é rejeitada, com o modelo sendo considerado significativo.

3. Metodologia

O objetivo deste trabalho é estudar qual função de aproximação do MMQ melhor se ajusta ao conjunto de pontos referentes à: concentração de CO_2 ao longo do tempo; as anomalias de temperatura ao longo do tempo; as anomalias de temperatura em relação a concentração de CO_2 . As avaliações dos ajustes das funções de aproximação do MMQ serão realizadas a partir dos cálculos do coeficiente de determinação, teste t de Student e teste F ANOVA.

Para este estudo os dados de concentração do CO_2 e das anomalias de temperatura serão obtidos a partir de [8] e [9], respectivamente. A partir disso, os bancos de dados serão filtrados para o período de janeiro de 1980 a dezembro de 2023, sendo combinados em uma única tabela a partir de suas datas correspondentes. Após os dados serem combinados, serão utilizados para o estudo somente a média mensal de concentração de CO_2 , ano, mês, média mensal de anomalia de temperatura, e a data decimal. A data decimal representa o ano e sua fração transcorrida, sendo utilizada no estudo por apresentar maior exatidão em relação ao tempo decorrido mensalmente. Um recorte dos dados pode ser visto a partir da Tabela 1 abaixo:

Ano	Mês	Data Decimal	CO_2 (ppm)	Anomalias de Temperatura
1980	1	1980,042	338,55	0,37
1980	2	1980,125	339,26	0,46
1980	3	1980,208	339,59	0,35
1980	4	1980,292	339,99	0,34
1980	5	1980,375	340,44	0,39
1980	6	1980,458	339,99	0,27

Tabela 1. Tabela de dados do CO_2 e Anomalias de temperatura. Fontes: [8] e [9]

A partir da coleta e tratamento dos dados, serão definidas as variáveis dependentes e independentes de cada problema e construídos os gráficos de dispersão dos pontos, sendo realizado transformações nos dados para linearizar as funções logarítmica, exponencial e potencial. Em seguida, será calculada a função de aproximação $\varphi(x)$, seguida do cálculo do coeficiente de determinação, teste t de Student e do teste F com ANOVA. Após os cálculos serão construídos os gráficos com as funções de aproximação com os pontos tabelados e, finalizando com a análise dos resultados. A sequência de etapas desenvolvidas nesse trabalho pode ser visualizada a partir do fluxograma seguir:

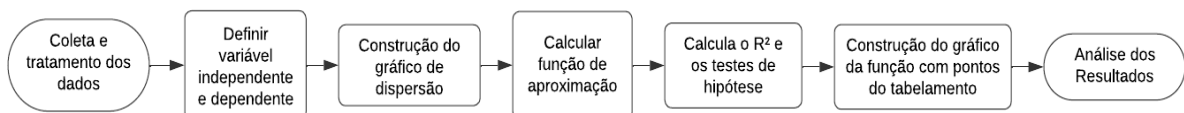


Figura 2. Fluxograma das etapas do estudo. Fonte: autoria própria

As etapas do trabalho serão realizadas computacionalmente utilizando a plataforma Anaconda, versão 24.1.2, e o software Jupyter Notebook, versão 7.0.8, com a linguagem de programação Python versão 3.11.7. Os códigos contarão com o suporte das bibliotecas *pandas*, *statsmodels*, *numpy*, *scipy* e *matplotlib*, em uma máquina com processador Intel Core i5 de 11ª geração, 16 GB de RAM, e sistema operacional Linux Ubuntu 22.04.4 LTS de 64 bits.

De maneira geral, a biblioteca *pandas* será utilizada para a extração, limpeza, segmentação e integração dos dados, enquanto que biblioteca *numpy* irá padronizar as variáveis obtidas pela biblioteca *pandas* e auxiliar na realização das operações necessárias para a linearização das funções de aproximação e a deslinearização dos coeficientes. A biblioteca *statsmodels* será utilizada para calcular os coeficientes das funções de aproximação, o coeficiente de determinação, realizar o teste t de Student para os coeficientes e o teste F com ANOVA. Para gerar os gráficos será utilizada a biblioteca *matplotlib*.

Problema 1: Concentração de $\text{CO}_2 \times$ Tempo

A partir dos dados da concentração do CO_2 ao longo do período de janeiro de 1980 até dezembro de 2023, que foram extraídos de [8] e são visualizados no gráfico de dispersão na Figura 3, queremos saber qual a função de aproximação do MMQ que melhor se ajusta ao conjunto de pontos tabelados, levando em consideração o

coeficiente de determinação, o teste t de Student e o teste F de ANOVA com nível de significância α de 1%.

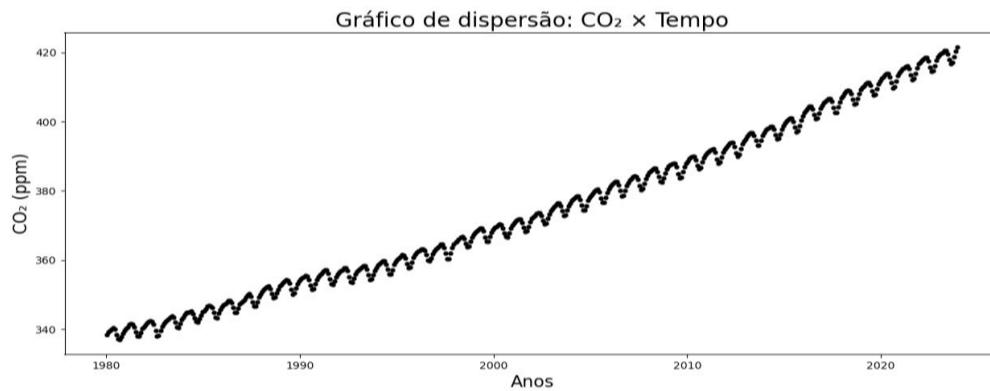


Figura 3 – Dispersão: CO₂ × Tempo. Fonte: Autoria própria

Problema 2: Anomalia de Temperatura × Tempo

A partir dos dados das anomalias de temperatura ao longo do período de janeiro de 1980 até dezembro de 2023, extraídos de [9] e visualizados no gráfico de dispersão na Figura 4, queremos saber qual a função de aproximação do MMQ que melhor se ajusta ao conjunto de pontos tabelados levando em consideração o coeficiente de determinação, e os testes t de Student e F de ANOVA com nível de significância α de 1%.

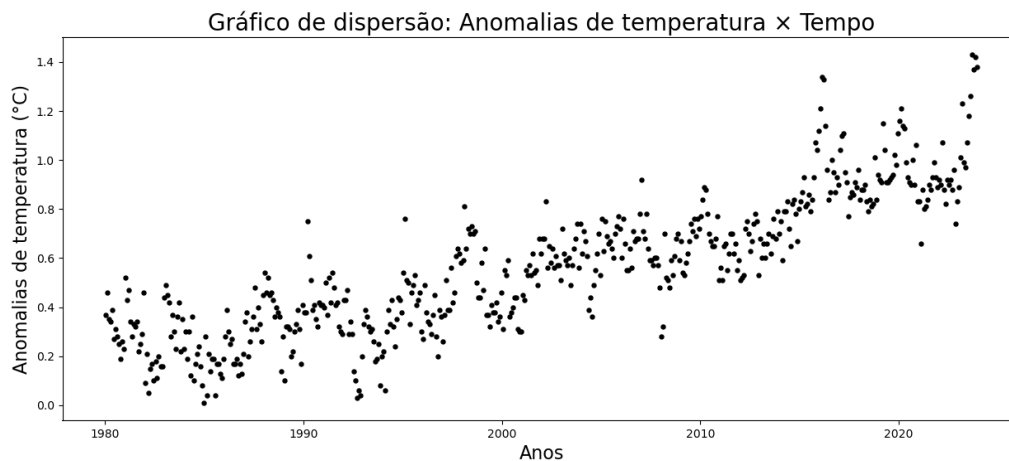


Figura 4 – Dispersão: Anomalias de Temperatura × Tempo. Fonte: Autoria própria

Problema 3: Anomalia de Temperatura × Concentração de CO₂

A partir dos dados de concentração de CO₂ e anomalias de temperatura do período de janeiro de 1980 até dezembro de 2023, extraídos de [8] e [9] respectivamente e visualizados no gráfico de dispersão da Figura 5, queremos saber qual a função de aproximação do MMQ que melhor se ajusta ao conjunto de pontos tabelados levando em consideração o coeficiente de determinação, e os testes t de Student e F de ANOVA com nível de significância α de 1%.

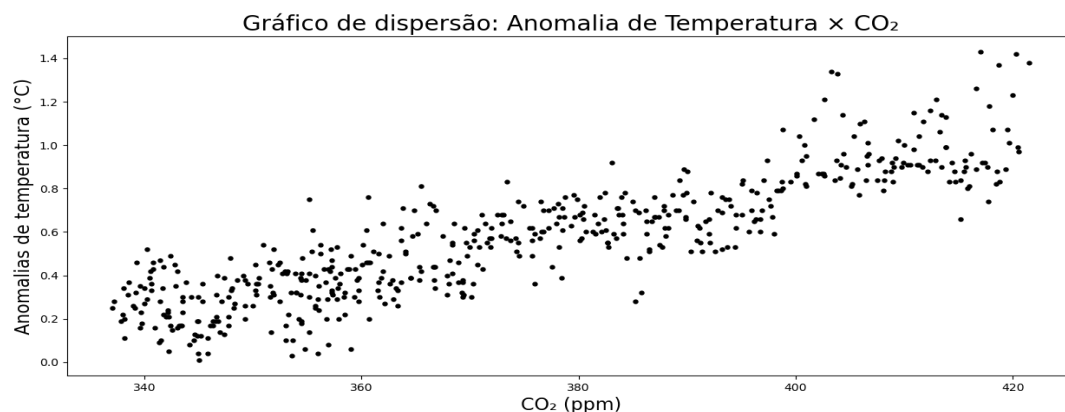


Figura 5 – Dispersão: Anomalias de Temperatura × Concentração de CO₂. Fonte: Autoria própria

4. Resultados e Discussões

Neste capítulo serão apresentados os resultados das funções de aproximação do MMQ com o respectivo coeficiente de determinação, teste t de Student e o teste F de ANOVA, para os três problemas apresentados na seção anterior, com o objetivo de identificar qual das funções de aproximação melhor se ajusta ao seu respectivo tabelamento. Para os testes t de Student e F de ANOVA, será analisado se os resultados de seu valor-p são menores que o nível de significância estabelecido de 1%.

Nas Tabelas 2, 3 e 4 serão apresentadas as funções de aproximação do MMQ calculadas para cada conjunto de pontos, juntamente a seu respectivo coeficiente de determinação R^2 e valor-p dos testes de t de Student e F de ANOVA. O valor-p do teste t de Student para cada coeficiente das funções de aproximação será representado por valor-p(α_1), valor-p(α_2) e valor-p(α_3) enquanto para o teste F de ANOVA está representado por valor-p(F).

Problema 1

Os resultados das funções de aproximação do MMQ com seu respectivo coeficiente de determinação e os testes t de Student e F de ANOVA, para os dados de Concentração de CO_2 ao longo do tempo, estão representados na Tabela 2.

Função de Aproximação	R^2	teste t de Student			teste F de ANOVA
		valor-p(α_1)	valor-p(α_2)	valor-p(α_3)	valor-p(F)
$1,8641x - 3357,2834$	0,986	0,000	0,000	-	0,000
$0,0160x^2 - 62,3537x + 60922,0958$	0,995	0,000	0,000	0,000	0,000
$3730,7570 \ln(x) - 27986,1452$	0,985	0,000	0,000	-	0,000
$0,0182e^{(0,0050x)}$	0,990	0,000	0,000	-	0,000
$(6,2379 \cdot 10^{-31})x^{(9,9283)}$	0,990	0,000	0,000	-	0,000

Tabela 2. Resultados para Concentração de $\text{CO}_2 \times$ Tempo. Fonte: autoria própria

Conforme a Tabela 2, podemos verificar que os resultados para os testes t de Student e F da ANOVA foram reportados como zero devido ao truncamento causado pela precisão da máquina, indicando que esses valores são extremamente próximos de zero, além da capacidade de precisão do software. Esses resultados confirmam a significância de todas as funções de aproximação e todos os seus coeficientes, demonstrando que as funções são apropriadas para estudar os dados, além de que todos os seus coeficientes são relevantes.

Analisando o coeficiente de determinação fica evidente que todas as funções de aproximação tiveram um ajuste satisfatório ao conjunto de dados, visto que os valores de R^2 são muito próximos de 1. Dentre todas as funções, temos que a função polinomial de segundo grau se destacou levemente em relação as demais de acordo com o R^2 , demonstrando ser o modelo mais ajustado ao comportamento desse conjunto de pontos. A função polinomial de segundo grau pode ser observada na Figura 6.

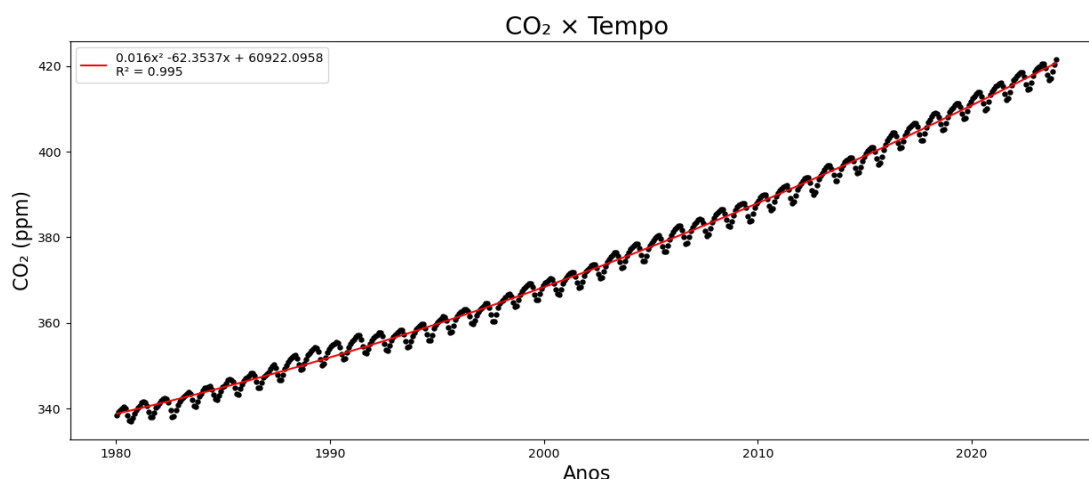


Figura 6 – Função Quadrática para $\text{CO}_2 \times$ Tempo. Fonte: Autoria própria

Problema 2

Os resultados das funções de aproximação do MMQ com seu respectivo coeficiente de determinação e os testes t de Student e F de ANOVA, para o conjunto de pontos das Anomalias de Temperatura ao longo do tempo, estão representados na Tabela 3.

Função de Aproximação	R ²	teste t de Student			teste F de ANOVA
		valor-p(α_1)	valor-p(α_2)	valor-p(α_3)	valor-p(F)
$0,0189x - 37,2769$	0,749	$1,884.10^{-157}$	$5,097.10^{-160}$	-	$5,097.10^{-160}$
$0,000285071x^2 - 1,1225x + 1105,2468$	0,771	$1,841.10^{-11}$	$8,793.10^{-12}$	$4,066.10^{-12}$	$8,928.10^{-169}$
$37,8219\ln(x) - 286,9528$	0,748	$2,385.10^{-159}$	$1,100.10^{-159}$	-	$1,100.10^{-159}$
$(6,7240.10^{-35})e^{(0,0389x)}$	0,599	$7,236.10^{-108}$	$1,382.10^{-106}$	-	$1,382.10^{-106}$
$(2,0352.10^{-258})x^{(77,9578)}$	0,600	$8,839.10^{-107}$	$1,304.10^{-106}$	-	$1,304.10^{-106}$

Tabela 3. Resultados para Anomalias de Temperatura × Tempo. Fonte: autoria própria

Conforme a Tabela 3, podemos verificar que os resultados do teste t de Student e F da ANOVA não sofreram truncamento e ficaram abaixo do nível de significância de 1%. Assim, estes resultados foram satisfatórios e demonstraram a significância de todas as funções de aproximação, levando em consideração todos os seus respectivos coeficientes. Em contrapartida, o coeficiente de determinação não teve um desempenho satisfatório em nenhuma das funções de aproximação, embora os testes t de Student e F de ANOVA se mostraram satisfatórios.

Os resultados do R² para esse problema indicam a necessidade de explorar outras funções de aproximação do MMQ afim de obter uma melhora no ajuste aos dados. Entretanto, entre todas as funções de aproximação abordadas no estudo, a que melhor se adequou ao conjunto de pontos das anomalias de temperatura ao longo do tempo foi a função polinomial de segundo grau. O destaque para a função quadrática pode ser atribuído à inserção do termo quadrático α_3 , uma vez que o maior número de termos em $\varphi(x)$ permite uma melhor aproximação das variações dos dados e melhora seu ajuste, como pode ser observado na Figura 7.

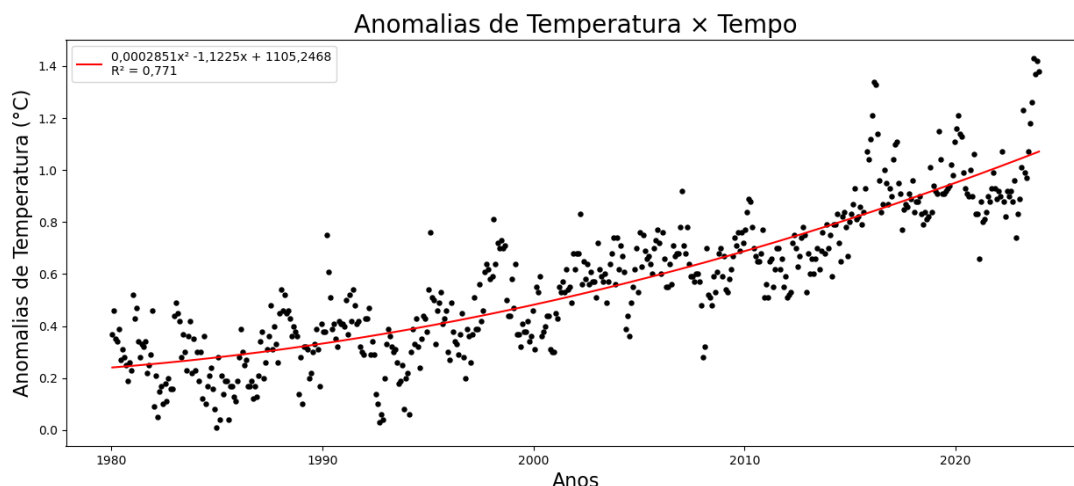


Figura 7 – Função Quadrática para Anomalias de Temperatura × Tempo. Fonte: Autoria própria

Problema 3

Os resultados das funções de aproximação do MMQ com seu respectivo coeficiente de determinação e os testes t de Student e F de ANOVA para o conjunto de pontos das Anomalias de Temperatura em função da Concentração de CO₂, estão representados na Tabela 4.

Função de Aproximação	R ²	teste t de Student			teste F de ANOVA
		valor-p(α_1)	valor-p(α_2)	valor-p(α_3)	valor-p(F)
$0,0102x - 3,2626$	0,771	$3,029.10^{-143}$	$1,251.10^{-170}$	-	$1,251.10^{-170}$
$(2,7631.10^{-5})x^2 - 0,0107x + (0,6641)$	0,774	0,678	0,210	0,014	$3,275.10^{-170}$
$3,8401\ln(x) - 22,1836$	0,768	$1,227.10^{-164}$	$4,923.10^{-169}$	-	$4,923.10^{-169}$
$0,0002e^{(0,0207x)}$	0,596	$4,302.10^{-118}$	$1,162.10^{-105}$	-	$1,162.10^{-105}$
$(3,6320.10^{-121})x^{(7,8215)}$	0,601	$3,683.10^{-109}$	$5,300.10^{-107}$	-	$5,300.10^{-107}$

Tabela 4. Resultados para Anomalias de Temperatura × CO₂. Fonte: autoria própria

De acordo com a Tabela 4, é possível observar que para o teste t de Student, os coeficientes de todas as funções de aproximação foram estatisticamente significativos, com exceção da função polinomial de segundo grau que apresentou os três coeficientes não significativos. Entretanto, para o teste F de ANOVA, todas as funções de

aproximação apresentaram significância estatística, incluindo a função polinomial de segundo grau. Analisando somente o coeficiente de determinação, todas apresentaram valores considerados insatisfatório, porém as funções que tiveram melhor desempenho foi a polinomial de segundo grau, seguido da função linear e logarítmica.

Os resultados dos testes t de Student e F de ANOVA para a função quadrática podem indicar um problema chamado multicolinearidade, que é uma forte correlação dos termos da função que torna difícil de identificar a contribuição individual de cada variável. Entretanto, mesmo que os testes de significância de cada coeficiente individual indiquem que esses termos sozinhos não são estatisticamente relevantes, o valor-p de α_3 apresentou um valor próximo ao nível de significância estabelecido, sugerindo que sua presença e contribuição é suficiente para um maior valor de R^2 e para a significância do teste F de ANOVA, respectivamente. Essa função pode ser observada na Figura 8.

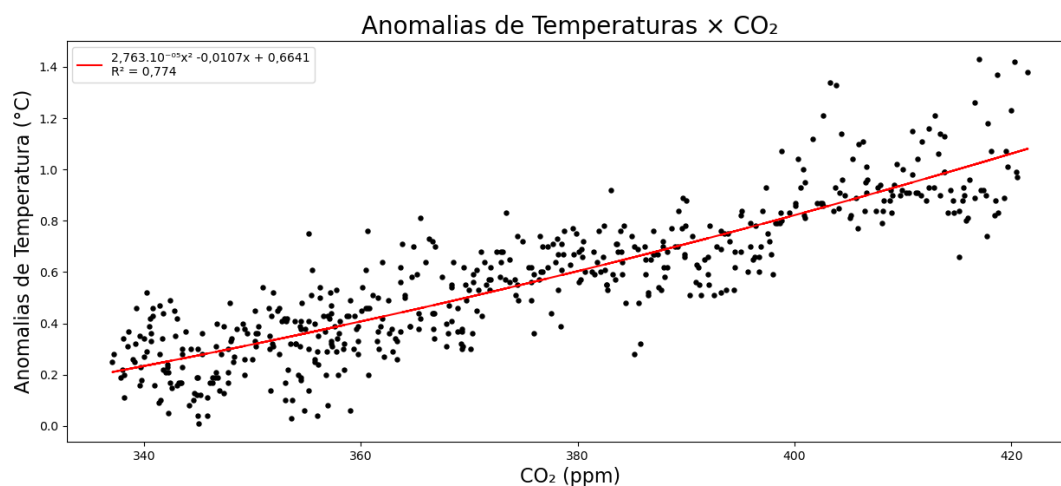


Figura 8 – Função Quadrática para Anomalias de Temperatura $\times$ CO_2 . Fonte: Autoria própria

Os resultados para os Problemas 1, 2 e 3 demonstraram grande disparidade, especialmente devido a natureza de cada problema e aos diferentes métodos utilizados para avaliá-los. O R^2 foca na capacidade explicativa geral do modelo e seus resultados para as funções dos Problemas 2 e 3, mostram limitações nas tentativas de se adequar a variabilidade dos dados. Por outro lado, o teste t de Student revelou a contribuição de cada variável individualmente, destacando que certas funções poderiam estar incluindo termos irrelevantes, como observado somente nos coeficientes da função quadrática do Problema 3. Em contrapartida, independente do resultado do teste t e do coeficiente de determinação, o teste F de ANOVA indicou a significância estatística para todas as funções de aproximação de todos os problemas, uma vez que seu foco de análise é a contribuição conjunta de todos os coeficientes da função para explicar a variabilidade dos dados.

5. Conclusão

O presente estudo realizou a análise de dados históricos de concentração de CO_2 e anomalias de temperaturas em diferentes contextos e a partir da aplicação do MMQ com diferentes funções de aproximação foi possível analisar qual das funções de aproximação melhor se ajustou ao conjunto de dados em cada problema, avaliando a qualidade do ajuste por meio do coeficiente de determinação, teste t de Student para a significância dos coeficientes das funções e teste F de ANOVA para significância das funções. Os resultados do estudo indicaram a significância estatística de diferentes funções de aproximação do MMQ e de seus coeficientes, demonstrando a aplicabilidade do método para todos os casos estudados.

Para a concentração de CO_2 os resultados obtidos foram considerados satisfatórios e demonstraram uma excelente forma de analisar esses dados. Entretanto, o estudo mostrou a partir da análise do coeficiente de determinação os resultados se mostraram pouco satisfatórios aos casos que envolviam dados de anomalias de temperatura, o que indica uma baixa capacidade de explicação das funções de aproximação para as variações nesses dados, sendo recomendado a aplicação de mais funções de aproximação do MMQ, além de técnicas de análise de séries temporais afim de analisar esses dados por uma abordagem que forneça resultados melhores.

Os resultados obtidos demonstram também uma tendência de crescimento a partir de uma função quadrática crescente tanto para a concentração de CO_2 como para as Anomalias de Temperatura ao longo do tempo. Essa tendência revela uma intensificação das mudanças climáticas, configurando um cenário preocupante para os próximos anos caso essa curva de crescimento permaneça. Portanto, é urgente e fundamental a adoção medidas que ajudem a reverter essa tendência de crescimento de forma a diminuir as consequências dessas mudanças.

Por fim, a execução deste trabalho reforça a relevância da análise de dados climáticos e atmosféricos no atual cenário de mudanças climáticas. Compreender melhor o comportamento dessas variáveis é essencial para a busca de soluções que mitiguem os impactos das alterações no clima global.

6. Referências Bibliográficas

- [1] CORTESE, Tatiana Tucunduva P.; NATALINI, Gilberto. *Mudanças Climáticas: Do Global ao Local*. Barueri: Editora Manole, 2014. E-book. ISBN 9788520446607. Disponível em: <<https://app.minhabiblioteca.com.br/#/books/9788520446607/>>. Acesso em: 27 jul. 2024.
- [2] SILVA, Cleyton Martins da; ARBILLA, Graciela. *Emissões atmosféricas e mudanças climáticas*. 1. ed. Rio de Janeiro: Freitas Bastos, 2022. E-book. Disponível em: <<https://plataforma.bvirtual.com.br>>. Acesso em: 24 jul. 2024.
- [3] FRANCO, Neide B. *Cálculo Numérico*. São Paulo: Pearson Prentice Hall, 2006.
- [4] RUGGIERO, Marcia A. G.; LOPES, Vera L. R. *Cálculo numérico: aspectos teóricos e computacionais*. 2. ed. São Paulo: Pearson Education do Brasil, 2013.
- [5] ARENALES, S.; DAREZZO, A. *Cálculo Numérico: Aprendizado com apoio de software*. 2. ed. São Paulo: Cengage Learning, 2015.
- [6] MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. *Introduction to Linear Regression Analysis*. 5. ed. Hoboken: Wiley, 2012.
- [7] MORETTIN, Pedro A.; BUSSAB, Wilton de O. *Estatística básica*. São Paulo: SRV Editora LTDA, 2017. E-book. ISBN 9788547220228. Disponível em: <<https://app.minhabiblioteca.com.br/#/books/9788547220228/>>. Acesso em: 30 jul. 2024.
- [8] Global Monitoring Laboratory – Carbon Cycle Greenhouse Gases. Disponível em: <gml.noaa.gov/webdata/ccgg/trends/co2/co2_mm_gl.txt>. Acesso em: 27 ago. 2024.
- [9] Climate at a Glance | Global Time Series | National Centers for Environmental Information (NCEI). Disponível em: <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/global/time-series/globe/land_ocean/1/0/1979-2024>. Acesso em: 27 ago. 2024.