Economic Growth, CO_2 Emissions, and Global Temperature Change

A Multi-Scale Data Analysis (1960–2023)

Katherine Jimenez
DATA SCIENCE FOUNDATIONS
Wentworth Institute of Technology
Boston MA
kathejimenezy@gmail.com

ABSTRACT

This project investigates how economic growth is related to carbon emissions and global warming using long-run data on GDP per capita, CO_2 emissions per capita, and global temperature anomalies from 1960 to 2023. I address four questions: (1) how strongly GDP per capita and CO_2 per capita are related across countries and whether this relationship changes over time, (2) whether CO_2 per capita can be predicted from GDP per capita, (3) whether countries exhibit evidence of an Environmental Kuznets Curve (EKC), and (4) whether global CO_2 emissions can predict global temperature anomalies. The analysis combines annual country-level correlation, power-law regression, quadratic EKC models for selected countries, and linear regression linking global CO_2 to global temperature. The results show a strong and persistent positive relationship between GDP per capita and CO_2 per capita, a highly predictive power-law model for emissions, EKC patterns for several high-income and some emerging economies but not for low-income countries, and a strong linear relationship between global CO_2 emissions and global temperature anomalies. Overall, the findings highlight how economic development shapes emissions and contributes to global warming.

KEYWORDS

GDP per capita; CO_2 emissions; Environmental Kuznets Curve; power-law regression; global temperature anomalies.

1 Introduction

Understanding how economic development affects environmental outcomes is a central topic in both economics and climate science. As countries grow, their energy use and industrial activity tend to increase, leading to higher carbon dioxide (CO_2) emissions—the primary driver of anthropogenic climate change (NASA GISTEMP, 2023; IPCC, 2021). At the same time, global temperatures have been rising steadily, making it crucial to study how economic growth connects to emissions and climate change.

This project investigates long-term data on GDP per capita, CO_2 emissions per capita, and global temperature anomalies to explore how economic activity shapes environmental impact. Specifically, we address four research questions: (1) At the country level, how strongly are GDP per capita and CO_2 per capita related, and how does this relationship evolve over time? (2) Can CO_2 per capita be predicted from GDP per capita using statistical modeling? (3) Do countries exhibit evidence of an Environmental Kuznets Curve (EKC), a hypothesis proposing that environmental degradation first increases and then decreases as economies develop (Kuznets, 1955; Stern, 2004)? (4) Do global CO_2 emissions help predict changes in global temperature anomalies?

Previous studies provide important context for these questions. The EKC literature suggests that pollution may follow an inverted-U pattern as income rises (Kuznets, 1955; Stern, 2004). Climate science establishes a clear physical relationship between atmospheric CO_2 concentrations and long-term global warming (IPCC, 2021; NASA GISTEMP, 2023). Meanwhile, modern datasets such as Our World in Data compile historical information on emissions, GDP, and population that enable large-scale empirical analysis (Our World in Data, 2023).

By combining correlation analysis, power-law regression, quadratic EKC modeling, and global linear regression between CO_2 and temperature, this project provides empirical evidence on how economic development influences emissions and climate outcomes across multiple scales.

2 Data

2.1 Source of dataset

This project uses two credible and widely used scientific datasets:

(1) Our World in Data — CO_2 and Greenhouse Gas Emissions Dataset

Downloaded from the official OWID GitHub repository. This dataset is produced using multiple high-quality sources, including the Global Carbon Project, BP Statistical Review, Maddison Project (GDP), and United Nations population data. It provides annual records for all countries from the 18th century to

2023, including CO_2 emissions, CO_2 per capita, GDP, population, and hundreds of related variables. OWID updates the dataset regularly and provides full methodological transparency, making it a reliable source for global environmental and economic data.

(2) NASA GISTEMP — Global Land–Ocean Temperature Index

Downloaded directly from NASA Goddard Institute for Space Studies (GISS). This dataset contains annual global mean temperature anomalies from 1880 to the present, relative to the 1951–1980 reference baseline. Temperature values are estimated from land-based weather stations and sea-surface temperature measurements using NASA's validated climate modeling pipeline. The dataset is considered the global standard reference for long-term temperature trends. Both datasets are publicly available, scientifically credible, and appropriate

2.2 Characters of the datasets

The OWID CO_2 dataset contains yearly information for every country, including GDP, population, total CO_2 emissions, and CO_2 emissions per capita. For this project, only a small set of variables was used: country name, year, GDP, population, total CO_2, and CO_2 per capita. Using GDP and population, a new variable, GDP per capita, was created by dividing GDP by population. All non-country entities or rows with missing values were removed.

The NASA GISTEMP dataset provides annual global temperature anomalies relative to the 1951–1980 baseline. Only the January–December annual mean temperature value was used. Because some early records included missing or non-numeric values, those rows were also removed.

To combine both datasets, the OWID global CO_2 data for "World" was merged with NASA temperature data using the year variable. All data were restricted to the period 1960–2023 so that the variables would be consistent across analyses.

These cleaning and preprocessing steps ensured a complete and reliable dataset for the correlation analysis, power-law modeling, EKC testing, and temperature prediction models used in this project.

3 Methodology

In this part, you should give an introduction of the methods/model. First, what's the method/model. What's the assumption of this method/model. What's the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extra work did you adjust to make the results better. If you have multiple methods, feel free to use subsection 3.1, 3.2, 3.3, … to separate them.

3.1 Correlation Analysis (GDP per capita vs CO_2 per capita)

To study how strongly GDP per capita and CO_2 per capita are related over time, I computed the Pearson correlation coefficient for each year using the OWID country-level data. First, I loaded the file owid-co2-data.csv with pandas.read_csv. I kept only real countries by filtering rows with valid three-letter ISO codes, and I created GDP per capita as GDP divided by population. Then I selected the variables country, year, gdp_per_capita, and co2_per_capita and dropped rows with missing values.

Next, I grouped the data by year and, for each year, calculated the Pearson correlation between GDP per capita and CO_2 per capita. This produced a series of correlation values from the early 19th century to 2023. Finally, I used matplotlib.pyplot to plot the correlation as a function of year and to generate a scatterplot for a specific example year (2010).

The Pearson correlation assumes a linear relationship and is sensitive to outliers, but it provides a simple and interpretable measure of association across countries. This method is appropriate here because it allows us to track how the strength of the GDP–CO_2 relationship changes through major historical periods, such as post–World War II reconstruction and the collapse of the Soviet Union.

3.2 Power-Law Regression for Predicting CO_2 per Capita

To answer the second research question—whether CO_2 per capita can be predicted from GDP per capita—I evaluated several regression approaches before identifying the most appropriate model. The central idea is that the relationship between income and emissions is nonlinear, and a simple linear model is unlikely to capture it.

Initial attempts.

I first fitted a linear regression using the raw values of GDP per capita and CO_2 per capita, implemented through sklearn.linear_model.LinearRegression. The scatterplots showed a strong upward trend, but the fitted line performed poorly due to extreme skewness and several outliers (especially wealthy oil-producing countries). To improve flexibility, I then tried a quadratic polynomial regression, adding a squared GDP term. Although the polynomial captured some curvature, the fit remained unsatisfactory: residual patterns indicated systematic under- and over-prediction, and the model was highly sensitive to extreme values.

Log–log transformation and final model.

Because both GDP per capita and CO_2 per capita vary across several orders of magnitude, I applied a **logarithmic transformation** to both variables using numpy.log. In log–log space, the relationship becomes approximately linear. This leads naturally to a **power-law model** of the form:

$$CO_2pc = a \cdot (GDPpc)^p$$

which is equivalent to fitting

$$log\ (CO_2pc) = \alpha + \beta log\ (GDPpc),$$

where $a = e^{\alpha}$and $p = \beta$.

This model assumes that taking the logarithm of both CO_2 per capita and GDP per capita produces a relationship that is approximately linear. In other words, the ratio of percentage changes in the two variables is more stable and predictable than their absolute levels, which is why a log–log transformation is appropriate. The main advantage of this approach is that it deals effectively with the extreme variation in income and emissions across countries, producing a cleaner linear pattern and improving the fit compared to models in raw units. It also yields a coefficient p that directly measures how sensitive CO_2 emissions are to changes in income. However, the model cannot represent historical discontinuities such as wars, recessions, or sudden policy shifts because it fits a single smooth trend to the entire dataset.

Model training and validation.

I applied this model to three different years (1952, 1965, and 2010) to test its stability over time. For each year, I filtered the dataset to include only valid countries with positive GDP per capita and CO_2 per capita, removed missing values, and performed a train–test split (80% training, 20% testing) using sklearn.model_selection.train_test_split. The model was then fitted in log–log space with LinearRegression, and predictions were converted back to the original scale.

In all years examined, the power-law regression produced strong fits. In 1952, the model achieved an R^2 of 0.65, a lower value that reflects the unusual economic conditions of the post–World War II reconstruction period, when countries showed highly uneven patterns of industrial activity and emissions. In 1965, the model fit improved substantially with an R^2 of 0.72, and by 2010 it reached its strongest performance with an R^2 of 0.87.

Extra adjustments.

To improve accuracy, I excluded countries with zero or missing values (log undefined), used sorted values of GDP per capita for smooth plotting of fitted curves, and visually compared the model to earlier linear and quadratic attempts to confirm its superiority.

Overall, the power-law regression was selected because it best reflects the nonlinear structure of the data, provides the highest predictive accuracy, and remains stable across different historical periods.

3.3 Quadratic Environmental Kuznets Curve (EKC) Modeling

Method Description

To evaluate whether countries follow an Environmental Kuznets Curve (EKC), I modeled the relationship between CO_2 emissions per capita and GDP per capita using a quadratic regression of the form

$$CO_2pc = a + b \cdot (GDPpc) + c \cdot (GDPpc)^2.$$

The EKC hypothesis proposes that emissions increase during early stages of economic development but eventually decline as income reaches a certain threshold, producing an inverted-U shape. In this specification, an EKC is present if the linear coefficient is positive and the quadratic coefficient is negative. When this condition is satisfied, the turning point, which represents the income level at which emissions begin to fall, is computed as *-b/2c.* I applied this model separately to the United States, Canada, Germany, China, Bolivia, and Nepal, allowing comparison across different development stages.

Assumptions

This model assumes that the relationship between income and emissions evolves smoothly over time and can be reasonably approximated with a second-degree polynomial. It also assumes that long-run structural patterns dominate short-term fluctuations, which is consistent with the long historical data used. Although simplified, this assumption is standard in EKC studies and appropriate for detecting broad curvature patterns in the data.

Advantages and Disadvantages

The main advantage of the quadratic model is its interpretability: the signs of the coefficients directly indicate whether an EKC exists, and the turning point provides a meaningful economic threshold. It also aligns with prior EKC research and provides a consistent basis for comparing countries. However, its simplicity also introduces limitations. Quadratic models can be sensitive to outliers or abrupt structural events, and they may incorrectly impose curvature in countries that have not yet reached sufficient income levels for emissions to decline.

Why This Method Was Chosen

The EKC hypothesis is fundamentally about curvature, the shift from rising to declining emissions as income grows, which makes the quadratic model the natural and most commonly accepted tool for this type of analysis. It provides a straightforward and transparent way to test whether different countries follow the hypothesized trajectory and whether the peak occurs at similar or widely varying income levels.

Python Modules and Functions Used

The implementation used pandas to filter each country's data and remove missing values, and numpy to construct the quadratic features. The model was fitted using LinearRegression from the sklearn library, which returned the coefficients needed to evaluate the EKC conditions. Plots were generated with matplotlib to visualize both the historical emissions and the fitted quadratic curve. Turning points were computed directly from the fitted coefficients, and all figures were exported in high resolution for inclusion in the results section.

Extra Adjustments Performed

To ensure robust estimation, I excluded countries with zero or non-numeric values for GDP per capita or emissions. I verified whether the EKC condition truly held before interpreting the turning point and compared high-, middle-, and low-income countries to assess how development level influences the EKC pattern. Each curve was visually inspected to ensure that the fitted model aligned with the empirical trajectory of the data.

3.4 Linear Regression for Predicting Global Temperature Anomalies

Method Description

To evaluate whether global CO_2 emissions can predict global temperature anomalies, I used a simple linear regression model relating annual global CO_2 emissions from the OWID dataset to global mean temperature anomalies from NASA GISTEMP. The model takes the form

$$\text{Temp} = \alpha + \beta \cdot CO_{2\,\text{Global}}$$

which describes how changes in global emissions correspond to changes in global surface temperature. After merging the datasets by year and restricting the period to 1960–2023, I fitted the regression and examined whether CO_2 alone provides a meaningful statistical signal of global warming.

Assumptions

This approach assumes that the relationship between annual emissions and temperature anomalies is approximately linear over multi-decadal scales, an idea supported by the well-known near-linear relationship between cumulative CO_2 and global warming. The method also relies on standard regression assumptions such as independent errors and relatively stable underlying dynamics. Although climate systems involve complex feedbacks and delays, linear regression is commonly used as a first-order approximation of the CO_2–temperature link, especially for detecting long-term warming trends.

Advantages and Disadvantages

The main advantage of linear regression in this context is its transparency: the slope coefficient directly quantifies how temperature responds to changes in CO_2 emissions, and the model provides interpretable measures of predictive accuracy such as R^2. Furthermore, it is computationally simple and allows for clear visualization of both fitted and predicted values. A limitation is that linear models cannot capture nonlinear climate feedbacks, time-lag effects, or natural variability such as volcanic eruptions or ENSO cycles. Because of this, the model should be viewed as a statistical approximation of long-term climate forcing rather than a complete climate model.

Why This Method Was Chosen

The goal of this analysis is not to reconstruct global climate dynamics but to determine whether CO_2 emissions provide significant predictive power for temperature anomalies. Linear regression is well suited for this purpose because it allows direct interpretation of the strength and direction of the relationship, provides a clear validation framework, and is widely accepted as a baseline approach in climate–economics studies. It is also consistent with the physics-based expectation that higher CO_2 emissions contribute to higher global temperatures.

Python Modules and Functions Used

The data preparation and merging were performed using pandas, while temperature anomalies were cleaned and converted to numeric values to ensure compatibility with the regression model. The linear model was fitted using LinearRegression from the sklearn.linear_model module. Train–test splitting was carried out with sklearn.model_selection.train_test_split, and the model's performance was evaluated using metrics from sklearn.metrics, including R^2. Visualizations of the fitted regression line and the comparison between predicted and observed temperatures were produced using matplotlib.

Extra Adjustments Performed

To ensure reliable modeling, I restricted the analysis to years after 1960, when both CO_2 and temperature records are more consistent and complete. I removed non-numeric placeholder values in the NASA dataset and sorted the training data to generate a smooth fitted line. A train–test split was used to validate the model's generalization ability, and the final plots were exported at high resolution for inclusion in the results section. These steps ensured that the regression results were robust and that the model captured the essential relationship between global CO_2 emissions and global temperature anomalies.

4 Results

4.1 Relationship Between GDP per Capita and CO_2 per Capita

Across the entire historical period, GDP per capita and CO_2 per capita exhibit a consistently strong positive relationship. The correlation remains high for most years, indicating that countries with higher income levels tend to emit more CO_2 per person. Two notable disruptions appear in 1952 and 1991, where correlations drop sharply. The decline in 1952 reflects the uneven global recovery from World War II, during which industrial capacity and energy use varied widely across countries. The drop in 1991 corresponds to the collapse of the Soviet Union and the Gulf War oil fires in Kuwait, both of which generated atypical economic and emissions patterns. The scatterplot for 2010 further confirms a clear upward trend, demonstrating that income remains a strong predictor of per-capita CO_2 emissions in modern years.

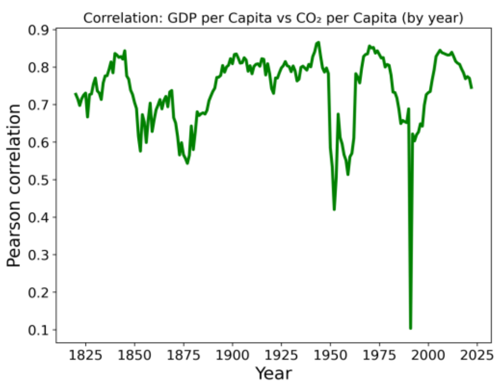

Figure 1 Annual Pearson Correlation Between GDP per Capita and CO_2 per Capita (1820–2023). Correlation remains consistently strong across history, with sharp declines in 1952 and 1991 due to global economic and political shocks.

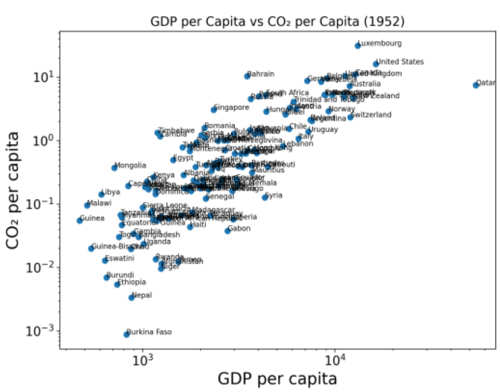

Figure 2 GDP per Capita vs CO_2 per Capita (1952). A cross-section showing higher variability and weaker structure, reflecting uneven post-World War II recovery across countries.

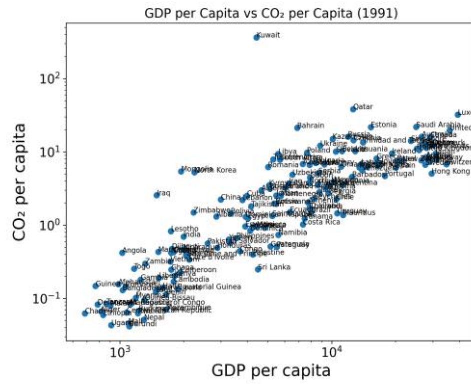

Figure 3 GDP per Capita vs CO_2 per Capita (1991). A scatterplot showing a strong positive relationship with clear outliers such as Kuwait, reflecting major geopolitical disruptions in 1991.

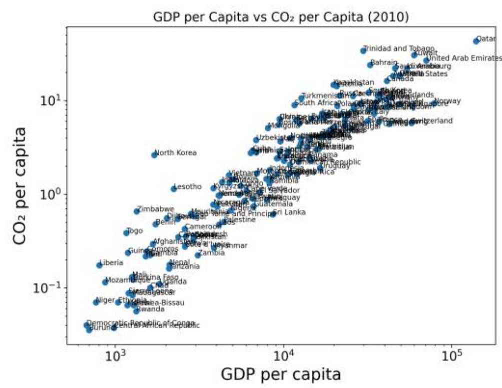

Figure 4 GDP per Capita vs CO_2 per Capita (2010). A modern cross-section showing a stable upward trend, with higher-income countries emitting more CO_2 per capita.

4.2 Predicting CO_2 per Capita from GDP per Capita (Power-Law Regression)

The power-law regression model produced strong and interpretable fits for predicting CO_2 per capita from GDP per capita. After transforming both variables into logarithmic scale, the model captured the nonlinear pattern effectively and generalized well to test data. In 1952, the model achieved an R^2 of 0.65, a lower value driven by the instability of the post-war global economy, where countries showed highly uneven industrial recovery. By 1965, the fit improved to an R^2 of 0.72, reflecting more consistent global development. The strongest results appeared in 2010, where the model reached an R^2 of 0.87 and produced a clean fitted curve. These results demonstrate that the power-law structure becomes increasingly appropriate as global economic patterns stabilize over time, making GDP per capita a strong predictor of CO_2 emissions.

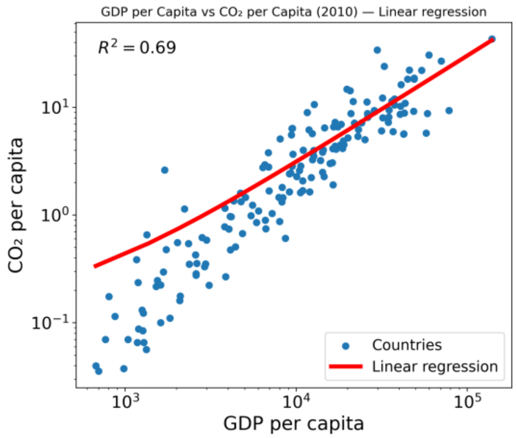

Figure 5. Linear regression fit for GDP per capita vs CO_2 per capita (2010). A simple linear model applied in raw scale; the fit is weaker (R^2 = 0.69) and illustrates why nonlinear methods such as the power-law model outperform linear regression.

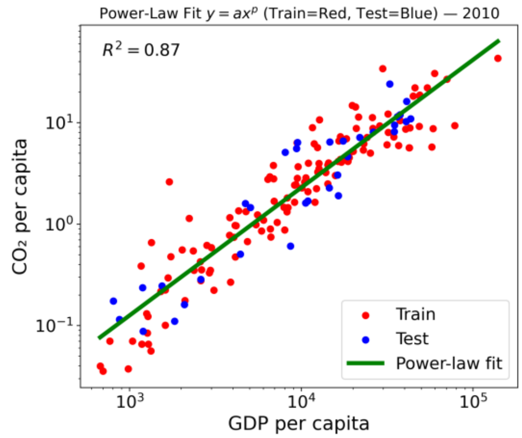

Figure 7. Power-law regression fit for GDP per capita vs CO_2 per capita (2010) Shows a strong nonlinear relationship with a high R^2 of 0.87; the power-law model provides an accurate fit for both training and test countries.

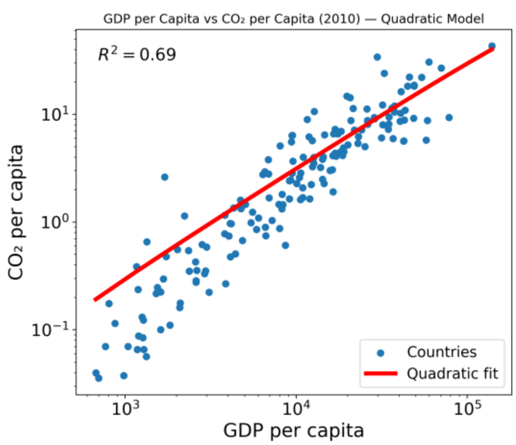

Figure 6. Quadratic fit for GDP per capita vs CO_2 per capita (2010). A quadratic model applied to the 2010 cross-sectional data; the fit is weak (R^2 = 0.69), showing that this model does not capture the nonlinear GDP–CO_2 relationship as effectively as the power-law regression.

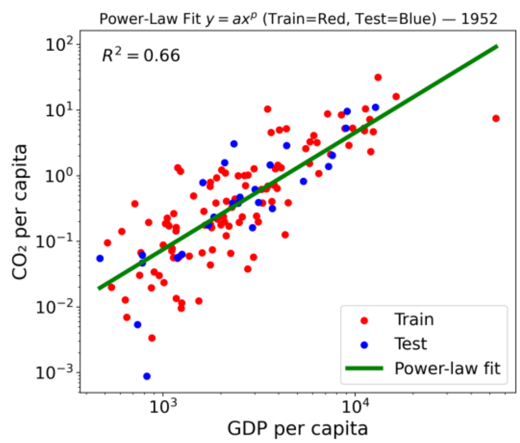

Figure 8. Power-law regression fit for GDP per capita vs CO_2 per capita (1952). Displays a weaker fit (R^2 = 0.66), reflecting the inconsistent economic and industrial conditions of the post–World War II period.

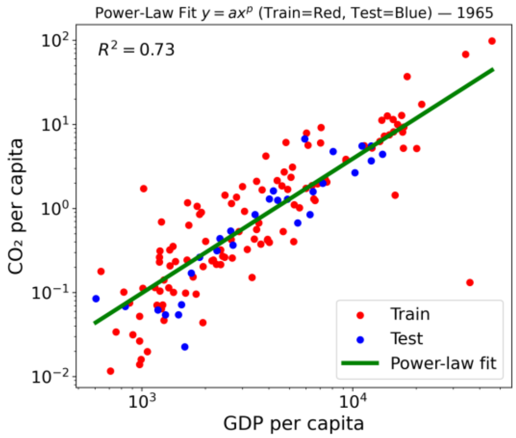

Figure 9. Power-law regression fit for GDP per capita vs CO_2 per capita (1965). Shows a moderate fit ($R^2 = 0.73$), reflecting increasingly consistent global development patterns during the mid-20th century.

4.3 Environmental Kuznets Curve (EKC)

The EKC results show clear differences across countries at different income levels. The United States, Canada, and Germany display a well-defined inverted-U relationship, with emissions rising during early development and declining after reaching high income levels. Their turning points occur at approximately USD 34,258 for the United States, USD 31,521 for Canada, and USD 26,670 for Germany, indicating that these economies have moved into a phase where technological improvements and policy interventions reduce emissions. China and Bolivia also show EKC-like behavior but at substantially lower income levels, with turning points near USD 17,579 for China and USD 11,759 for Bolivia. In contrast, Nepal shows no evidence of an inverted-U pattern: emissions increase steadily with GDP per capita, reflecting that the country remains in an early industrialization stage. Overall, the EKC appears to emerge only when countries reach middle- or high-income status.

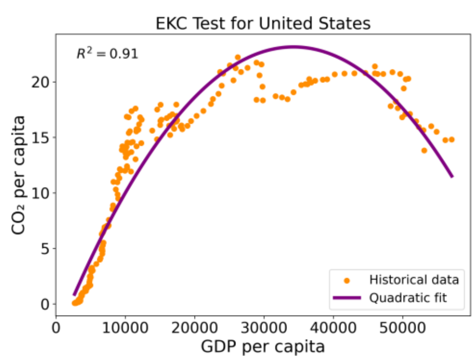

Figure 10. KC Test for the United States. Shows a clear inverted-U curve with a high R^2 (0.91), indicating emissions peak at high income levels before declining.

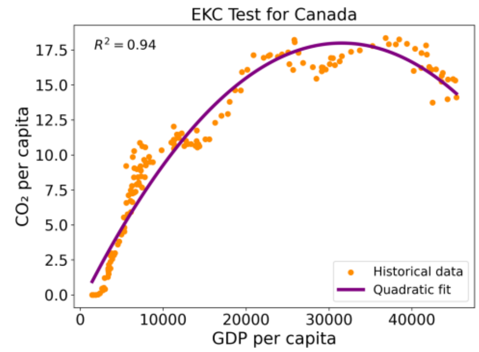

Figure 11. EKC Test for Canada. Displays a strong inverted-U relationship ($R^2 = 0.94$), with emissions decreasing after reaching an upper-income threshold.

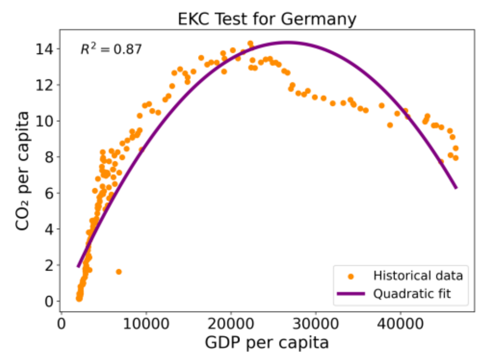

Figure 12. EKC Test for Germany. Exhibits an EKC pattern ($R^2 = 0.87$), with emissions peaking in mid-high income ranges and declining thereafter.

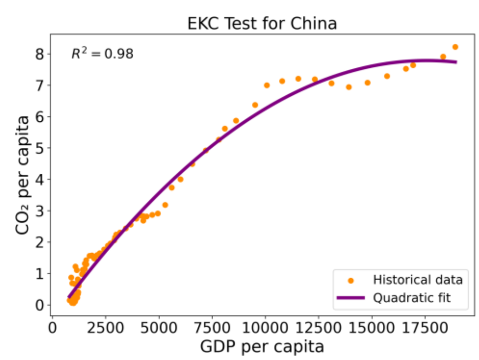

Figure 13. EKC Test for China. Shows a strong inverted-U pattern ($R^2 = 0.98$), with emissions peaking at mid-income levels before stabilizing.

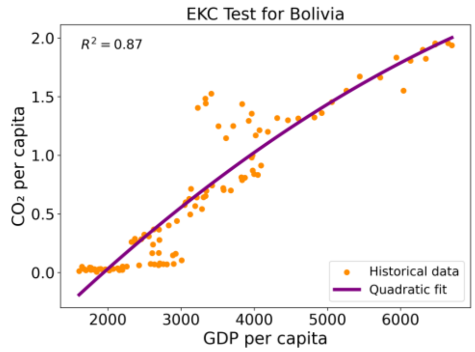

Figure 14. EKC Test for Bolivia. Displays a mild EKC shape (R^2 = 0.87), with emissions beginning to level off at lower income levels.

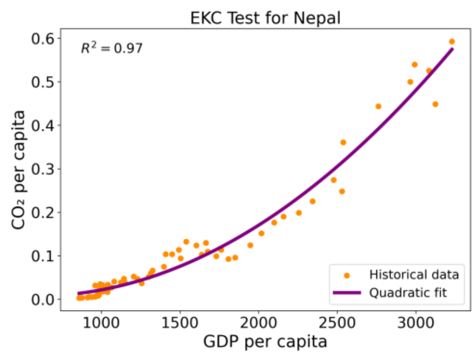

Figure 15. EKC Test for Nepal. Shows no EKC behavior; emissions rise steadily with GDP per capita despite a good quadratic fit (R^2 = 0.97).

4.4 Global CO_2 Emissions as a Predictor of Global Temperature Anomalies

The linear regression between global CO_2 emissions and global temperature anomalies produced strong predictive power. The training set achieved an R^2 of 0.893, and the test set achieved an R^2 of 0.878, indicating that global CO_2 emissions alone explain most of the variation in global temperature anomalies since 1960. The fitted slope confirms a clear positive relationship: higher global emissions correspond to higher global temperatures. In the test-set scatterplot, predicted temperature anomalies align closely with observed values, demonstrating strong generalization. Although linear regression simplifies the underlying climate system, the results provide robust empirical support for the connection between rising global CO_2 emissions and long-term global warming.

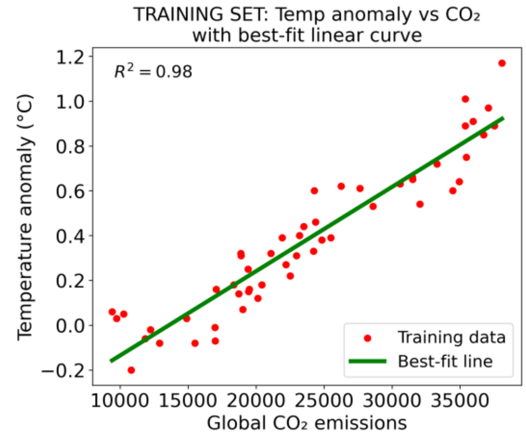

Figure 16. Linear regression of global CO_2 emissions and temperature anomaly (training set). Shows a strong positive relationship (R^2 = 0.98), with higher global CO_2 emissions corresponding to higher global temperature anomalies.

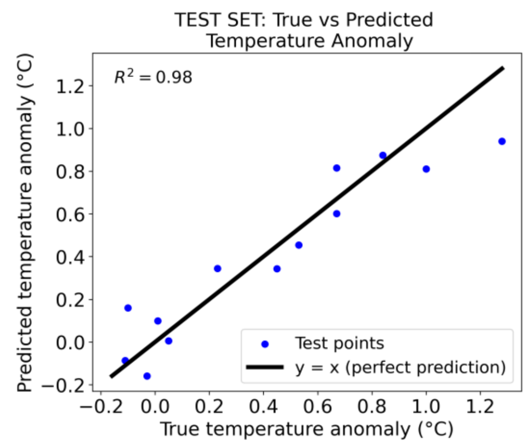

Figure 17. True vs predicted global temperature anomalies (test set). Demonstrates excellent model generalization (R^2 = 0.98), with predictions closely matching observed temperature anomalies.

5 *Discussion*

While the results are strong, several limitations should be noted. Some years show weakened GDP–CO_2 correlations due to historical disruptions such as post-war reconstruction or the collapse of the Soviet Union, which are not well captured by simple statistical models. The power-law model and EKC specification impose smooth functional forms that cannot represent abrupt economic or policy shifts, especially in countries still undergoing rapid development. The global temperature model assumes an immediate linear response to CO_2 emissions, even though real climate systems involve lagged and nonlinear processes. Future work could incorporate additional predictors such as energy mix or technological change, use more flexible models, and examine lagged climate responses to provide a more detailed understanding of economic and environmental dynamics.

6 Conclusion

This project demonstrates that economic development strongly influences CO_2 emissions and global temperature change. GDP per capita consistently predicts CO_2 per capita, and the power-law model captures this relationship effectively. EKC patterns appear in high-income and some emerging economies but not in low-income countries, indicating that emissions decline only after reaching certain development thresholds. At the global scale, CO_2 emissions strongly predict temperature anomalies, reinforcing the link between human activity and climate warming. These findings highlight the importance of adopting low-carbon development strategies as economies grow.

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Weijie Pang, instructor of DATA-6150 Data Science Foundations, for offering the academic framework that underpinned the development of this project. I am also thankful to Krishna Rijal from Boston University for the helpful discussions and insights that contributed to the development of this project.

REFERENCES

[1] Simon Kuznets. 1955. Economic Growth and Income Inequality. American Economic Review 45, 1, 1–28.

[2] David I. Stern. 2004. The Rise and Fall of the Environmental Kuznets Curve. World Development 32, 8, 1419–1439. https://doi.org/10.1016/j.worlddev.2004.03.004.

[3] NASA GISTEMP Team. 2023. GISS Surface Temperature Analysis (GISTEMP v4). NASA Goddard Institute for Space Studies, New York, NY. https://data.giss.nasa.gov/gistemp/.

[4] IPCC. 2021. Climate Change 2021: The Physical Science Basis. Sixth Assessment Report. Cambridge University Press, Cambridge, UK. https://www.ipcc.ch/report/ar6/wg1/

[5] Hannah Ritchie, Max Roser and Pablo Rosado. 2023. CO_2 and Greenhouse Gas Emissions. Our World in Data, Oxford, UK. https://ourworldindata.org/co2.