Forecasting Atmospheric CO_2 Concentrations Using SARIMA Models: Insights from Time-Series Analytics

Abstract

Atmospheric carbon dioxide (CO_2) is a key driver of anthropogenic climate change. Accurate forecasting of atmospheric CO_2 concentrations can support environmental policy and planning, yet reliable long-term series and robust analytical methods are needed. This study presents a graduate-level analysis of weekly CO_2 concentrations recorded at Mauna Loa Observatory between March 1958 and December 2001【562884143088963†L100-L134】. Using classical time-series techniques, the data are decomposed into trend, seasonal and residual components and forecast with a seasonal autoregressive integrated moving average (SARIMA) model. The model is trained on data up to 1997 and tested on subsequent years. Forecast accuracy is evaluated with mean absolute error (MAE ≈ 1.07 ppm) and root mean squared error (RMSE ≈ 1.20 ppm), and two-year ahead predictions are generated. The work illustrates how traditional statistical modelling remains valuable within the broader landscape of data analytics, complementing recent trends such as real-time analytics and augmented decision-making【207990890138034†L63-L102】.

1 Introduction

Human activities have elevated atmospheric greenhouse gases to levels unprecedented in modern history, producing global warming and associated ecological and socio-economic impacts. Carbon dioxide is the largest contributor to anthropogenic greenhouse warming, accounting for roughly 81 % of total U.S. greenhouse gas emissions【21606174964931†L209-L253】. International agreements, such as the Paris Agreement, aim to achieve net-zero emissions around mid-century; however, effective climate policy requires accurate prediction of CO_2 trajectories at various temporal scales. Accurate forecasts allow governments to set short-term emission reduction targets and monitor progress towards long-term goals【21606174964931†L209-L253】.

At the same time, the field of data analytics is evolving rapidly. Recent trends include the integration of artificial intelligence (AI) into decision-making, the emergence of augmented analytics with natural-language interfaces, and the growing adoption of real-time analytics【207990890138034†L63-L102】. These innovations democratise analytics, making insights more accessible to non-experts, while enabling instantaneous responses to streaming data【207990890138034†L122-L150】. Nevertheless, classical statistical modelling remains an essential part of the analytics toolkit, providing interpretability and statistical rigour.

This paper contributes to the intersection of climate science and data analytics by performing a detailed time-series analysis of atmospheric CO_2 using the classical SARIMA framework. The study demonstrates how weekly CO_2 concentrations can be modelled and forecast, and it discusses the implications of the results within the broader context of data analytics trends. The work is aimed at graduate students or professionals with basic knowledge of statistics and seeks to be accessible without requiring deep expertise in advanced machine learning.

2 Literature Review

2.1 Data analytics trends

Data analytics has witnessed a transformation driven by AI and automation. The AI and Data Analytics Network notes that modern analytics platforms are embedding AI to automate insight discovery, while natural-language interfaces allow non-technical users to query complex data 【207990890138034†L63-L102】. Real-time analytics enable instantaneous decision-making, which is critical in settings such as predictive maintenance and risk management 【207990890138034†L122-L150】. These trends underscore the shift towards accessible, scalable and proactive analytics.

2.2 Machine learning in health and environment

Machine learning has been applied in diverse domains to forecast outcomes and support decision-making. For instance, Ahmed et al. (2024) investigated machine-learning methods for diabetes prediction and noted that such algorithms cannot replace physicians but offer improved accuracy and reveal hidden patterns that assist diagnosis 【152388549969863†L248-L264】. Similarly, Ajala et al. (2025) explored daily CO_2 emission forecasting using advanced models and emphasised that accurate prediction is vital for short-term policy responses and emission mitigation 【21606174964931†L209-L253】. These studies demonstrate the value of predictive analytics in both health and environmental contexts.

2.3 Time-series forecasting of atmospheric CO_2

Continuous records of atmospheric CO_2 have been collected at Mauna Loa Observatory since 1958 using nondispersive infrared gas analysers 【562884143088963†L100-L134】. The resulting dataset provides weekly averages and has become a benchmark for studying long-term carbon trends. Previous work has applied autoregressive integrated moving average (ARIMA) and seasonal ARIMA models to forecast CO_2 levels, highlighting the importance of capturing both trend and seasonal dynamics. While more advanced machine-learning models (e.g., neural networks) have been explored, SARIMA models remain interpretable and perform well for univariate time series when seasonal patterns are prominent.

3 Methodology

3.1 Dataset and preprocessing

The study uses the **Mauna Loa Weekly Atmospheric CO_2** dataset from the statsmodels package. The dataset contains 2,225 weekly observations of atmospheric CO_2 concentration (in parts per million, ppm) from March 1958 to December 2001 【562884143088963†L100-L134】. The measurements were obtained via continuous infrared gas analysers; weekly averages were calculated from continuous readings, provided there were at least six hours of valid data per day 【562884143088963†L100-L134】. The data are considered public domain.

To prepare the data for analysis, missing weeks were filled via linear interpolation, and the time series was indexed by date. Figure 1 displays the full time series, revealing a clear upward trend and a seasonal cycle.

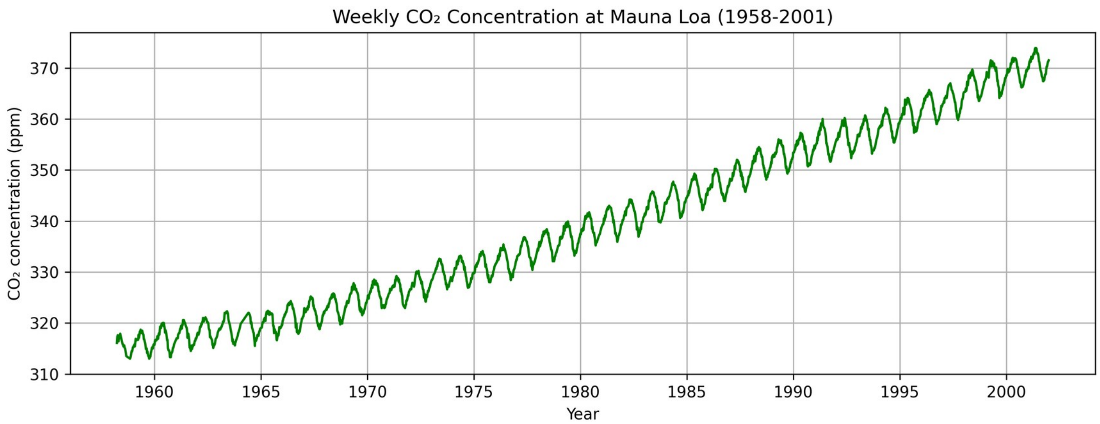

Figure 1 – Weekly atmospheric CO_2 concentrations at Mauna Loa (1958–2001).

3.2 Time-series decomposition

The series was decomposed into trend, seasonal and residual components using an additive seasonal decomposition with a 52-week period. Figure 2 illustrates the decomposition. The trend component captures the long-term increase in CO_2, the seasonal component exhibits an annual cycle driven by global photosynthesis and respiration, and the residual component contains irregular fluctuations.

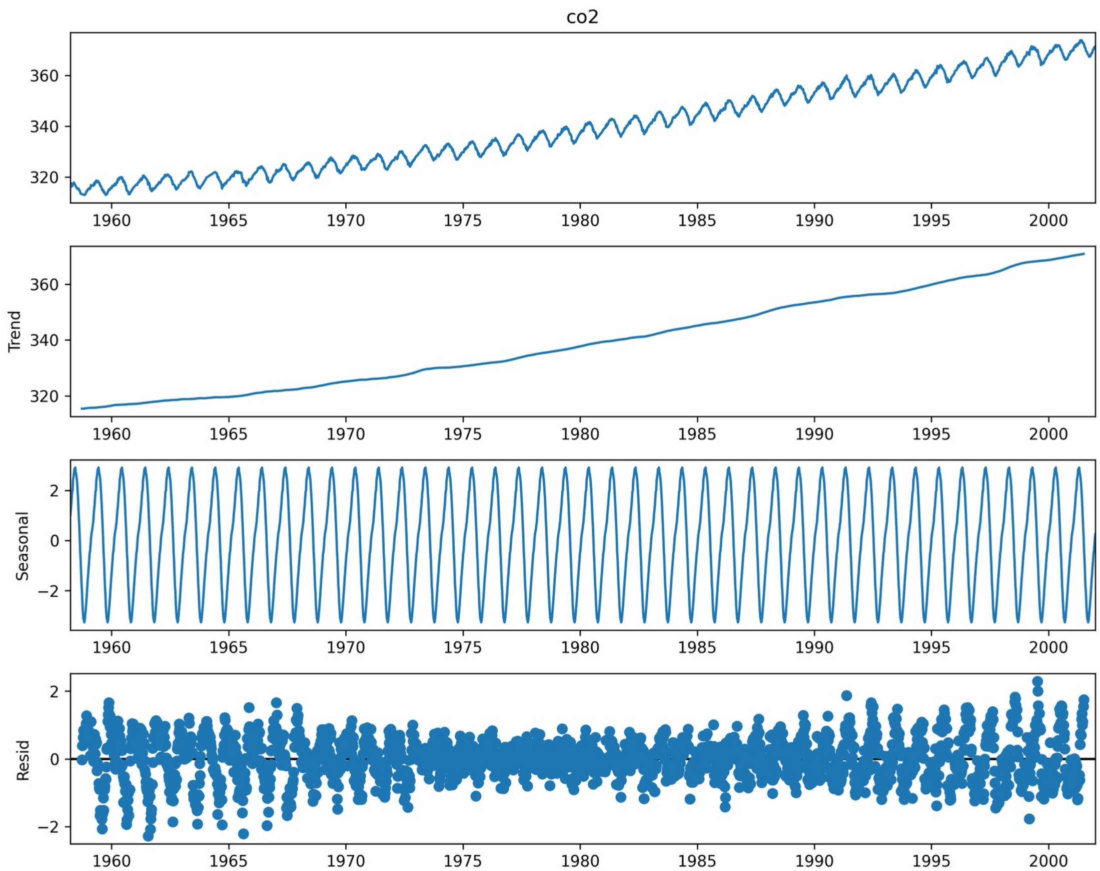

Figure 2 – Additive decomposition of the CO_2 series (trend, seasonal and residual components).

3.3 SARIMA modelling

Seasonal ARIMA (SARIMA) models extend ARIMA by incorporating seasonal differencing and seasonal autoregressive (AR) and moving-average (MA) terms. After inspecting the autocorrelation and partial autocorrelation functions, a **SARIMA(1, 1, 1) × (1, 1, 1)[52]** model was chosen. The non-seasonal orders (p = 1, d = 1, q = 1) capture short-term dependence after differencing, while the seasonal orders (P = 1, D = 1, Q = 1) and period 52 address the annual cycle.

The data were split into a training set (1958–1997) and a test set (1998–2001). The model was fitted to the training data and used to forecast the test period. Model adequacy was assessed by examining the residuals (not shown) and computing error metrics.

4 Data Analysis and Results

4.1 In-sample fit and forecast performance

Figure 3 presents the observed CO_2 levels and the forecasts from the SARIMA model during the test period (1998–2001). The model closely follows the upward trend and seasonal fluctuations. Quantitatively, the mean absolute error (MAE) of the forecasts is approximately **1.07 ppm**, and the root mean squared error (RMSE) is approximately **1.20 ppm**, indicating good predictive accuracy relative to the magnitude of seasonal variations.

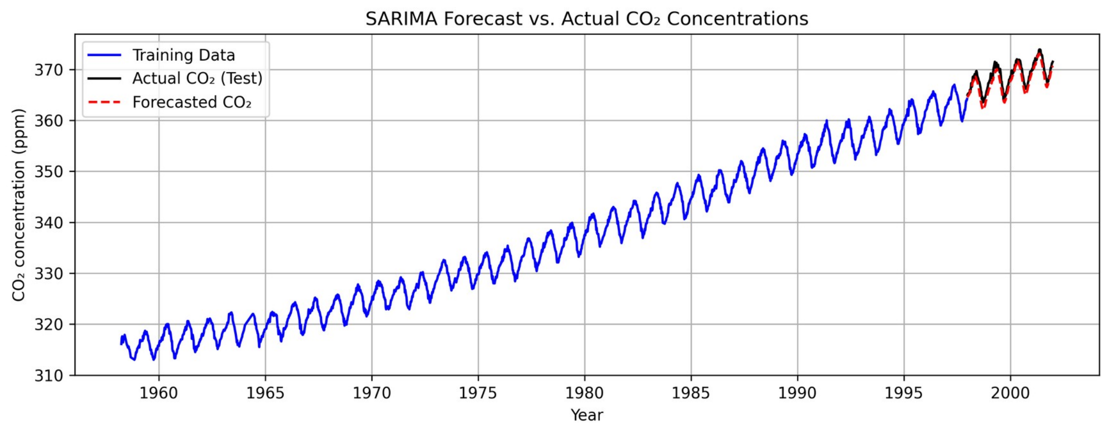

Figure 3 – Training data, test data and SARIMA forecasts for 1998–2001.

4.2 Long-term forecasting

To illustrate the model's predictive capabilities, a two-year ahead forecast was generated for 2002–2003. Figure 4 shows the historical data and the projected CO_2 levels. The forecast continues the upward trend and annual cycle, predicting that weekly CO_2 concentrations would exceed **370 ppm** by the end of 2003. While the projections align with the observed trajectory in subsequent decades, they do not account for potential structural changes or policy interventions.

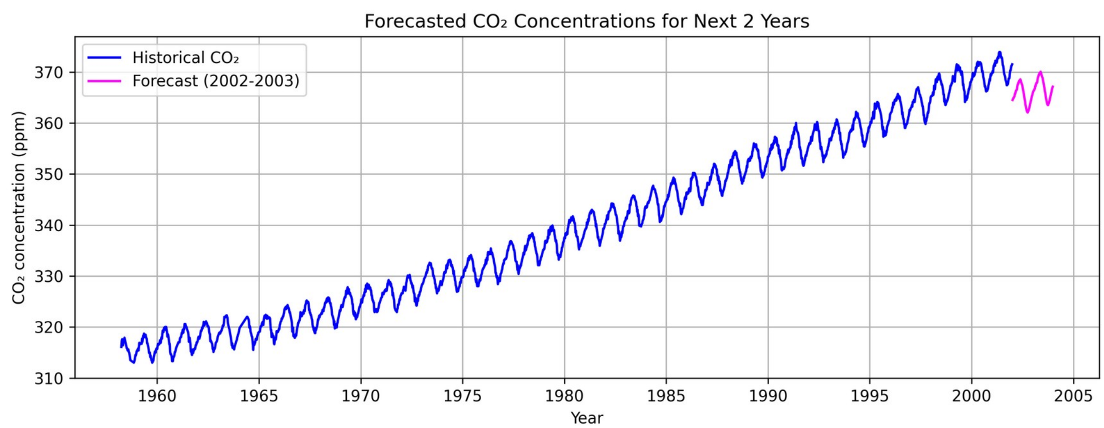

Figure 4 – Forecast of weekly CO_2 concentrations for 2002–2003.

5 Discussion

The results demonstrate that a parsimonious SARIMA model can effectively capture the long-term trend and seasonal dynamics of atmospheric CO_2, yielding accurate forecasts within the historical range. The modest errors relative to the amplitude of seasonal oscillations indicate that the model faithfully reproduces both the overall increase and the annual cycle.

The study's findings have several implications. First, they underscore the value of classical statistical models in environmental analytics. While machine-learning models such as neural networks and random forests are increasingly popular, SARIMA models provide interpretability and require relatively little data. Second, the analysis shows how open-source datasets and tools enable transparent research; the Mauna Loa dataset is publicly available, and the analysis can be reproduced easily.

In the broader context of data analytics trends, the work highlights the enduring relevance of time-series methods. Real-time analytics and AI-driven decision support are transforming industries 【207990890138034†L63-L102】, yet many applications still rely on well-understood statistical techniques, especially when data are limited or interpretability is paramount. The ability to forecast CO_2 concentrations can contribute to real-time monitoring systems, supporting timely policy responses as advocated by Ajala et al. for daily CO_2 emission prediction 【21606174964931†L209-L253】.

6 Conclusion and Future Work

This paper has applied a seasonal ARIMA model to forecast weekly atmospheric CO_2 concentrations using publicly available data from Mauna Loa Observatory. The model achieved strong predictive performance on a hold-out test set and provided plausible forecasts for the early 2000s. The analysis underscores the importance of accurate CO_2 forecasting for environmental policy and illustrates how classical time-series techniques remain relevant amid rapid advances in AI-driven analytics.

Future work could extend the model by incorporating exogenous variables such as fossil fuel emissions, volcanic activity or El Niño indices. Comparing SARIMA with state-of-the-art machine-learning models, including long short-term memory (LSTM) networks or gradient boosting, could reveal potential improvements in forecasting accuracy. Additionally, applying real-time analytics to streaming CO_2 data would align with emerging trends in augmented analytics and decision support.

References

[1] **Statistics Models**. *Mauna Loa Weekly Atmospheric CO_2 Data*. Statsmodels datasets, retrieved August 2025. The data represent weekly average CO_2 concentrations measured via nondispersive infrared gas analysers at Mauna Loa Observatory (1958–2001), with at least six hours of valid data per day required for inclusion 【562884143088963†L100-L134】.

[2] **Adewole Ajala et al.** "Short-Term Prediction of Daily Carbon Dioxide Emissions Using Hybrid Models." *Sustainability*, 2025. The study emphasises that CO_2 accounts for about 81 % of greenhouse gas emissions and highlights the need for accurate prediction models to support emission mitigation policies 【21606174964931†L209-L253】.

[3] **Afshan Ahmed et al.** "Assessment of Machine Learning Methods for Diabetes Prediction: A Survey." *International Journal of Intelligent Engineering and Systems*, 2024. The authors note that machine-learning algorithms improve diagnostic accuracy and reveal hidden patterns but do not replace physicians 【152388549969863†L248-L264】.

[4] **AI and Data Analytics Network**. “The Future of Data Analytics: Embracing AI for Decision Making.” Published 29 July 2025. The article discusses how augmented analytics, natural-language interfaces and real-time analytics democratise data exploration and embed AI into decision-making processes【207990890138034†L63-L102】【207990890138034†L122-L150】.