University of Leeds

MATH5747M: Learning Skills Through Case Studies

School of Mathematics

Investigating the Relationship Between Atmospheric CO_2 Concentrations and Global Temperature Anomalies

Author:
Matthew Gordon

S.I.D: [redacted]
April 11, 2025

Contents

1 Introduction 1

2 Literature Review 1

3 Data and Methodology 2
- 3.1 Global Temperature Anomalies (NASA GISTEMP) . . . 2
- 3.2 Atmospheric CO_2 Concentrations (NOAA Mauna Loa Observatory) . . . 3
- 3.3 Methods and Justifications . . . 3
  - 3.3.1 Scatter Plot with Regression Line . . . 3
  - 3.3.2 Pearson and Spearman Correlation Coefficients . . . 3
  - 3.3.3 Simple Linear Regression Model . . . 4
  - 3.3.4 Residual Diagnostics . . . 4
  - 3.3.5 Standardised Time Series Plot . . . 4
  - 3.3.6 Cross-Correlation Function (CCF) . . . 4

4 Results 5
- 4.1 Correlation Analysis . . . 5
- 4.2 Regression Modelling . . . 6
- 4.3 Standardised Time Series . . . 6
- 4.4 Cross-Correlation Function . . . 7
- 4.5 Residual Analysis . . . 8
  - 4.5.1 Histogram of Residuals . . . 9
  - 4.5.2 Q–Q Plot . . . 9
  - 4.5.3 Residuals vs Fitted Values . . . 10
  - 4.5.4 Formal Statistical Tests . . . 10

5 Discussion 11
- 5.1 Interpretation of Results . . . 11
  - 5.1.1 Correlation Analysis . . . 11
  - 5.1.2 Regression Model . . . 11
  - 5.1.3 Standardised Time Series . . . 11
  - 5.1.4 Cross-Correlation Function . . . 12
  - 5.1.5 Residual Analysis . . . 12
- 5.2 Limitations and Assumptions . . . 13
- 5.3 Future Work . . . 13
- 5.4 Summary and Concluding Remarks . . . 14

References 15

1 Introduction

The Earth's climate is undergoing rapid change, with global temperatures rising at a pace unprecedented in the historical record. Since the late 19th century, the average global surface temperature has increased by more than 1 °C, with most of this warming occurring in the last 50 years [5]. This trend is widely attributed to the enhanced greenhouse effect — a process in which certain atmospheric gases, such as carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), trap outgoing infrared radiation and re-emit it toward the Earth's surface. While the greenhouse effect is a natural and necessary phenomenon for maintaining a habitable planet, the dramatic increase in anthropogenic greenhouse gas emissions, especially CO_2, has intensified this process, contributing to the current era of global warming [3].

Carbon dioxide is of particular concern due to its long atmospheric lifetime and its strong radiative forcing potential. Measurements from the Mauna Loa Observatory in Hawaii — the longest continuous record of atmospheric CO_2 — show a rise from approximately 316 ppm in 1959 to over 420 ppm in recent years [6]. This increase closely mirrors the post-industrial acceleration in fossil fuel combustion, deforestation, and cement production. Concurrently, global temperature anomalies have also risen, raising questions about the extent to which rising CO_2 levels directly influence global temperatures.

Understanding the relationship between atmospheric CO_2 and global temperature anomalies is critical not only for affirming the physical science basis of climate change but also for informing future policy decisions and climate projections. Although climate models and physical simulations provide robust predictions, accessible data analysis using real-world observations offers valuable insights — particularly when transparent, reproducible statistical methods are used. This is especially important for bridging the gap between the scientific community and the public or policy spheres, where trust in data and methods plays a central role in shaping climate response.

In this report, we use two modern, high-quality datasets — global temperature anomalies from NASA's Goddard Institute for Space Studies (GISTEMP) and annual mean atmospheric CO_2 concentrations from NOAA's Mauna Loa Observatory — to investigate the statistical relationship between CO_2 and temperature from 1959 to the present. The aim is to quantify this relationship using a combination of correlation analysis, linear regression modelling, time series techniques and cross-correlation analysis. Additionally, diagnostic tests are performed to ensure model validity, including assessments of residual normality and homoscedasticity. While previous research has widely explored the CO_2–temperature link, this study contributes a focused, data-driven investigation using simple, transparent methods to reinforce and communicate key climate science findings.

2 Literature Review

Scientific consensus strongly supports the role of atmospheric carbon dioxide (CO_2) as a primary driver of recent global warming. CO_2 is a greenhouse gas that absorbs and re-emits infrared radiation, effectively trapping heat within the Earth's atmosphere. This mechanism, first proposed in the 19th century and later quantified by [1], has since become foundational to modern climate science. According to the Intergovernmental Panel on Climate Change (IPCC), the sharp increase in anthropogenic CO_2 emissions since the Industrial Revolution is the dominant cause of the observed rise in global surface temperatures [3].

Numerous empirical studies have confirmed a strong correlation between atmospheric CO_2 levels and global temperatures over both short and long timescales. For example, studies using ice core data reveal that historical CO_2 concentrations and temperature anomalies have co-varied over glacial cycles [4], while modern instrumental datasets show an accelerating trend in both variables over the past century. NASA's Goddard Institute for Space Studies (GISS) and NOAA's Mauna Loa Observatory provide reliable, long-term records of global temperature anomalies and atmospheric CO_2 concentrations, respectively. These datasets serve as the empirical basis for much of the recent work investigating climate trends.

While the physical mechanism linking CO_2 and temperature is well understood, statistical analysis provides a critical layer of validation. Studies such as [7] and [2] have used regression models and correlation analysis to quantify the relationship, typically finding strong positive associations. However, establishing causality remains challenging due to the presence of confounding natural variability, including oceanic cycles (e.g., El Niño–Southern Oscillation) and volcanic activity, which can obscure short-term trends.

Cross-correlation methods and residual diagnostics have increasingly been used to investigate the temporal structure of the CO_2–temperature relationship and test the validity of regression models. These techniques help evaluate whether CO_2 leads temperature changes (indicating a causal influence), or whether feedback loops, such as ocean outgassing, also contribute to the observed co-variation. Such methodological approaches complement the physical understanding of the carbon cycle by quantifying statistical relationships in observed data.

While the statistical relationship between atmospheric CO_2 concentrations and global temperature anomalies has been widely studied, much of the existing literature either focuses on large-scale multi-factor climate models or spans long paleoclimate timescales. In contrast, this report conducts a focused, accessible statistical investigation using modern post-1959 instrumental data from NASA and the NOAA. By applying a combination of linear modelling, correlation analysis, and cross-correlation techniques, this study aims to independently verify the strength and direction of the CO_2–temperature relationship using transparent, reproducible methods. This approach not only reinforces scientific consensus but also demonstrates how basic statistical techniques can effectively quantify climate trends using publicly available data.

3 Data and Methodology

In this section, we analyse the reliability of the two datasets used throughout this report: NASA's GISTEMP dataset for global temperature anomalies and NOAA's Mauna Loa Observatory data for atmospheric CO_2 concentrations. Both are publicly available and widely recognised for their scientific credibility.

3.1 Global Temperature Anomalies (NASA GISTEMP)

The global temperature anomaly dataset used in this study was obtained from NASA's Goddard Institute for Space Studies (GISTEMP). This dataset provides annual global surface temperature anomalies, expressed in degrees Celsius (°C), relative to a 1951–1980 baseline period. The full dataset spans the period from 1880 to 2016 and is widely regarded as a reliable source for long-term temperature trends due to its combination of land-based weather station records and sea surface temperature data. The GISTEMP dataset is regularly updated, peer-reviewed, and publicly accessible via NASA's climate data portal [5].

For the purposes of this study, the dataset was restricted to the years 1959 to 2016 in order to align with the temporal coverage of the atmospheric CO_2 dataset, enabling year-on-year comparisons. The filtered dataset contained no missing values, and each entry consisted of two columns: `Year` and `Temperature_Anomaly`, with the anomaly values expressed relative to the long-term climatological baseline.

3.2 Atmospheric CO_2 Concentrations (NOAA Mauna Loa Observatory)

Atmospheric carbon dioxide (CO_2) data was obtained from the NOAA Global Monitoring Laboratory, specifically from the Mauna Loa Observatory in Hawaii. While NOAA collects daily CO_2 measurements at this site, they also publish an annual dataset in which the mean CO_2 concentration is calculated from the daily observations for each calendar year. As the temperature anomaly data is annual, this pre-aggregated annual dataset was used to ensure temporal alignment.

Mauna Loa was chosen as the monitoring site due to its remote location and minimal local contamination, making it one of the most consistent and uncontaminated CO_2 records available [6]. The selected data spanned from 1959 to 2024; however, to match the temporal window of the temperature anomaly dataset, only values from 1959 to 2016 were used in the final analysis. No missing values were present in this range.

The dataset consisted of two columns: **Year** and **CO2**, representing the annual average CO_2 concentration for each year. By aligning the two datasets over a shared time period (1959–2016) and confirming complete data coverage, a robust foundation was established for statistical comparison between atmospheric CO_2 levels and global temperature anomalies.

3.3 Methods and Justifications

To investigate the relationship between atmospheric CO_2 concentrations and global temperature anomalies, a suite of statistical and time series methods were employed. These methods were selected to explore both the strength and nature of the association between the two variables, as well as to assess whether the observed relationship is consistent with a direct or temporally lagged effect. This section outlines each method and provides a rationale for its use within the context of the research question.

3.3.1 Scatter Plot with Regression Line

A scatter plot was produced to visually assess the relationship between global temperature anomalies and atmospheric CO_2 concentrations. This method provides an intuitive first step in exploring whether a linear pattern exists between the two variables. A line of best fit was overlaid using simple linear regression to illustrate the overall trend, helping to visually confirm the appropriateness of subsequent correlation and regression modelling.

3.3.2 Pearson and Spearman Correlation Coefficients

To quantify the strength of the relationship between the two variables, both Pearson and Spearman correlation coefficients were calculated. Pearson's coefficient measures the degree of linear association, assuming both variables are continuous and normally distributed. In contrast, Spearman's rank correlation assesses monotonic relationships without requiring linearity or normality. Using both metrics allowed for a more robust evaluation of the association, helping to confirm whether temperature anomalies consistently increase as CO_2 levels rise, regardless of the exact shape of the relationship.

3.3.3 Simple Linear Regression Model

A simple linear regression model was fitted with CO_2 concentration as the independent variable and temperature anomaly as the dependent variable. This approach enabled formal quantification of the relationship, specifically estimating how much global temperature anomalies increase for each unit rise in CO_2. The model also provided an R^2 value to indicate the proportion of variability in temperature explained by CO_2, supporting a statistical basis for causality.

3.3.4 Residual Diagnostics

To validate the assumptions of the linear regression model, residual diagnostics were performed. A histogram and Q–Q plot of the residuals were used to assess normality, while the Shapiro–Wilk test provided a formal statistical test. Homoscedasticity (constant variance of residuals) was evaluated using a residuals vs. fitted values plot and was formally measured using a studentised Breusch-Pagan test. These checks ensure that the conclusions drawn from the regression model are statistically valid, particularly regarding p-values. Validating these assumptions enhances the reliability of the model's interpretation in the context of this study.

3.3.5 Standardised Time Series Plot

Both variables were standardised using z-score normalisation to produce a dual time series plot on a common scale. This method allowed direct visual comparison of the temporal trends in CO_2 and temperature anomalies, helping to identify whether their patterns evolved in tandem over the study period. By removing the influence of differing units and magnitudes, standardisation provided a clear way to assess co-movement between the variables across time.

3.3.6 Cross-Correlation Function (CCF)

The cross-correlation function was applied to assess the lead-lag relationship between atmospheric CO_2 and temperature anomalies. The CCF calculates correlation coefficients at a range of lags, allowing us to determine whether CO_2 changes tend to precede or follow changes in temperature. This method is particularly useful for identifying temporal structure in time series data and probing whether CO_2 might act as a causal driver of temperature changes, or whether the relationship is more complex.

Collectively, these methods provide a comprehensive framework for analysing the relationship between atmospheric CO_2 and global temperature anomalies. By combining visual exploration, correlation analysis, predictive modelling, and diagnostic testing, the analysis is designed to be both statistically rigorous and accessible. The following section presents the results obtained from each method and forms the basis for subsequent interpretation and discussion.

4 Results

This section presents the outputs of the statistical analyses described in the methodology. Visualisations and summary statistics are provided to support each method used, including scatter plots, correlation coefficients, regression results, standardised time series, and cross-correlation outputs.

4.1 Correlation Analysis

To explore the relationship between atmospheric CO_2 concentrations and global temperature anomalies, a scatter plot was produced with a fitted linear regression line. The plot, shown in Figure 1, reveals the overall trend and spatial distribution of the data points over the observed period.

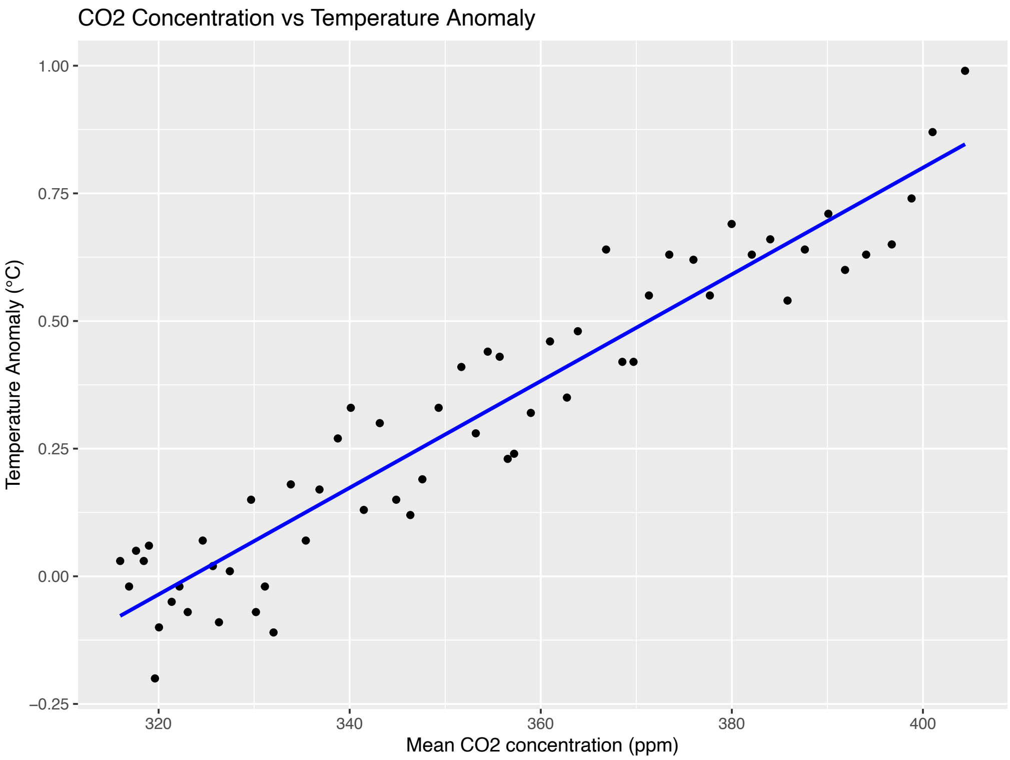

Figure 1: Scatter plot of atmospheric CO_2 concentration (ppm) versus global temperature anomaly (°C) with fitted regression line.

In addition to the visual representation, both the Pearson and Spearman correlation coefficients were computed to quantify the strength and direction of the relationship. The **Pearson correlation coefficient** was found to be **0.948**, indicating a very strong positive linear association between the two variables. The **Spearman rank correlation coefficient** was calculated as **0.931**, supporting a strong monotonic relationship, independent of any assumptions about linearity.

These results suggest a high degree of agreement between the variables across both correlation measures.

4.2 Regression Modelling

A simple linear regression model was fitted with atmospheric CO_2 concentration (ppm) as the independent variable and global temperature anomaly (°C) as the dependent variable. The model aimed to estimate the extent to which changes in CO_2 levels could explain variations in temperature anomalies. The regression results are summarised in Table 1.

Table 1: Summary of Linear Regression Results

Metric	**Value**
Intercept	-3.379
CO_2 Coefficient	0.01045
CO_2 p-value	$< 2e{-}16$
Multiple R-squared	0.8979
Adjusted R-squared	0.8961
Residual Standard Error	0.0932
F-statistic	492.6 ($p < 2.2e{-}16$)

The model produced a statistically significant positive coefficient for atmospheric CO_2, with an estimated increase of 0.01045 °C in global temperature anomaly for every 1 ppm increase in CO_2 concentration. The p-value associated with the CO_2 term was less than 2e-16, indicating that the relationship is highly significant. The model achieved a Multiple R-squared value of 0.8979, suggesting that approximately 89.8% of the variance in temperature anomalies can be explained by changes in atmospheric CO_2 levels. The residual standard error was 0.0932 °C, indicating relatively small deviations from the fitted line.

4.3 Standardised Time Series

Figure 2 displays the standardised time series of atmospheric CO_2 concentrations and global temperature anomalies from 1959 to 2016. Both variables were normalised using z-score standardisation to allow for direct comparison on a common scale.

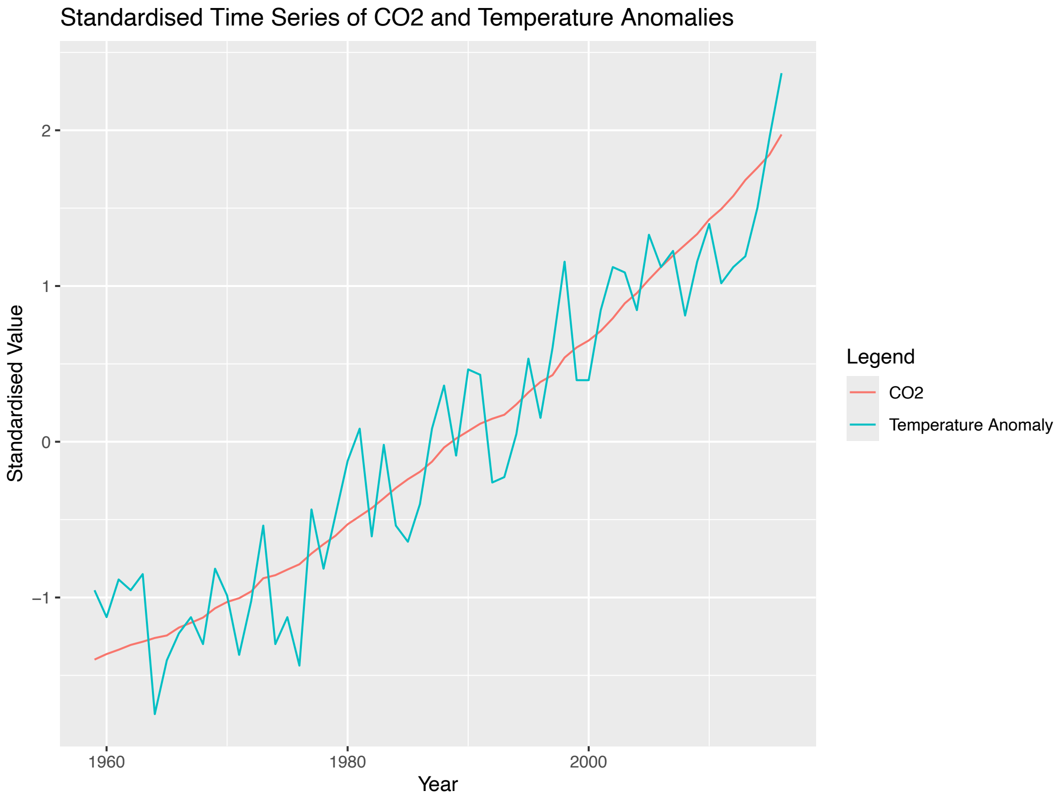

Figure 2: Standardised time series of atmospheric CO_2 concentrations and global temperature anomalies (1959–2016). CO_2 shown in red; temperature anomaly in blue.

The plot reveals that both CO_2 concentrations and temperature anomalies have followed a broadly similar upward trajectory over the six-decade period. The CO_2 series rises smoothly and almost linearly, reflecting the steady increase in emissions from anthropogenic sources. In contrast, the temperature anomaly series exhibits greater short-term variability, likely influenced by natural fluctuations such as El Niño–Southern Oscillation events and volcanic activity. Despite these year-to-year variations, the overall shape of the temperature trend closely aligns with the CO_2 trend, particularly from the mid-1970s onward.

4.4 Cross-Correlation Function

To assess the potential lead–lag relationship between atmospheric CO_2 concentrations and global temperature anomalies, a cross-correlation function (CCF) was computed. This method evaluates the correlation between the two time series across a range of lags. The resulting plot is shown in Figure 3.

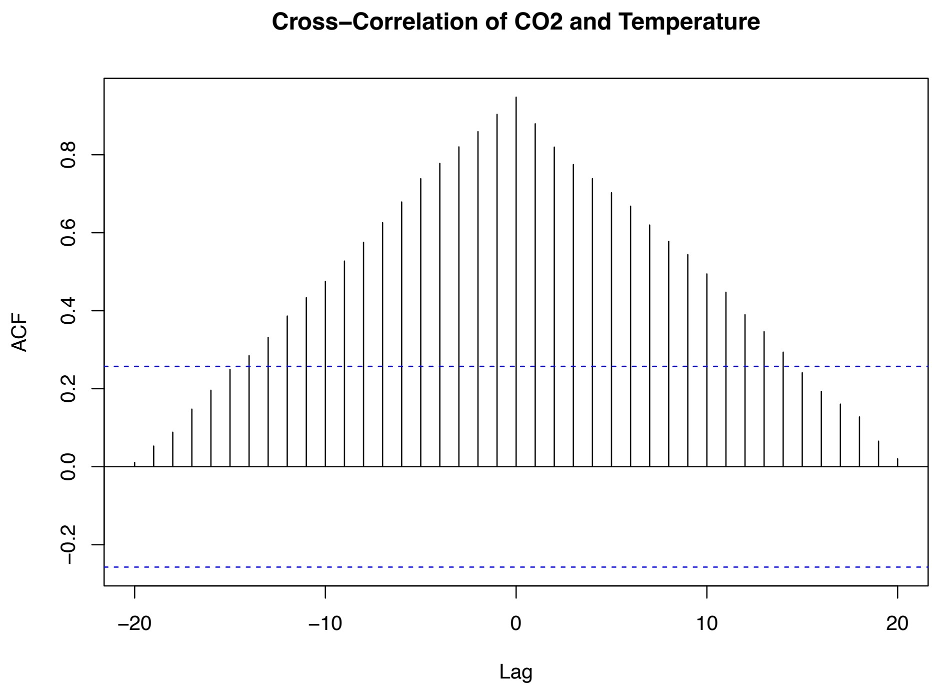

Figure 3: Cross-correlation function (CCF) between atmospheric CO_2 and global temperature anomalies. Lag 0 represents simultaneous correlation; positive lags indicate CO_2 leads temperature.

The maximum correlation occurred at lag zero, indicating that CO_2 concentrations and temperature anomalies are most strongly aligned in the same year. Correlations remained statistically significant and relatively high for a range of small positive and negative lags, suggesting that increases in CO_2 could plausibly precede short-term increases in global temperature and vice-versa. The correlation structure around lag zero appeared largely symmetric, with negative lags displaying small systematic dominance. All correlations near lag zero exceeded the significance bounds, confirming that the observed relationships are statistically significant and unlikely to have occurred by chance.

4.5 Residual Analysis

To evaluate the validity of the linear regression model, diagnostic checks were conducted on the residuals. These checks assess whether the model satisfies the key assumptions of linearity, normality, and homoscedasticity — all of which are essential for ensuring that the model's estimates and statistical inference are reliable. This section presents and discusses three separate diagnostic plots: a histogram of residuals, a normal Q–Q plot, and a residuals versus fitted values plot. In addition to these visual diagnostics, formal statistical tests were conducted, including the Shapiro–Wilk test for normality and the studentised Breusch–Pagan test for homoscedasticity, to provide further validation of model assumptions.

4.5.1 Histogram of Residuals

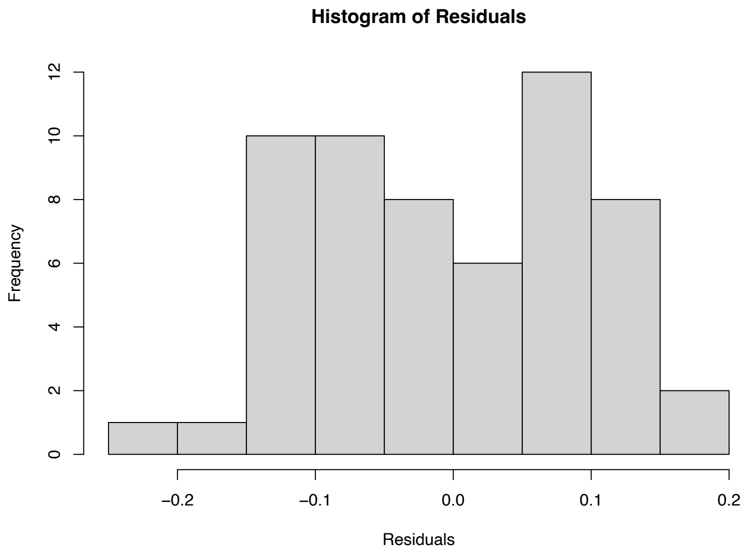

Figure 4: Histogram of residuals from the linear regression model.

The histogram of residuals appears reasonably symmetric and bell-shaped, centred around zero. Most residuals fall within the range of approximately -0.2 to 0.2, suggesting that the residuals are approximately normally distributed. There is no evidence of strong skewness or outliers. This supports the assumption of normality, and the centred distribution indicates that the model does not consistently over or under-predict.

4.5.2 Q–Q Plot

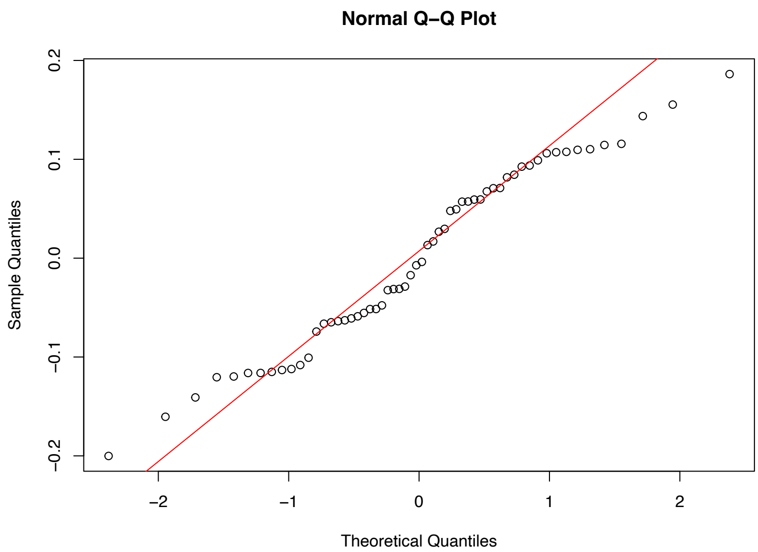

Figure 5: Normal Q–Q plot of residuals from the regression model.

The Q–Q plot shows that most residuals lie close to the theoretical 45-degree reference line, especially in the central region. This indicates that the central residuals follow a normal distribution closely. Slight deviations are observed at both tails, where the points curve away from the line, suggesting mild departures from normality in the extremes. However, these deviations are minor and are unlikely to substantially affect inference in a model of this scale.

4.5.3 Residuals vs Fitted Values

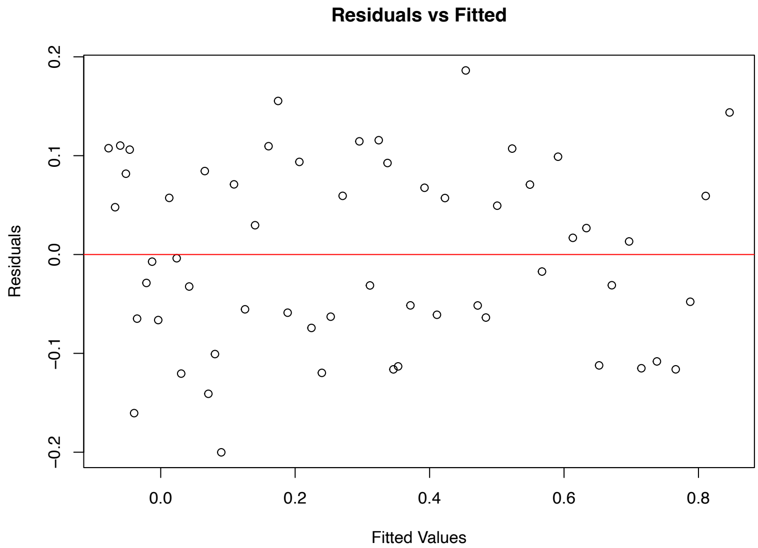

Figure 6: Residuals vs fitted values plot from the regression model.

The residuals vs fitted values plot reveals a random scatter of residuals around the zero line, with no apparent structure, curvature, or trend. This randomness suggests that the linearity assumption holds. The variance of residuals appears relatively constant across the range of fitted values, with no evident funnel shape or clustering. There may be a slight narrowing of spread at higher fitted values, but not to an extent that indicates heteroscedasticity. No extreme outliers are visible.

4.5.4 Formal Statistical Tests

In addition to visual diagnostic plots, two formal statistical tests were conducted to evaluate the assumptions of normality and homoscedasticity of the regression residuals.

The **Shapiro–Wilk test** was used to assess whether the residuals were approximately normally distributed. This test compares the order statistics of the sample to those expected under a normal distribution. The null hypothesis is that the data are drawn from a normal distribution. For this model, the test yielded a W-statistic of 0.9638 and a p-value of 0.0812. Since the p-value exceeds the conventional 0.05 threshold, there is insufficient evidence to reject the null hypothesis. This supports the assumption of normality, aligning with the visual findings from the histogram and Q–Q plot.

To assess the assumption of homoscedasticity (constant variance of residuals), the **studentised Breusch–Pagan test** was performed. This test evaluates whether the variance of the residuals depends on the fitted values of the model. The null hypothesis is that the residuals exhibit constant variance (i.e., homoscedasticity). The test returned a test statistic of 0.2232 with a p-value of 0.6366. As the p-value is well above 0.05, there is no significant evidence to suggest heteroscedasticity is present in the model.

Both statistical tests confirm the visual assessments and indicate that the residuals satisfy the key assumptions of linear regression to an acceptable degree.

5 Discussion

This section reflects on the findings presented in the results, with a focus on interpreting their implications in the context of the research question: whether atmospheric CO_2 concentrations are a direct influencer of global temperature anomalies. The preceding analyses consistently indicated a strong positive association between the two variables. However, it is important to consider these results not only in statistical terms, but also in light of the broader physical mechanisms and potential limitations that may influence the observed relationships.

The following subsection provides a detailed interpretation of the results and examines possible explanations for their patterns, including known climate feedback mechanisms. Subsequent parts of this discussion evaluate the strengths and limitations of the methods used, assess the robustness of the conclusions, and propose directions for future work.

5.1 Interpretation of Results

This section provides a critical interpretation of the results presented previously. Each analytical method is revisited with a view to understanding its implications, evaluating consistency across findings, and identifying potential sources of bias or alternative explanations.

5.1.1 Correlation Analysis

The high Pearson and Spearman correlation coefficients (**0.948** and **0.931**, respectively) provide strong evidence of a consistent, positive relationship between atmospheric CO_2 and global temperature anomalies. The agreement between both metrics reinforces the validity of the relationship, whether assessed linearly or more broadly as a monotonic trend.

5.1.2 Regression Model

The simple linear regression model yielded an R^2 value of **0.8979**, indicating that approximately 89.8% of the variation in temperature anomalies can be explained by CO_2 concentrations alone. The highly significant coefficient $< 2e{-}16$ confirms the strength of this relationship. While the intercept has no meaningful physical interpretation (as $CO_2 = 0$ is not realistic), the slope of 0.01045 suggests that even modest increases in CO_2 levels result in measurable warming.

5.1.3 Standardised Time Series

The standardised time series plot shows that both CO_2 and temperature anomalies have risen in broadly similar trajectories since 1959. The CO_2 trend is smooth and nearly linear, consistent with steady anthropogenic emissions. Temperature anomalies exhibit more short-term variability, but their long-term rise aligns closely with CO_2, particularly from the mid-1970s onwards. This reinforces the overall strength of the observed relationship, despite the presence of natural variability.

5.1.4 Cross-Correlation Function

The cross-correlation function (CCF) analysis revealed that the strongest correlation between atmospheric CO_2 and global temperature anomalies occurred at lag zero. This indicates that the two variables rise simultaneously within the dataset, suggesting a strong real-time association. Importantly, significant correlations were also observed at small positive lags, where changes in CO_2 precede changes in temperature. This supports the hypothesis that rising atmospheric CO_2 may contribute to subsequent increases in global temperatures over short time periods.

Conversely, slightly stronger correlations at small negative lags suggest that temperature anomalies may also influence atmospheric CO_2 concentrations. This implies the potential presence of climate feedback mechanisms, whereby rising temperatures stimulate natural processes that release additional greenhouse gases into the atmosphere.

One such mechanism is **ocean outgassing** — the process by which warmer ocean waters release previously absorbed CO_2. As the solubility of gases decreases with temperature, a warming ocean can no longer retain as much dissolved CO_2, leading to its release into the atmosphere. This can be likened to a can of fizzy drink left out on a hot day: as the drink warms, it fizzes more, releasing the carbon dioxide previously held in solution. Similarly, rising ocean temperatures reduce the ocean's ability to act as a carbon sink, resulting in increased atmospheric CO_2 levels.

Another independent process is **permafrost thawing** — the degradation of ancient frozen tundra found in Arctic and subArctic regions. As temperatures increase, permafrost thaws and releases carbon that has been trapped in frozen soil for thousands of years. This carbon, often in the form of CO_2 and methane (CH_4), was preserved in undecomposed organic matter and becomes available for atmospheric release once thawed, contributing to further warming.

A third feedback mechanism is **increased microbial activity in soils**. As temperatures rise, microbial metabolism accelerates, leading to faster decomposition of organic matter already present in active soil layers. This decomposition releases additional greenhouse gases, primarily CO_2, into the atmosphere. Unlike permafrost thawing, this mechanism involves current rather than ancient carbon, and is driven by temperature-enhanced biological activity rather than the release of long-stored material.

Each of these mechanisms is distinct, yet all contribute to the amplification of atmospheric CO_2 levels in response to warming. Their presence helps explain why the cross-correlation function shows slightly stronger correlations at negative lags — suggesting that rising temperatures may, in part, drive further increases in CO_2.

5.1.5 Residual Analysis

The three diagnostic plots provide strong evidence that the regression model satisfies the core assumptions of linear regression. The histogram and Q–Q plot support the normality of residuals, while the residuals vs fitted values plot supports homoscedasticity and linearity. Although slight deviations were noted in the residual tails, they are minor and tolerable in the context of this analysis. Together with the results from the Shapiro-Wilks and Studentised Breusch-Pagan tests, these findings justify the use of linear regression and support the validity of its results and interpretations.

5.2 Limitations and Assumptions

While the results of this study provide strong evidence of a relationship between atmospheric CO_2 concentrations and global temperature anomalies, several limitations and assumptions should be acknowledged.

First, the analysis was constrained by the temporal resolution of the available temperature anomaly data, which was recorded on an annual basis. Although atmospheric CO_2 measurements are available at a daily resolution, this higher-frequency data could not be utilised due to the lack of equivalently granular temperature data. The use of annual averages may have obscured short-term dynamics or lagged relationships that might be detectable with finer-resolution datasets.

Second, although global temperature anomaly records extend back to 1880, the atmospheric CO_2 dataset used in this study only begins in 1959. While proxy-based reconstructions or extrapolations could theoretically be employed to estimate earlier CO_2 concentrations, such methods were excluded due to concerns over reliability and model uncertainty. As a result, a substantial portion of the earlier temperature data had to be excluded from analysis, potentially omitting long-term patterns or shifts in the CO_2–temperature relationship.

Conversely, the CO_2 dataset extends to 2024, whereas the temperature anomaly data ends in 2016. To ensure consistency and maintain the integrity of year-on-year comparisons, later CO_2 observations were also excluded. This necessarily limits the ability to incorporate the most recent data or assess whether the observed relationships have persisted or changed in the last decade.

Finally, although the residual diagnostics for the linear regression model generally supported the normality and homoscedasticity assumptions, they were not perfectly satisfied. Minor deviations in the tails of the Q–Q plot and slight narrowing of residual spread at higher fitted values were observed. While these issues were not severe enough to invalidate the model, they suggest some caution is warranted when interpreting inferential statistics derived from the regression.

These limitations do not undermine the main findings but highlight areas where the use of more comprehensive or higher-resolution data — or alternative modelling approaches — could enhance future analyses.

5.3 Future Work

While this study has provided strong statistical evidence supporting the relationship between atmospheric CO_2 concentrations and global temperature anomalies, several avenues for future work could help expand upon and refine these findings.

First, future analyses could incorporate higher-resolution datasets. The availability of daily or monthly CO_2 measurements already exists, and if equivalent high-frequency temperature data were accessible, more granular time series techniques — such as lagged regression models or Granger causality tests — could be applied to explore short-term dynamics and delayed effects in greater depth.

Second, the temporal scope of the data could be extended by incorporating reconstructed historical CO_2 data. Ice core records and paleoclimate reconstructions, while subject to greater uncertainty, offer valuable insight into long-term climate-carbon dynamics and could allow for analysis of trends stretching back hundreds or even thousands of years. Combining these with modern observational data would support a more comprehensive, long-range investigation into how the CO_2–temperature relationship has evolved over different geological periods.

Additionally, future work could employ multivariate models to account for other influential climate variables such as methane (CH_4), solar irradiance, aerosol concentrations, and oceanic oscillations (e.g., El Niño–Southern Oscillation). Doing so would help isolate the specific contribution of CO_2 while accounting for natural variability and other anthropogenic drivers.

Beyond analytical refinement, future research could also explore predictive modelling. By training machine learning models on historical CO_2 data, it may be possible to forecast future atmospheric concentrations under various emissions scenarios. These projections could then be used to estimate future global temperature anomalies. When combined with datasets on extreme weather events and sea level rise, such modelling could help paint a data-driven picture of how the climate may evolve if current trends continue. These forecasts could serve not only as scientific evidence but also as a compelling communication tool — one that helps policymakers, stakeholders, and the wider public understand the urgency of immediate climate action. Delaying change by even a decade may drastically limit the ability to avoid irreversible environmental and societal consequences.

5.4 Summary and Concluding Remarks

This study has demonstrated a strong and statistically significant relationship between atmospheric CO_2 concentrations and global temperature anomalies using a combination of visual, statistical, and time series analyses. While some limitations in temporal resolution and data availability were acknowledged, the findings are consistent with the broader scientific consensus that anthropogenic CO_2 emissions are a primary driver of recent global warming. Through careful model diagnostics and interpretation, this work reinforces the urgency of climate action and highlights the value of transparent, reproducible analysis in supporting evidence-based decision-making. Future extensions of this work could provide not only deeper insights, but also more impactful predictions — strengthening the case for immediate and sustained mitigation efforts.

References

[1] Arrhenius, S. 1896. On the influence of carbonic acid in the air upon the temperature of the ground. Philosophical Magazine and Journal of Science. [Online]. [Accessed 12 May 2025]. Available from: https://doi.org/10.1080/14786449608620846

[2] Foster, G. and Rahmstorf, S. 2011. Global temperature evolution 1979–2010. Environmental Research Letters. [Online]. 6(4). [Accessed 12 May 2025]. Available from: https://doi.org/10.1088/1748-9326/6/4/044022

[3] Intergovernmental Panel on Climate Change (IPCC). 2021. Climate Change 2021: The Physical Science Basis. [Online]. Sixth Assessment Report. Geneva: World Meteorological Organization. [Accessed 12 May 2025]. Available from: https://www.ipcc.ch/report/ar6/wg1/

[4] 3. Lüthi, D., Le Floch, M., Bereiter, B. and Stocker, T.F. 2008. Lüthi, D., Le Floch, M., Bereiter, B. and Stocker, T.F. 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. Nature. [Online]. 453(7193), pp.379–382. [Accessed 12 May 2025]. Available from: https://doi.org/10.1038/nature06949

[5] NASA Goddard Institute for Space Studies. 2024. Global Temperature Trends: 1880 to Present. [Online]. [Accessed 12 May 2025]. Available from: https://data.giss.nasa.gov/gistemp/

[6] NOAA Global Monitoring Laboratory. 2024. Trends in Atmospheric Carbon Dioxide – Mauna Loa Observatory. [Online]. [Accessed 12 May 2025]. Available from: https://gml.noaa.gov/ccgg/trends/

[7] Rohde, R., Muller, R., Jacobsen, R. and Wurtele, J. 2013. A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview. [Online]. [Accessed 12 May 2025]. Available from: https://doi.org/10.4172/2327-4581.1000101