Journal of Computer Science

Research Article

Predicting CO_2 Emissions Using Multivariate Regression: Assessing the Impact of Pandemics and Industrial Revolutions

[1] **Youssef Mekki,** [2] **Chouaib Moujahdi,** [1] **Noureddine Assad and** [1] **Aziz Dahbi**

[1] *Information Technology Laboratory, National School of Applied Sciences, Chouaib Doukkali University, El Jadida, Morocco*
[2] *Department of Botany and Plant Ecology, Scientific Institute, Mohammed V University in Rabat, Morocco*

Article history
Received: 20-10-2024
Revised: 05-01-2025
Accepted: 21-01-2025

Corresponding Author:
Youssef Mekki
Information Technology Laboratory, National School of Applied Sciences, Chouaib Doukkali University, El Jadida, Morocco
Email: mekki.y@ucd.ac.ma

Abstract: This paper explores the prediction of CO_2 emissions using Multivariate Regression models, incorporating the influence of significant historical events such as pandemics and industrial revolutions. While existing research primarily focuses on carbon emissions alone in forecasting models, this study emphasizes the importance of incorporating multiple factors for a comprehensive understanding of CO_2 dynamics. Beyond emissions, factors such as socioeconomic indicators, industrial activities, environmental policies, and historical events play significant roles. This multidimensional methodology is essential for developing robust prediction models capable of capturing the complex dynamics of CO_2 emissions. This paper introduces the notion of regression flags to mark these impactful events, revealing the intricate relationship between CO_2 emissions and external factors. By integrating these flags into a Multivariate Regression model, we uncover how different historical contexts shape emissions trends over time. Experimental results underscore the effectiveness of our model in accurately predicting CO_2 emissions dynamics amidst varying historical conditions. The experimental results demonstrate a robust ability of the used Multivariate Regression model with the notion of events flags to achieve precise predictions of CO_2 emissions over time.

Keywords: CO_2 Emissions, Forecasting, Multivariate Regression, Artificial Intelligence

Introduction

The pressing challenge of our era lies in climate change, where anthropogenic carbon dioxide (CO_2) emissions stand out as a key driver, requiring a thorough grasp of their dynamics for effective mitigation strategies, which in turn call for both monitoring and predictive capacities. The ability to forecast CO_2 emissions accurately is crucial for informing policy interventions and guiding sustainable development efforts (Faruque *et al.*, 2022). To this end, sophisticated analytical techniques such as polynomial regression and others offer promising avenues for modeling the complex relationships between CO_2 emissions and various influencing factors (Konya and Nematzadeh, 2024).

In addition to ongoing anthropogenic activities, significant historical events, such as pandemics and industrial revolutions, have demonstrable impacts on CO_2 emissions. These events disrupt societal and economic systems, leading to fluctuations in energy consumption, industrial output, and transportation patterns, all of which influence carbon emissions (Khan *et al.*, 2022). Recognizing the need to account for such events in predictive models, the incorporation of flags or indicators becomes imperative. Flags serve as markers for periods characterized by extraordinary circumstances, enabling the delineation of distinct temporal segments within datasets and facilitating more nuanced analyses.

While existing research predominantly focuses on carbon emissions as the sole predictor in their forecasting models, our study underscores the imperative of integrating multiple factors to achieve a comprehensive understanding of CO_2 dynamics. By broadening the scope beyond carbon emissions alone, our approach advocates for a more holistic perspective that incorporates various criteria crucial for accurate CO_2 predictions. Factors such as socioeconomic indicators, industrial activities, environmental policies, and historical events like pandemics and industrial revolutions can significantly influence CO_2 emissions (Konya and Nematzadeh, 2024). Recognizing the interconnectedness of these variables is paramount for developing robust prediction models that capture the nuanced dynamics of CO_2 developments. Our work underscores the necessity of adopting a multidimensional approach to forecast CO_2 emissions, thereby offering insights crucial for informed decision-making and effective climate mitigation strategies.

In this context, we believe that pandemics have historically played a pivotal role in influencing global CO_2 emissions, often resulting in abrupt. For instance, during the COVID-19 pandemic, global daily CO_2 emissions experienced a significant decline of approximately 17% at the height of global confinement measures in April 2020, compared to the same period in 2019 (Le Quéré *et al.*, 2020). This reduction was attributed to widespread lockdowns, diminished industrial activity, and reduced transportation demand. Similarly, historical pandemics, such as the 1918 influenza pandemic, likely caused regional disruptions in emissions, although comprehensive datasets from earlier periods are sparse.

The transient nature of these reductions underscores the limited capacity of pandemics to effectuate sustained emission decreases without systemic policy interventions. Recognizing this, recent studies have incorporated pandemics as a variable in predictive models to better understand their long-term implications on CO_2 emissions. For example, Liu *et al.* (2020) used machine learning algorithms to project emissions trajectories under scenarios that factor in potential pandemic events. Such models are instrumental in elucidating the interplay between health crises and climate dynamics, enabling policymakers to anticipate and mitigate the rebound effects observed post-pandemic. These predictive approaches not only highlight the immediate impacts of pandemics but also serve as critical tools for planning sustainable recovery strategies that align with global climate objectives.

Brzezinski (2021) explored how past pandemics influenced CO_2 emissions and the shift toward renewable energy. The study found that historical pandemics triggered a short-term decline in CO_2 emissions by approximately 3.4–3.7% while also contributing to a rise in the share of electricity generated from renewable sources over the following five years.

In the forthcoming sections, we will delve into our methodology, explaining the utilization of polynomial regression, as a chosen prediction model for this study, and the notion of event flags. By incorporating historical contexts, our research seeks to offer a nuanced comprehension of the impact of pandemics and industrial revolutions on CO_2 emissions. Through this analysis, our goal is to improve the precision and utility of CO_2 emissions prediction models within a constantly evolving global landscape.

The rest of this paper is organized as follows: In coming paragraphs of this introduction, we embark on a comprehensive review of the current literature pertaining to our research domain. The Materials and Methods section delves into the specifics of our methodology, encompassing a detailed exposition of the used dataset, the processes involved in data integration and preprocessing, as well as the formulation of the used emissions prediction model and its evaluation criteria. The Results section is dedicated to the presentation and discussion of our experimental results, shedding light on the insights gleaned from our analyses. Finally, the Conclusion section encapsulates our concluding remarks and outlines potential avenues for future research and exploration in this field.

Literature Review

Understanding the factors influencing carbon dioxide (CO_2) emissions is crucial for mitigating climate change. Multivariate polynomial regression offers a valuable tool for modeling CO_2 emissions and assessing the impact of various factors. This study aims to predict CO_2 emissions using multivariate polynomial regression and the notion of flag, specifically focusing on the role of pandemics and industrial revolutions. We emphasize that modeling without analyzing the relationships between carbon emissions and various factors cannot reliably support forecasts, as these factors influence carbon emissions over time. Thus, it is crucial to establish a carbon emissions forecasting model capable of uncovering these interactions.

Indeed, must of previous CO_2 forecasting models may not comprehensively depict the intricate interplay between emissions and diverse factors, especially nonlinear relationships. Moreover, existing studies lacks the integration of event flags, such as pandemics and industrial revolutions, into their analyses, emphasizing the need for a nuanced approach. Existing CO_2 forecasting models in the literature can generally be classified into two main types. The first category includes models that assess the influence of a single factor on carbon emission trends. The second category consists of models that analyze the effects of multiple factors on CO_2 emissions independently.

In the context of the first category, Houghton (2003) analyzed land-use changes in the United States and China, as well as the latest FAO estimates of deforestation and tropical afforestation, to calculate the annual flow of carbon between terrestrial ecosystems and the atmosphere, representing direct human activities, resulting in a global release of 156 PgC into the atmosphere from 1850 to 2000. Bouznit and Pablo-Romero (2016) examined the correlation between Algerian GDP and CO emissions, revealing a significant association between the two variables. In Saleh *et al.* (2016), an SVM model is proposed to predict carbon emission expenditure based on energy consumption, with a focus on monitoring CO_2 emissions in business operations. Bokde *et al.* (2021) proposed a novel short-term CO_2 emissions forecast for intelligent scheduling of flexible electricity consumption, aiming to minimize emissions. They compare two time series decomposition methods against state-of-the-art models, achieving a 25%

lower mean absolute percentage error. Alam and AlArjani (2021) conducted a comparative analysis of CO_2 emission forecasts in Gulf countries. Their study employed three forecasting models—Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), and Holt-Winters Exponential Smoothing (HWES)—to predict annual CO_2 emissions in the region. Similarly, Zhao *et al.* (2018) introduced a hybrid forecasting approach, integrating the mixed data sampling (MIDAS) regression model with a Back Propagation (BP) neural network, known as the MIDAS-BP model, to estimate carbon dioxide emissions in the United States. Prakash and Singh (2023) focused on predicting carbon dioxide (CO_2) emissions using various time-series, machine learning, and deep learning models applied to a dataset from the central electricity authority. Köne and Büke (2010) underscores the importance of trend analysis in modeling and forecasting energy-related CO_2 emissions, particularly focusing on the top-25 emitting countries. Through regression analyses, statistically significant trends in CO_2 emissions are identified for eleven countries and the world total, enabling the development of models for future CO_2 emission projections. Libao *et al.* (2017) establishes a method for predicting CO_2 emissions from coal-fired power plants based on combustion analysis and proximate fuel data, aiding in online monitoring.

For the second category where studies examine the impact of several factors on carbon emissions, we can cite first namely Faruque *et al.* (2022) that explores CO_2 emissions, GDP, and energy usage, employing advanced modeling techniques for enhanced predictive accuracy and highlights the concerning upward trend in CO_2 emissions for Bangladesh. Between 1970 and 2010, Ye *et al.* (2022) introduced a dynamic time-delay discrete grey forecasting model, designed to enhance accuracy in predicting China's carbon emissions. This model leverages lagged relationship analysis and impulse response assessment to improve its applicability. In a similar approach, Aftab *et al.* (2021) used the AutoRegressive Distributed Lag (ARDL) model and Johansen cointegration analysis to explore the relationship between CO_2 emissions, energy consumption, and economic growth in Pakistan over the period from 1971 to 2019, employing both bivariate and multivariate techniques. Meanwhile, Jamel and Derbali (2016) analyzed time-series data from 1991 to 2013 across eight Asian countries, uncovering a strong link between economic development, energy consumption, and CO_2 emissions, with economic growth and energy use playing a significant role in environmental degradation.

Always in the context of the second category, Zhang and Lin (2012) analyzed the impact of multiple economic factors on pollution levels, particularly CO_2 emissions, in China between 1995 and 2010. Using the fixed effects model and least squares generalized linear regression, they examined key economic indicators such as demographic intensity, urbanization, GDP, industrial and service production, and energy consumption. Their study found that demographic intensity, GDP, industrial output, and energy consumption had a significant influence on CO_2 emissions.

Likewise, Ozturk and Acaravci (2013) performed a time-series analysis to examine the relationship between financial development, energy consumption, economic growth, trade openness, and CO_2 emissions in Turkey over the period from 1960 to 2007. Their findings indicated that while economic growth and trade openness had a notable impact on environmental pollution, financial development did not play a significant role in determining environmental quality.

Additionally, Salari *et al.* (2021) investigated the link between CO_2 emissions, energy consumption, and GDP across different U.S. states from 1997 to 2016. Their analysis, based on both static and dynamic models, revealed a persistent long-term relationship between various forms of energy consumption and CO_2 emissions. In this context Machine learning algorithms have shown proficiency in understanding intricate associations directly from the dataset.

Linardatos *et al.* (2023) used a multivariate dataset comprising IoT sensor readings of various environmental parameters to predict CO levels in a smart-city setting. This hybrid approach integrates the AutoRegressive Integrated Moving Average (ARIMA) statistical technique with Temporal Fusion Transformers (TFT), a deep learning method. (Xiong *et al.*, 2021) introduced an advanced multi-variable grey model (GM(1,N)), derived from a linear time-varying parameter discrete grey model (TDGM(1,N)), to enhance the accuracy of carbon emission predictions. Similarly, Wang and Ye (2017) developed a nonlinear grey multivariable model incorporating power exponential terms to forecast China's carbon emissions from fossil fuel consumption. Their approach demonstrated greater precision than traditional models, offering valuable insights for energy planning and environmental policymaking.

Chiu *et al.* (2020) designed an innovative Multivariate Grey Prediction Model (MGPM) for CO_2 emissions forecasting, integrating grey relational analysis for feature selection and a neural-network-based residual model for improved accuracy. Validated against real CO_2 emission data, their model outperformed existing MGPMs. Additionally, Huang *et al.* (2022) proposed a nonlinear multivariate grey model (ENGM(1,4)) specifically tailored for predicting carbon emissions from the transportation sector. Their model incorporated the IPAT equation and the STIRPAT model, providing a comprehensive framework for analyzing emissions in this domain.

To enhance model accuracy, they employed the derivation method along with the Quantum Particle

Swarm Optimization (QPSO) algorithm for parameter optimization. The model was validated using 18 years of carbon emission data from China, the USA, and Japan. Comparative analysis demonstrated that the ENGM(1,4) model outperformed other models in prediction accuracy, forecasting an upward trend in carbon emissions for China and the USA, while projecting a decline for Japan between 2019 and 2025.

Similarly, Ding *et al.* (2023) introduced an advanced grey multivariate coupled model (CTGM(1,N)) for carbon emission forecasting. This model integrates the Choquet fuzzy integral with a grey multivariate delay framework, accounting for time-lag effects and interactions between influencing factors. To further enhance precision, the model's parameters were optimized using the whale optimization algorithm. Javanmard and Ghaderi (2022) developed a hybrid method that integrates machine learning and mathematical programming to accurately predict greenhouse gas emissions, utilizing data from Iran between 1990 and 2018. The study employed nine algorithms—ANN, AR, ARIMA, SARIMA, SARIMAX, RF, SVR, KNN, and LSTM—for forecasting, followed by evaluation using five performance indicators.

Wei *et al.* (2018) proposed a hybrid model combining random forest and extreme learning machine for CO_2 emission forecasting. The random forest was used for identifying key influential factors, while the extreme learning machine handled the prediction task. The model's initial weights and biases were optimized using moth-flame optimization.

Ding *et al.* (2017) introduced a novel grey multivariable model to predict CO_2 emissions in China, aiming to overcome the limitations associated with traditional forecasting models.

It introduces optimizations in background value prediction, variable trends incorporation, and adjustment coefficient optimization. Lin *et al.* (2018) addressed the critical issue of forecasting carbon dioxide emissions, crucial for understanding environmental impact and guiding policy. It introduces a novel two-stage approach combining multivariable grey forecasting and genetic programming, overcoming limitations of conventional methods and achieving higher accuracy in predicting emissions trends.

Wu *et al.* (2015) investigated the interplay between, urban population, energy consumption, economic growth, and CO_2 emissions in BRICS countries from 2004 to 2010 using a novel multi-variable grey model. Findings suggest varying effects of economic growth on CO_2 emissions across countries, with Brazil and Russia showing a decreasing trend, while India, China, and South Africa exhibit an increasing trend. Wen and Cao (2020) aimed to accurately predict residential energy-related CO_2 emissions in China by analyzing influential factors. Initial indicators are identified using grey relational analysis, followed by principal component analysis to extract main components for Support Vector Machine (SVM) input. Otherwise Ahmed *et al.* (2020) utilizes Grey system theory to predict CO_2 emissions based on various variables such as environmental-related technological change, fossil fuel and renewable energy consumption, and economic output. The findings indicate a continued rise in emissions intensity in the BICS region, but also highlight the potential of environmental technology advancements to mitigate this intensity while supporting economic growth targets.

In this paper, we propose a CO_2 emission forecasting model based on Multivariate Polynomial Regression and the integration of event flags. Contrary to the other models of literature, the proposed one examine the impact of several factors, simultaneously, on carbon emissions.

Materials and Methods

In this section, we outline the used databases for our experiments, which encompass Our World in Data's repository (Ritchie *et al.*, 2022) and the integrated event flags. Following this, we present the predictive model for CO_2 emissions.

Data Preparation Procedures

The dataset used for this study draws from a reputable source known as "Our World in Data" a platform that curates and provides open-access data related to various global issues, including CO_2 and greenhouse gas emissions. Additionally, historical records of pandemics and industrial revolutions were obtained from the List of Epidemics and Pandemics from Our World in Data' s Historical Pandemics page (Dattani *et al.*, 2023) as well. These records encompassed details of major pandemics, including their onset dates, duration, and geographical spread.

Indeed, adding the Pandemics Flag highlights periods of widespread disease outbreaks, aiding in the analysis of CO_2 emissions during such crises. Furthermore, to facilitate a comprehensive examination of the influence of industrial revolutions on CO_2 emissions, we introduced flags as indicators of significant historical events. These flags denote periods corresponding to major industrial revolutions, such as the First Industrial Revolution in the late 18th to early 19th century and subsequent industrial milestones. By delineating these epochs within our dataset, we sought to compare CO_2 emissions trends across different phases of industrial development, elucidating the long-term impacts of technological advancements and societal transformations on carbon emissions dynamics.

Through rigorous analysis, we aimed to deduce whether a single factor, such as a pandemic or an industrial revolution, is sufficient for accurate emissions prediction, or if multiple criteria are necessary to

DOI: 10.3844/jcssp.2025.1705.1718

precisely determine the trajectory of emissions in the years to follow. Incorporating these diverse datasets and methodologies, our analysis aimed to provide a nuanced understanding of the interplay between pandemics, industrial revolutions, and CO_2 emissions, offering valuable insights for climate change mitigation efforts and policy formulation. In order to seamlessly blend historical emissions data with event-specific information, the dataset preprocessing is required, including the incorporation of flags. This could entail aligning temporal periods, standardizing data formats, and potentially addressing any missing values. Flags are particularly important as they help mark significant events within the dataset, enabling a more nuanced understanding of emissions trends during and after these events. The successful integration of emissions data with event-related data, facilitated by flags, is crucial for accurately capturing the nuanced emissions trajectories during and following these events.

The dataset provided by "Our World in Data" consists of a total of 74 columns, with a datetime index ranging from 1750-01-01 to 2021-01-01. Notably, it includes various attributes related to CO_2 emissions, such as CO_2, cement CO_2, coal CO_2, oil CO_2 and flaring CO_2 among others. Each column represents different aspects of CO_2 emissions, including absolute values, per capita metrics, and cumulative emissions. Additionally, the dataset encompasses related indicators such as population, GDP, primary energy consumption, and greenhouse gas emissions. The dataset offers a comprehensive view of CO_2 emissions across countries and over time, facilitating detailed analyses and insights into global emissions trends and patterns.

In our study, we focused primarily on the CO_2 field within the dataset. This column provides crucial information regarding absolute CO_2 emissions, which served as the primary variable of interest in our analysis. The values unit is million-tones. During implementations, it was found that on the CO_2 field missing values were present in 15174 features out of 31349. Thus, the attributes have been filled with median values.

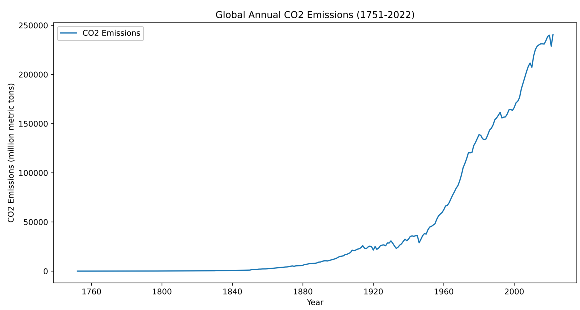

Fig. 1: Global Annual CO_2 Emissions spanning the years 1751 to 2022

Subsequently, we proceeded to visualize the Global Annual CO_2 Emissions spanning the years 1751 to 2022, this can be seen in Figure 1. This visualization offers a comprehensive overview of the historical trends and patterns in CO_2 emissions, enabling us to gain insights into the evolution of emissions over the past centuries. Through this visualization, we aim to elucidate the trajectory of global carbon emissions and identify significant shifts or trends that have occurred over time.

On the other hand, Epidemics and Pandemics information furnishes a comprehensive overview of significant disease outbreaks throughout history, encompassing details on various disease outbreaks that have transpired over time. It includes details such as the name of the epidemic or pandemic, the year it occurred, the affected regions, the estimated number of cases and deaths, and other relevant information. This dataset serves as a comprehensive repository of historical disease outbreaks, providing valuable insights into the patterns, impacts, and responses to infectious diseases over time.

The used list of epidemics & pandemics in this study consists of 272 entries and comprises five columns. Each entry provides information on epidemics or pandemics, including the category and the name of the event (Epidemics/Pandemics), the type of disease (Disease), the death toll (Death Toll), the date of occurrence (Date), and the location (Location). The dataset primarily contains categorical and textual data, with no missing values observed.

Table 1: List of Epidemics and Pandemics

Diseases	Date	Location
Third plague pandemic	1855–1960	Worldwide
Hong Kong flu	1968–1969	Worldwide
cholera pandemic	1846–1860	Worldwide
influenza epidemic	1847–1848	Worldwide
Third cholera pandemic	1846–1860	Worldwide
encephalitis lethargica pandemic	1915–1926	Worldwide
influenza pandemic ('Spanish flu')	1918–1920	Worldwide
psittacosis pandemic	1929–1930	Worldwide
influenza pandemic ('Asian flu')	1957–1958	Worldwide
Seventh cholera pandemic	1961–1975	Worldwide
Hong Kong flu	1968–1970	Worldwide
1977 Russian flu	1977–1979	Worldwide
HIV/AIDS pandemic	1981–present	Worldwide
SARS outbreak	2002–2004	Worldwide
swine flu pandemic	2009–2010	Worldwide
Middle East respiratory syndrome	2012–2021	Worldwide
Zika virus epidemic	2015–2016	Worldwide
COVID-19 pandemic	2019–present	Worldwide
hepatitis of unknown origin	2021–2022	Worldwide
monkeypox outbreak	2022–present	Worldwide

We constructed a comprehensive pandemic dataset. Subsequently, we performed a targeted extraction to create a sub-dataset encompassing only pandemics that exerted a global impact. This sub-dataset included well-documented historical cases such as cholera, the Spanish Flu, and the ongoing COVID-19 pandemic etc. The provided Table 1 presents the list of epidemics and pandemics along with their respective dates and locations.

We have divided the ‘Date’ column of the DataFrame into two separate columns: ‘Start Date’ and ‘End Date’. This division is achieved by identifying the start and end dates within each date range specified in the ’Date’ column. The resulting DataFrame contains then distinct columns for the start and end dates, facilitating easier analysis and interpretation of the data, because We aim to provide an insightful visualization that sheds light on the prevalence of diseases within our dataset. By employing a grouping and counting strategy, we have extracted and ranked the top 12 diseases based on their occurrence frequencies, this can be seen in Figure 2.

Moreover, we can visualize in Figure 3 the top 10 locations with the highest occurrence counts of epidemics and pandemics in the dataset, providing a clear overview of the distribution of epidemics and pandemics across different geographical areas.

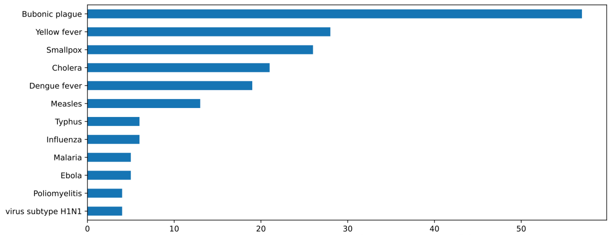

Fig. 2: Top 12 diseases based on their occurrence frequencies

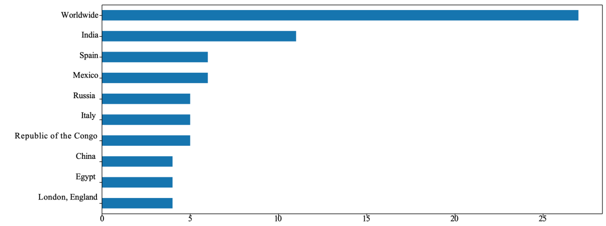

Fig. 3: Top 10 locations with the highest occurrence counts of epidemics and pandemics in the dataset

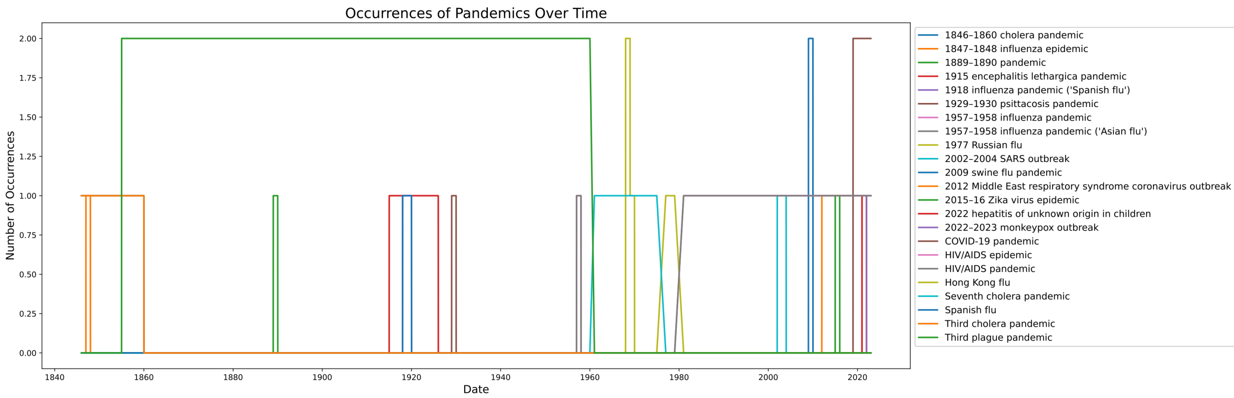

Fig. 4: The number of occurrences of each pandemic over time

Our experimental analysis is conducted on a comprehensive dataset derived from global epidemics, encompassing events that have significantly impacted populations worldwide. We observe that there have been a total of 27 worldwide epidemics documented in our dataset. On the other hand, we produce a time series plot in Figures 4 and 5 that shows the number and occurrences of active global pandemics for each date within the provided data range. This allows us to visualize how many pandemics were occurring simultaneously and identify potential patterns.

In the data preprocessing phase, a new column named Pandemics Flag was introduced to the CO_2 dataset. This column serves as a binary indicator, denoted by a value of 1, to signify the occurrence of a global pandemic in a specific year. Thus, whenever a global pandemic occurred in a particular year, the corresponding entry in the Pandemics Flag column was set to 1, indicating the presence of a pandemic during that period. This additional feature enhances the dataset by providing valuable contextual information about the occurrence of pandemics alongside CO_2 emissions data. It enables subsequent analyses to explore potential correlations or effects of pandemics on CO_2 emissions trends, facilitating a more comprehensive understanding of the dataset dynamics.

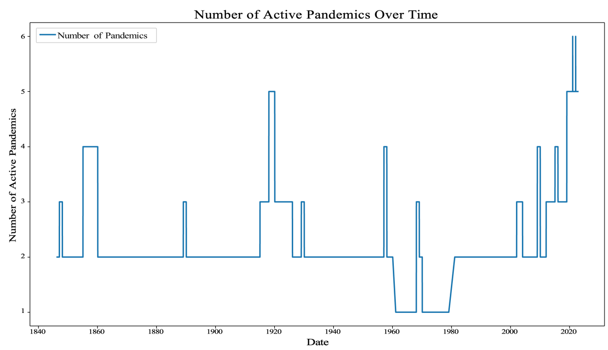

Fig. 5: The number of Active Pandemics Over Time

Furthermore, we enhance the CO_2 dataset by integrating flags indicating the periods of industrial revolutions:

1. Industrial Revolution (1760-1840)

DOI: 10.3844/jcssp.2025.1705.1718

2. Industrial Revolution (1870-1914)
3. Industrial Revolution (1950-1970)
4. Industrial Revolution (2000-present)

Figure 6 illustrates the schema depicting flags integration. Incorporating the Industrial Revolution's Flag into the CO_2 dataset was a pivotal step in augmenting the contextual information surrounding CO_2 emissions. This flag serves as a binary identifier, with a value of 1 denoting the presence of the Industrial Revolution during a particular timeframe. By integrating this additional feature, the dataset gains enhanced granularity, enabling researchers to discern the temporal relationship between industrialization and CO_2 emissions. This augmentation facilitates a more nuanced analysis of CO_2 emission trends, shedding light on the impact of industrialization on environmental dynamics. Consequently, we believe that the inclusion of the Industrial Revolution Flag enriches the dataset, empowering researchers to explore and elucidate the complex interplay between industrial activities and CO_2 emissions over time.

Fig. 6: Schema illustrating the integration of Pandemics Flag and Industrial Revolution Flag into the CO_2 dataset

Data Modeling

In this study, we have chosen to employ the Multivariate Polynomial Regression (Masry, 1996) which is a sophisticated methodology centered around the utilization of advanced Artificial Intelligence (AI) and Machine Learning (ML) techniques to forecast CO_2 emissions. By leveraging these powerful tools, we aim to develop predictive models capable of capturing the intricate dynamics of CO_2 emissions while considering the impacts of pandemic and industrial revolution events. This approach enables us to discern complex patterns and relationships within the data, facilitating more accurate predictions of future emissions trajectories. Through the integration of AI and events flags, we seek to enhance our understanding of the factors driving CO_2 emissions and contribute to informed decision-making in climate change mitigation efforts. Additionally, we investigate whether a single factor suffices to determine emissions trajectories or if multiple factors are required for a comprehensive understanding of CO_2 emissions dynamics.

Multivariate Polynomial Regression (Masry, 1996) is a machine learning technique employed to model the relationship between several independent variables (features) and a dependent variable. It builds on simple polynomial regression, which models the relationship between a single independent variable and the dependent variable, by extending it to cases involving multiple independent variables.

In multivariate polynomial regression, the model presumes a polynomial relationship between dependent and independent variables. It can capture nonlinear relationships between the features and the target variable by introducing polynomial terms of various degrees.

The general form of a multivariate polynomial regression model with n independent variables (features) and up to k-degree polynomial terms is given by Eq. (1):

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \cdots + \beta_{1k} x_1^k + \beta_{22} x_2^2 + \cdots + \beta_{nn} x_n^2 + \cdots + \epsilon \quad (1)$$

Where:

- $\hat{y}$ is the predicted output.
- $x_1, x_2, \ldots, x_n$ are the independent variables (features).
- $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients of the linear terms.
- $\beta_{11}, \beta_{12}, \ldots, \beta_{1k}, \beta_{22}, \ldots, \beta_{nn}$ are the coefficients of the polynomial terms.
- ε represents the error term.

The model learns the coefficients β_i and β_{ij} from the training data to minimize the difference between the actual and predicted values of the dependent variable.

Multivariate Polynomial Regression allows for more flexible modeling of complex relationships between features and the target variable, but it also increases the risk of over fitting, especially when using higher-degree polynomial terms. Regularization techniques can be employed to mitigate over fitting in such cases.

In this study, Multivariate Polynomial Regression is employed for forecasting, utilizing historical data of both dependent and independent variables to train the model. After training, the model can be applied to new data to predict future values of the dependent variable. The accuracy of the predictions is influenced by factors such as data quality, the selected polynomial degree, and the presence of outliers or other sources of error.

In this paper, the Polynomial Regression Models for CO_2 Emissions using Pandemics Flags are as follow:

For CO_2 Emissions:

$$\hat{y}_{CO_2} = \beta_{00} + \beta_{01} \cdot year + \beta_{02} \cdot year^2 + \beta_{03} \cdot year^3$$

For Pandemics Flag:

$$\hat{y}_{Pandemics} = \beta_{10} + \beta_{11} \cdot year + \beta_{12} \cdot year^2 + \beta_{13} \cdot year^3$$

DOI: 10.3844/jcssp.2025.1705.1718

Where:

- *year* represents the years.
- $\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}$ are the coefficients of the polynomial terms for CO^2 emissions.
- $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$ are the coefficients of the polynomial terms for the Pandemics Flag.

These equations represent the regression models where the outputs $\hat{y}_{CO_2}$ and $\hat{y}_{Pandemics}$ are functions of the years, including polynomial terms up to the third degree. The coefficients β_{ij}'s are estimated during the training process to minimize the error term.

The Polynomial Regression Models for CO_2 Emissions using industrial Revolution Flag are as follow:

$$\hat{y}_{CO_2} = \beta_{00} + \beta_{01} \cdot year + \beta_{02} \cdot year^2 + \beta_{03} \cdot year^3$$

$$\hat{y}_{Rev} = \beta_{10} + \beta_{11} \cdot year + \beta_{12} \cdot year^2 + \beta_{13} \cdot year^3$$

Where:

- $\hat{y}_{CO_2}$ and $\hat{y}_{Rev}$ are the predicted CO_2 emissions and predicted industrial revolution flags respectively.
- year represents the years.
- $\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}$ are the coefficients of the polynomial terms for CO_2 emissions.
- $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$ are the coefficients of the polynomial terms for industrial revolution flags.

These equations represent the regression models where the outputs $\hat{y}_{CO_2}$ and $\hat{y}_{Rev}$ are functions of the years, including polynomial terms up to the third degree. The coefficients β_{ij}'s are estimated during the training process to minimize the error term.

The used multivariate polynomial regression model can be described by the Eq. (2) below:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1 x_2 + \beta_5 x_2^2 + \beta_6 x_1^3 + \beta_7 x_1^2 x_2 + \beta_8 x_1 x_2^2 + \beta_9 x_2^3 + \epsilon \quad (2)$$

Where:

- y is the predicted CO_2 emissions.
- x_1 represents the Pandemics Flag.
- x_2 represents the Industrial Revolution Flag.
- β_0 is the intercept term.
- β_1 to β_9 are the coefficients associated with each term.
- ε is the error term.

Moreover, if we use only the prediction on the current values of the dataset until 2022 and the multivariate polynomial regression model with Ridge regularization expressed as follows we can predict more specifically carbon emissions.

The model includes polynomial terms up to the third degree for each feature, and Ridge regularization with a regularization parameter α = 1.0 is applied to mitigate over fitting. The Ridge regression model used can be represented by the following equation:

$$\hat{y} = \left(X^T X + \alpha I\right)^{-1} X^T y$$

Where:

- $\hat{y}$ represents the predicted CO_2 emissions.
- X is the matrix of input features (including polynomial features) with dimensions $n \times p$, where n is the number of samples and p is the number of features.
- y is the target variable vector with dimensions $n \times 1$.
- α is the regularization parameter, which controls the complexity of the model. It is multiplied by the identity matrix I and added to the covariance matrix $X^T X$ to penalize large coefficients
- T denotes the transpose operation.
- $^{-1}$ denotes the matrix inverse operation.

The purpose of Ridge regression is to minimize the residual sum of squares subject to the constraint that the L_2 norm of the coefficients (weights) is less than a certain value, determined by the regularization parameter α. This helps prevent overfitting by adding a penalty for large coefficients, effectively shrinking them towards zero.

In summary, Multivariate Polynomial Regression is a valuable tool for forecasting that allows for the modeling of nonlinear relationships between variables. By leveraging historical data and polynomial terms, it provides a flexible framework for predicting future outcomes with enhanced accuracy and granularity.

Model Validation

To evaluate the performance of the predictive model, the R-squared (R^2) value was used as a key metric. (R^2), or the coefficient of determination, is a statistical indicator that reflects the proportion of variance in the dependent variable that can be explained by the independent variable(s) in a regression model. Mathematically, it is defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where:

- SS_{res} is the sum of squares of residuals (the differences between actual and predicted values).
- SS_{tot} is the total sum of squares (the differences between actual values and the mean of the dependent variable).

The R^2 value ranges from 0 to 1, where:

- $R^2 = 0$ indicates that the model does not explain any of the variability of the response data around its mean.

DOI: 10.3844/jcssp.2025.1705.1718

- R^2 = 1 indicates that the model explains all the variability of the response data around its mean.
- An R^2 value closer to 1 indicates a better fit of the model to the data.

The Root Mean Squared Error (RMSE) metric (Karunasingha, 2022) is another method used to assess the model's performance. It is computed by taking the square root of the mean of the squared differences between the predicted and observed values. RMSE provides a single value that quantifies the typical magnitude of errors made by a predictive model. A lower RMSE indicates better predictive accuracy, as it shows that the model's predictions are closer to the actual observed values. The equation for calculating RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

Where:

- n represents the number of data points.
- y_i represents the actual values.
- $\hat{y}_i$ represents the predicted values.
- The summation symbol $\sum$ represents the sum of squared differences between actual and predicted values.
- The fraction $\frac{1}{n}$ represents the average.
- The square root symbol √ represents the square root of the average of squared differences.

Results

Forecasting CO_2 Emissions Using Flag

Our preliminary analysis revealed that the occurrence of global pandemics could lead to a significant reduction in CO_2 emissions. Based on historical data and predictive modeling, we estimated that a global pandemic could cause a reduction in CO_2 emissions of up to 10%. This highlights the potential environmental benefits of pandemic-related disruptions, which may offer valuable insights for policymakers and environmental scientists. However, the predictive models reveal an intriguing trend wherein emissions appear to resume their upward trajectory after a brief period following the epidemic. This phenomenon can be seen in the Table 2 which presents the CO_2 emissions (in million tonnes) for each year during the Spanish Flu period.

Table 2: CO_2 Emissions During the Spanish Flu Period

Year	CO_2 Emissions (million tonnes)
1916	24560.298
1917	25450.989
1918	24842.169
1919	21446.094
1920	24997.031

In the same way, we noticed that the observed resurgence of CO_2 emissions during the years 2021 and 2022, coinciding with the easing of restrictions and the gradual resumption of economic activities after COVID 19. The following Table 3 shows the evolution of emissions over these years These observations suggest the influence of additional factors beyond the immediate impact of the pandemic on CO_2 emissions.

Table 3: CO_2 Emissions after COVID 19

Year	CO_2 Emissions (in millions of tonnes)
2019	239925.740
2020	228766.484
2021	240609.645

In this context, we notice that the used model indicates an increase in flags since 1850, reaching a peak during the COVID-19 pandemic, followed by a subsequent decrease post-COVID. This observed trend aligns with the historical occurrences of pandemics and their aftermaths, suggesting a cyclical pattern in global health emergencies. The model's ability to capture such fluctuations underscores its effectiveness in forecasting the dynamics of pandemics flags over time, providing valuable insights into the temporal evolution of public health crises.

The Figure 7 illustrates the predicted CO_2 emissions and the predicted occurrence of pandemics over the forecasted period.

Furthermore, The analysis of CO_2 emissions and the impact of global pandemics revealed several important findings. Firstly, our predictive models achieved high accuracy in forecasting both CO_2 emissions and the occurrence of pandemics, as indicated by the high R-squared values obtained. The R-squared value for CO_2 emissions prediction was found to be 0.88, suggesting that our model can explain approximately 88% of the variance in CO_2 emissions. Similarly, the R-squared value for the prediction of pandemics was 0.82, indicating a good fit of the model to the data.

Table 4 presents the R-squared values obtained from the predictive models for CO_2 emissions and Pandemics Flag.

Table 4: R-squared Values

Variable	R-squared Value
CO_2 Emissions	0.8820
Pandemics Flag	0.8161

Table 5 presents the RMSE values obtained from the predictive models for CO_2 emissions and Pandemics Flag.

Table 5: RMSE Values

Variable	RMSE Value
CO_2 Emissions	9239.01
Pandemics Flag	31.17

For the CO_2 emissions model, the RMSE is approximately 9239.01. This value represents the

DOI: 10.3844/jcssp.2025.1705.1718

average error between the actual and predicted CO_2 emissions Lower RMSE values indicate better model performance, thus in this case, while the RMSE is quite large, it is dependent on the scale of the CO_2 emissions data, in our case the value of CO_2 emissions can reach 350000. For the Pandemics Flag model, the RMSE is approximately 3117. This indicates a very small average error between the actual and predicted values of the Pandemics Flag, proving a good performance of the model in predicting this variable.

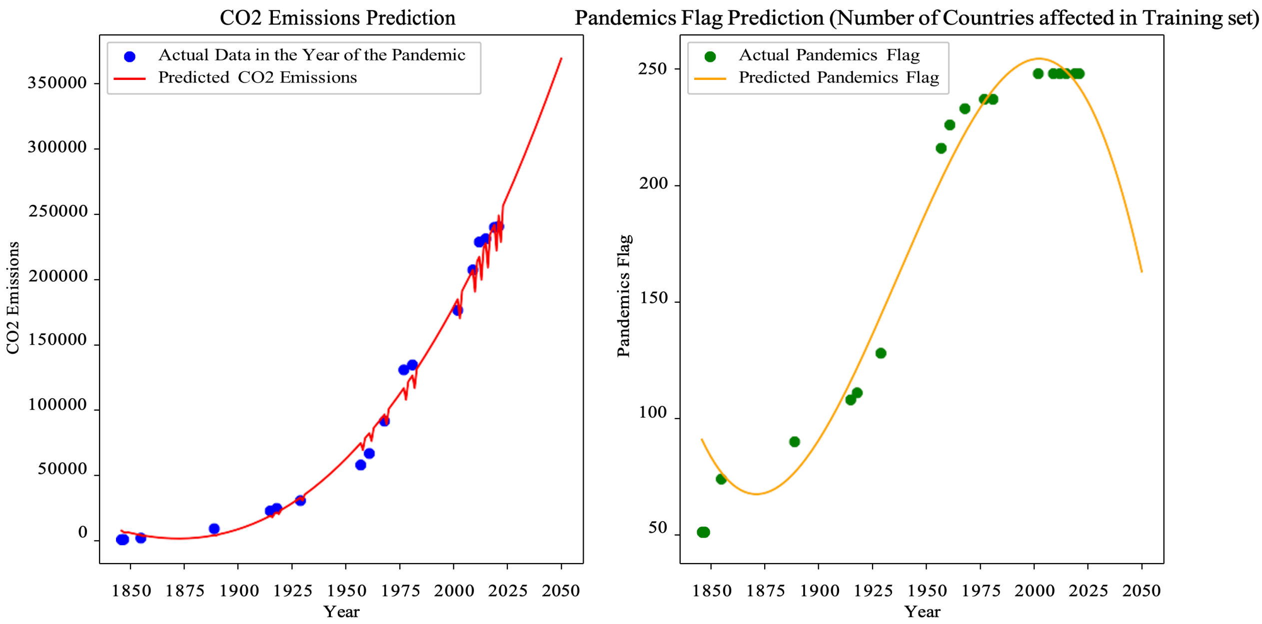

Fig. 7: CO_2 Emissions Prediction with pandemics Flag

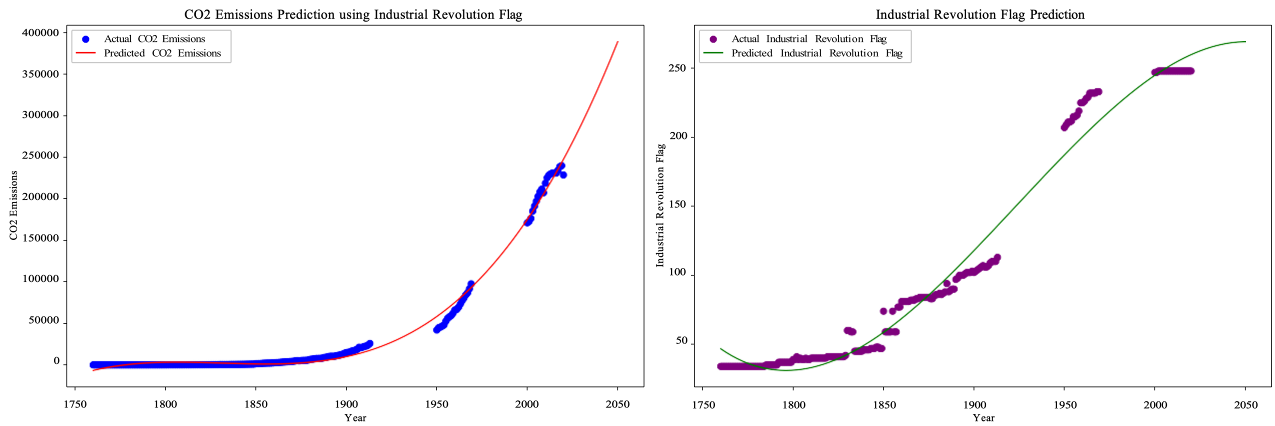

Fig. 8: CO_2 Emissions Prediction with Industrial Revolution Flag

These results prompted us to delve further into the underlying cause of this surge by the study of CO_2 emissions during industrial revolutions.

By investigating the influence of the industrial revolutions on CO_2 emissions, we observe an exponential increase in carbon emissions from the 90s, this is due to the industrial revolution which marked a turning point in the history of CO_2 emissions. Increased use of fossil fuels, deforestation, the development of new technologies and population growth have all contributed to increased CO_2 in the atmosphere, which continues to affect the planet's climate. In this context, the used predictive model indicates a continuous increase in industrial revolution flags, reaching its zenith towards the latter half of the 21st century. The Figure 8 depicts the predicted CO_2 emissions and the trend of industrial revolution flags over time.

The R-squared value for CO_2 Emissions (0.9911) shows that approximately 99.11% of the variance in CO_2 emissions can be explained by the predictors included in the model. This high R-squared value indicates that the model fits the data very well and has strong predictive power for CO_2 emissions. On the other hand, the R-squared value for the Industrial Revolution Flag (0.9741) indicates that around 97.41% of the variance in the occurrence of industrial revolutions can be explained by the predictors in the model. This proves that the model also fits the data well for predicting industrial revolution events based on the provided features.

Table 6 presents the R-squared values which indicate that the models have high explanatory power and are able to capture a significant portion of the variability in the respective target variables.

Table 6: R-squared Values

Variable	R-squared Value
CO_2 Emissions	0.9911
Industrial Revolution Flag	0.9741

Table 7 presents the RMSE values. For the CO_2 emissions model, the RMSE is approximately 5026.25. This value represents the average error between the actual and predicted CO_2 emissions. A relatively low RMSE like this indicates that model performance is good. For the Industrial Revolution Flag model, the RMSE is approximately 11.56. Similarly, this indicates a very small average error between the actual and predicted values of the Industrial Revolution Flag, proving ans excellent performance in predicting this variable by the used model.

Table 7: R-squared Values

Variable	RMSE Value
CO_2 Emissions	5026.25
Industrial Revolution Flag	11.56

In our next main experiment, we use the current values of the dataset until 2022 and the multivariate polynomial regression to predict CO_2 emissions based on two flags: Pandemics Flag and Industrial Revolution Flag. The Pandemics Flag indicates the occurrence of pandemics, while the Industrial Revolution Flag signifies the periods of industrial revolution. By incorporating these two flags as features in our multivariate regression model, we aim to capture their combined effect on CO_2 emissions over time.

In this scenario of test, the calculated R-squared value was 0.687. This value indicates that the model, using the Pandemics Flag and Industrial Revolution Flag as predictors, explains approximately 687% of the variance in CO_2 emissions. This proves that while the presence of a pandemic flag may not directly cause a decrease in CO_2 emissions, it does not solely dictate the trajectory of CO_2 emissions. Indeed, other factors, such as the industrial revolution flag, likely contribute to the observed variability in CO_2 emissions. The increase in CO_2 emissions despite the presence of a pandemic flag could be influenced by various factors associated with industrial activity, economic conditions, or policy decisions that impact emission levels.

In summary, the model's performance shows that while pandemics may not directly lead to a decrease in CO_2 emissions, there are other influential factors at play that affect emission trajectories. The Figure 9 presents predicted CO_2 Emissions with Multivariate Polynomial Regression using Pandemics and Industrial Revolution Flags.

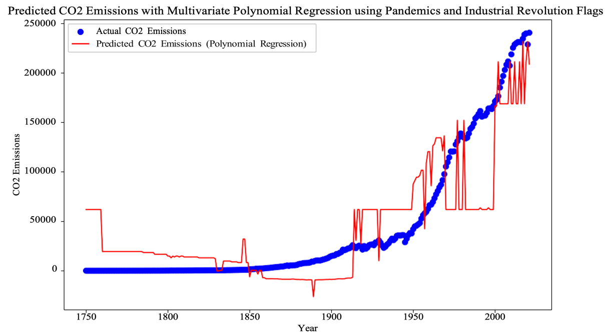

Fig. 9: CO_2 Emissions with Multivariate Polynomial Regression using Pandemics and Industrial Revolution Flags

To predict the future CO_2 emissions using the same methodology, we extended our forecasting model to 2050 as can be seen in the Figure 10. This last highlights the limitation of relying solely on the pandemic factor for precise predictions. Indeed, while pandemics may exert short-term influences on CO_2 emissions, long-term trends indicate a multitude of interconnected factors shaping emission trajectories. Thus, achieving accurate forecasts necessitates a holistic approach that considers diverse variables and their evolving dynamics over time.

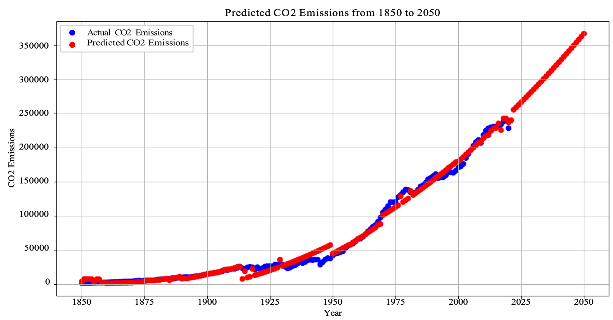

Fig. 10: Forecasting the CO_2 Emissions until 2050 with Multivariate Polynomial Regression using Pandemics and Industrial Revolution Flags with a regularization parameter

The calculated value of the R-squared in this scenario is of 0.992. This value indicates an exceptional level of variance in CO_2 emissions explained by the model, showing an accurate fit to the data. This signifies that the chosen predictors, including year, Pandemics Flag, and flag industrial revolution, along with their polynomial terms, effectively capture the underlying trends and fluctuations in CO_2 emissions.

In addition, with an RMSE of 6454.39, the used model demonstrates promising accuracy in forecasting CO_2 emissions. This level of error is considered acceptable within the context of our application, where the model's predictions are used to inform environmental policy decisions and guide sustainability initiatives. While further refinement and validation may be warranted in certain scenarios, the model's performance aligns well with our expectations and contributes valuable insights towards understanding and mitigating the impacts of CO_2 emissions on our environment.

The presented results demonstrate a robust ability of the used model / methodology to account for the intricate interplay between historical events such as pandemics and industrial revolutions, enabling precise predictions of CO_2 emissions over time. Overall, the model's strong predictive performance underscores its reliability for understanding and forecasting CO_2 emissions dynamics.

Discussion

In this study, we delved into the prediction of CO_2 emissions using a Multivariate Regression model, while taking into account major events like pandemics and industrial revolutions. Surprisingly, The achieved results challenge the common assumption that CO_2 emissions uniformly decline during pandemics; instead, various factors may counterbalance, potentially leading to an increase We highlighted the intricate impact of pandemics on CO_2 emissions, considering factors like lockdown measures during the COVID-19 pandemic By using regression flags to mark these events, we unveiled the complex interplay between CO_2 emissions and external factors, crucial for refining models and shaping effective emission mitigation policies in a changing world While many studies focus solely on CO_2 emissions in their forecasts, we argue for a broader methodology as presented.

By considering various factors such as economic indicators, industrial activities, and policy changes alongside CO_2 emissions, we gain a richer understanding of emission dynamics. The multidimensional methodology emphasizes the importance of recognizing the interconnectedness of these factors for developing robust prediction models and shaping effective climate policies.

Conclusion

This paper explores the complex relationship between historical events and their impact on CO_2 emissions. Employing a Multivariate Regression model, we scrutinize the multifaceted dynamics underlying carbon emissions. Our study underscores the significance of accounting for significant historical events, such as pandemics and industrial revolutions, in CO_2 prediction models. By integrating these events into our analysis, we reveal nuanced insights into the complex relationship between external factors and CO_2 emissions. This holistic methodology not only enhances our understanding of carbon emission dynamics but also furnishes valuable insights for policymakers and stakeholders striving to devise effective strategies for mitigating CO_2 emissions and addressing climate change challenges. Additionally, we evaluate the predictive accuracy of the used model using metrics like the R-squared and the RMSE, further enhancing the robustness of achieved results.

In this context, this paper suggest the following future perspectives: Implement real-time data integration mechanisms to continuously update predictive models, enabling timely adjustments in response to evolving events such as pandemics and industrial shifts. Additionally, explore the incorporation of machine learning algorithms to further refine CO_2 emission predictions, leveraging their capacity to adapt to changing datasets and capture complex relationships. Such methodology hold the potential to enhance the accuracy and reliability of CO_2 forecasting models, facilitating more effective decision-making in climate mitigation efforts.

Acknowledgment

The authors would like to acknowledge Hannah Ritchie, Pablo Rosado and Max Roser for sharing the dataset used in this work.

Funding Information

No funding was received for conducting this study.

Authors Contributions

All authors have contributed equally.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Data Availability

The used models and dataset in this paper can be downloaded from https://github.com/YoMekki/CO2ForecastingMultivariate.

References

Aftab, S., Ahmed, A., Chandio, A. A., Korankye, B. A., Ali, A., & Fang, W. (2021). Modeling the Nexus Between Carbon Emissions, Energy Consumption and Economic Progress in Pakistan: Evidence from Cointegration and Causality Analysis. *Energy Reports*, *7*, 4642-4658. https://doi.org/10.1016/j.egyr.2021.07.020

Ahmed, S., Ahmed, K., & Ismail, M. (2020). Predictive Analysis of CO_2 Emissions and the Role of Environmental Technology, Energy Use and Economic Output: Evidence from Emerging Economies. *Air Quality, Atmosphere & Health*, *13*(9), 1035-1044. https://doi.org/10.1007/s11869-020-00855-1

Alam, T., & AlArjani, A. (2021). A Comparative Study of CO_2 Emission Forecasting in the Gulf Countries Using AutoRegressive Integrated Moving Average, Artificial Neural Network, and Holt-Winters Exponential Smoothing Models. *Advances in Meteorology*, *2021*, 1-9. https://doi.org/10.1155/2021/8322590

Bokde, N. D., Tranberg, B., & Andresen, G. B. (2021). Short-Term CO Emissions Forecasting Based on Decomposition Approaches and Its Impact on Electricity Market Scheduling. *Applied Energy*, *281*, 116061. https://doi.org/10.1016/j.apenergy.2020.116061

Bouznit, M., & Pablo-Romero, M. del P. (2016). CO_2 Emission and Economic Growth in Algeria. *Energy Policy*, *96*, 93-104. https://doi.org/10.1016/j.enpol.2016.05.036

Brzezinski, M. (2021). The Impact of past Pandemics on CO_2 Emissions and Transition to Renewable Energy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3837444

Chiu, Y.-J., Hu, Y.-C., Jiang, P., Xie, J., & Ken, Y.-W. (2020). A Multivariate Grey Prediction Model Using Neural Networks with Application to Carbon Dioxide Emissions Forecasting. *Mathematical Problems in Engineering*, *2020*, 1-10. https://doi.org/10.1155/2020/8829948

Dattani, S., Rodés-Guirao, L., Mathieu, E., Ritchie, H., & Roser, M. (2023). Pandemics. *Our World in Data*. https://ourworldindata.org/pandemics

Ding, Q., Xiao, X., & Kong, D. (2023). Estimating Energy-Related CO_2 Emissions Using a Novel Multivariable Fuzzy Grey Model with Time-Delay and Interaction Effect Characteristics. *Energy*, *263*, 126005. https://doi.org/10.1016/j.energy.2022.126005

Ding, S., Dang, Y.-G., Li, X.-M., Wang, J.-J., & Zhao, K. (2017). Forecasting Chinese CO_2 Emissions from Fuel Combustion Using a Novel Grey Multivariable Model. *Journal of Cleaner Production*, *162*, 1527-1538. https://doi.org/10.1016/j.jclepro.2017.06.167

Faruque, Md. O., Rabby, Md. A. J., Hossain, Md. A., Islam, Md. R., Rashid, M. M. U., & Muyeen, S. M. (2022). A Comparative Analysis to Forecast Carbon Dioxide Emissions. *Energy Reports*, *8*, 8046-8060. https://doi.org/10.1016/j.egyr.2022.06.025

Houghton, R. A. (2003). Revised Estimates of the Annual Net Flux of Carbon to the Atmosphere from Changes in Land use and Land Management 1850-2000. *Tellus B*, *55*(2), 378-390. https://doi.org/10.1034/j.1600-0889.2003.01450.x

Huang, S., Xiao, X., & Guo, H. (2022). A Novel Method for Carbon Emission Forecasting Based on EKC Hypothesis and Nonlinear Multivariate Grey Model: Evidence from Transportation Sector. *Environmental Science and Pollution Research*, *29*(40), 60687-60711. https://doi.org/10.1007/s11356-022-20120-5

Jamel, L., & Derbali, A. (2016). Do Energy Consumption and Economic Growth Lead to Environmental Degradation? Evidence from Asian Economies. *Cogent Economics & Finance*, *4*(1), 1170653. https://doi.org/10.1080/23322039.2016.1170653

Javanmard, M. E., & Ghaderi, S. F. (2022). A Hybrid Model with Applying Machine Learning Algorithms and Optimization Model to Forecast Greenhouse Gas Emissions with Energy Market Data. *Sustainable Cities and Society*, *82*, 103886. https://doi.org/10.1016/j.scs.2022.103886

Karunasingha, D. S. K. (2022). Root Mean Square Error or Mean Absolute Error? Use their Ratio as Well. *Information Sciences*, *585*, 609-629. https://doi.org/10.1016/j.ins.2021.11.036

Khan, H., Weili, L., & Khan, I. (2022). Institutional quality, Financial Development and the Influence of Environmental Factors on Carbon Emissions: Evidence From a Global Perspective. *Environmental Science and Pollution Research*, *29*(9), 13356-13368. https://doi.org/10.1007/s11356-021-16626-z

Köne, A. Ç., & Büke, T. (2010). Forecasting of CO_2 Emissions From Fuel Combustion Using Trend Analysis. *Renewable and Sustainable Energy Reviews*, *14*(9), 2906-2915. https://doi.org/10.1016/j.rser.2010.06.006

Konya, A., & Nematzadeh, P. (2024). Recent Applications of AI to Environmental Disciplines: A Review. *Science of The Total Environment*, *906*, 167705. https://doi.org/10.1016/j.scitotenv.2023.167705

Le Quéré, C., Jackson, R. B., Jones, M. W., Smith, A. J. P., Abernethy, S., Andrew, R. M., De-Gol, A. J., Willis, D. R., Shan, Y., Canadell, J. G., Friedlingstein, P., Creutzig, F., & Peters, G. P. (2020). Temporary Reduction in Daily Global CO_2 Emissions During the COVID-19 Forced Confinement. *Nature Climate Change*, *10*(7), 647-653. https://doi.org/10.1038/s41558-020-0797-x

Libao, Y., Tingting, Y., Jielian, Z., Guicai, L., Yanfen, L., & Xiaoqian, M. (2017). Prediction of CO 2 Emissions Based on Multiple Linear Regression Analysis. *Energy Procedia*, *105*, 4222-4228. https://doi.org/10.1016/j.egypro.2017.03.906

Lin, C.-C., He, R.-X., & Liu, W.-Y. (2018). Considering Multiple Factors to Forecast CO_2 Emissions: A Hybrid Multivariable Grey Forecasting and Genetic Programming Approach. *Energies*, *11*(12), 3432. https://doi.org/10.3390/en11123432

Linardatos, P., Papastefanopoulos, V., Panagiotakopoulos, T., & Kotsiantis, S. (2023). CO_2 Concentration Forecasting in Smart Cities Using a Hybrid ARIMA-TFT Model on Multivariate Time Series IoT Data. *Scientific Reports*, *13*(1), 17266. https://doi.org/10.1038/s41598-023-42346-0

Liu, Z., Ciais, P., Deng, Z., Lei, R., Davis, S. J., Feng, S., Zheng, B., Cui, D., Dou, X., Zhu, B., Guo, R., Ke, P., Sun, T., Lu, C., He, P., Wang, Y., Yue, X., Wang, Y., Lei, Y., … Schellnhuber, H. J. (2020). Near-real-time monitoring of global CO_2 emissions reveals the effects of the COVID-19 pandemic. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18922-7

Masry, E. (1996). Multivariate Regression Estimation Local Polynomial Fitting for Time Series. *Stochastic Processes and Their Applications*, *65*(1), 81-101. https://doi.org/10.1016/s0304-4149(96)00095-6

Ozturk, I., & Acaravci, A. (2013). The Long-Run and Causal Analysis of Energy, Growth, Openness and Financial Development on Carbon Emissions in Turkey. *Energy Economics*, *36*, 262-267. https://doi.org/10.1016/j.eneco.2012.08.025

Prakash, A., & Singh, S. K. (2023). CO_2 emissions prediction from coal used in power plants using univariate and multivariate machine learning models. *Research Square*. https://doi.org/10.21203/rs.3.rs-3663119/v1

Ritchie, H., Roser, M., & Rosado, P. (2022). CO_2 and greenhouse gas emissions. *Our World in Data*. https://ourworldindata.org/CO_2-and-greenhouse-gas-emissions.

Salari, M., Javid, R. J., & Noghanibehambari, H. (2021). The Nexus Between CO_2 Emissions, Energy Consumption and Economic Growth in the U.S. *Economic Analysis and Policy*, *69*, 182-194. https://doi.org/10.1016/j.eap.2020.12.007

Saleh, C., Dzakiyullah, N. R., & Nugroho, J. B. (2016). Carbon Dioxide Emission Prediction Using Support Vector Machine. *IOP Conference Series: Materials Science and Engineering*, *114*, 012148. https://doi.org/10.1088/1757-899x/114/1/012148

Wang, Z.-X., & Ye, D.-J. (2017). Forecasting Chinese Carbon Emissions from Fossil Energy Consumption Using Non-Linear Grey Multivariable Models. *Journal of Cleaner Production*, *142*, 600-612. https://doi.org/10.1016/j.jclepro.2016.08.067

Wei, S., Yuwei, W., & Chongchong, Z. (2018). Forecasting CO_2 Emissions in Hebei, China, Through Moth-Flame Optimization Based on the Random Forest and Extreme Learning Machine. *Environmental Science and Pollution Research*, *25*(29), 28985-28997. https://doi.org/10.1007/s11356-018-2738-z

Wen, L., & Cao, Y. (2020). Influencing Factors Analysis and Forecasting of Residential Energy-Related CO_2 Emissions Utilizing Optimized Support Vector Machine. *Journal of Cleaner Production*, *250*, 119492. https://doi.org/10.1016/j.jclepro.2019.119492

Wu, L., Liu, S., Liu, D., Fang, Z., & Xu, H. (2015). Modelling and Forecasting CO 2 Emissions in the BRICS (Brazil, Russia, India, China and South Africa) Countries Using a Novel Multi-Variable Grey Model. *Energy*, *79*, 489-495. https://doi.org/10.1016/j.energy.2014.11.052

Xiong, P., Xiao, L., Liu, Y., Yang, Z., Zhou, Y., & Cao, S. (2021). Forecasting Carbon Emissions Using a Multi-Variable GM (1,N) Model Based on Linear Time-Varying Parameters. *Journal of Intelligent & Fuzzy Systems*, *41*(6), 6137-6148. https://doi.org/10.3233/jifs-202711

Ye, L., Yang, D., Dang, Y., & Wang, J. (2022). An Enhanced Multivariable Dynamic Time-Delay Discrete Grey Forecasting Model for Predicting China's Carbon Emissions. *Energy*, *249*, 123681. https://doi.org/10.1016/j.energy.2022.123681

Zhang, C., & Lin, Y. (2012). Panel Estimation for Urbanization, Energy Consumption and CO_2 Emissions: A Regional Analysis in China. *Energy Policy*, *49*, 488-498. https://doi.org/10.1016/j.enpol.2012.06.048

Zhao, X., Han, M., Ding, L., & Calin, A. C. (2018). Forecasting Carbon Dioxide Emissions Based On a Hybrid of Mixed Data Sampling Regression Model and Back Propagation Neural Network in the USA. *Environmental Science and Pollution Research*, *25*(3), 2899-2910. https://doi.org/10.1007/s11356-017-0642-6