Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ25888

Time Series Forecasting of Carbon Dioxide Concentration Based on Machine Learning Method

Jiahan Weng

School of Data Science, Zhejiang University of Finance & Economics, Hangzhou, China
wjh110780627@163.com

Abstract: The rapid proliferation of industrial activities and urban development has markedly intensified anthropogenic climate change, underscoring the critical need for effective mitigation strategies targeting greenhouse gas emissions. While most existing carbon emission forecasts focus on quarterly or annual scales, accurate daily-level prediction is essential. This study utilizes daily atmospheric CO_2 concentration data from January 25, 2015, to February 13, 2025, to forecast future CO_2 levels. Two machine learning-based time series models—Prophet and Long Short-Term Memory (LSTM)—are employed to perform daily-scale carbon emission forecasting. Descriptive statistical analysis reveals a clear upward trend and seasonal fluctuations in the data. Results indicate that while the Prophet model performs well on the training set, it exhibits limited generalization ability on the test set. In contrast, the LSTM model performs better on the test set, showing stronger adaptability to unseen data. This study provides insights into CO2 concentration trends, aiding climate change assessment, informing emission mitigation strategies, and guiding the selection of time series forecasting models.

Keywords: Machine Learning, CO_2 concentration Prediction, Time Series, Prophet Model, LSTM Model

1. Introduction

Since 1750, rapid industrialization and urbanization have markedly elevated greenhouse gas (GHG) concentrations and emissions due to intensified human activities [1], drawing worldwide attention. Building on this trend, the mid-20th century witnessed an even more dramatic expansion of the global economy, predominantly driven by extensive energy consumption [2]. This surge in energy consumption has led to significant emissions of greenhouse gases, thereby exacerbating climate-related issues such as global warming, melting glaciers, sea-level rise, and extreme weather events [3]. Controlling greenhouse gas emissions and mitigating the greenhouse effect are essential for sustainable development and the ongoing enhancement of human productivity [4].

Carbon emission predictions are predominantly long-term (annual, quarterly, or monthly) and tend to provide trend-based or approximate estimations [5]. Keeping an accurate track of the same in similar metrics so they can be combined in an overall entity is an overwhelming task [6]. These long-term forecasting models often rely on aggregated data, overlooking the high-frequency fluctuations in daily emissions. They fail to effectively model short-term carbon emission patterns

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0
(https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.BJ25888

under varying conditions, such as different dates or climate scenarios, and are inadequate for supporting rapid decision-making in mechanisms like carbon trading, cap-and-trade, or peak emission controls, thus restricting their predictive accuracy.

Therefore, improving the granularity of carbon emission forecasts to the daily level is crucial for enhancing prediction accuracy and timely decision-making. This research applies time series analysis techniques, employing machine learning models, to conduct daily-scale forecasts of global carbon emissions based on a single variable: historical carbon emission data.

Traditional regression models often fail to capture the dynamic and time-variant nature of such data effectively. This research adopts the Prophet model and the Long Short-Term Memory (LSTM) neural network to address this limitation and predict future carbon emissions. The study assesses both models' performance, analyzing their strengths and limitations, and discusses future directions in carbon emission forecasting. Accurate predictions are crucial for policymakers to evaluate the effectiveness of energy-saving and emission-reduction strategies.

2. Dataset description

This study employs a dataset of daily atmospheric CO_2 concentrations from January 25, 2015, to February 13, 2025, obtained via the RapidAPI Daily Atmosphere Carbon Dioxide Concentration API, ensuring data accuracy and reliability.

Table 1 presents the basic descriptive statistics of the seasonal cycle in CO_2 concentration (measured in parts per million, ppm) during this period.

Table 1. Basic descriptive statistics

	CO_2 concentration
Count	3673
mean	411.61
Std	7.38
min	396.50
25%	405.68
50%	411.61
75%	417.61
max	426.77

From the calculation of various descriptive statistics of this dataset, it can be analyzed and concluded that the data volume of CO_2 concentration is 3673, with a mean of 411.61ppm, and the average and median are equal, indicating that the data distribution is symmetrical. The standard deviation is 7.38 ppm, indicating the degree of fluctuation of the data points around the average value. The range between the maximum and minimum values is 30.27 ppm, showing the overall variation range of the data.

Figure 1 shows the time series plot and the time series decomposition of CO_2 concentrations：

DOI: 10.54254/2755-2721/2025.BJ25888

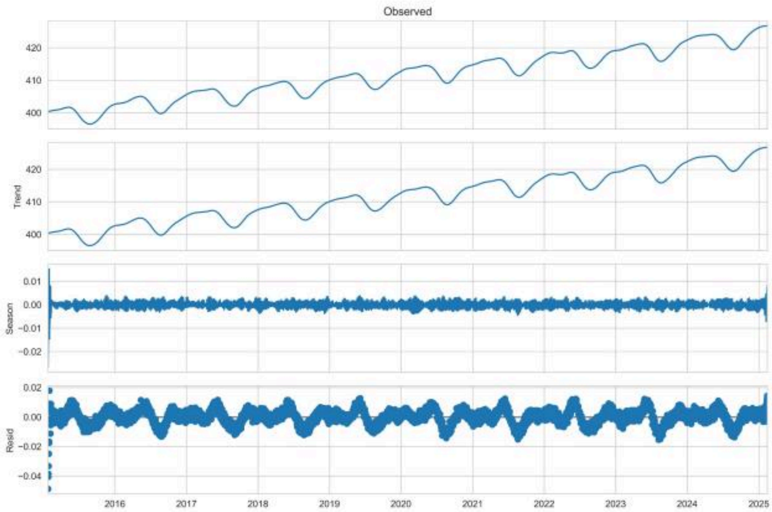

Figure 1. The time series decomposition plot of CO_2 concentration

The observed time series exhibits a clear upward trend, accompanied by distinct seasonal fluctuations in CO_2 concentrations. The second plot, representing the trend component, demonstrates a rising trajectory that aligns closely with the global context of continuously increasing atmospheric CO_2 levels. The seasonal component displays relatively minor fluctuations centered around zero, indicating regular and predictable seasonal variation. The residual component captures the random variations remaining after removing trend and seasonal influences. The residuals fluctuate within a specific range, suggesting that some stochastic factors continue to affect the time series even after accounting for trend and seasonality. All four of these figures offer valuable insight into selecting appropriate forecasting models.

3. Carbon concentration prediction based on machine learning

3.1. Prophet model

3.1.1. Introduction

This study employs the Prophet model to capture both the trend and seasonality observed in the CO_2 concentration time series. Facebook developed this model, which is available in Python and R [7]. Prophet is well-suited for datasets with strong seasonal effects and historical data spanning several seasons, and it is robust to missing data and outliers. The main reasons for building this model are its three main features, i.e., trend, seasonality, holidays [8] and demand for high-quality forecasting.

DOI: 10.54254/2755-2721/2025.BJ25888

3.1.2. Application

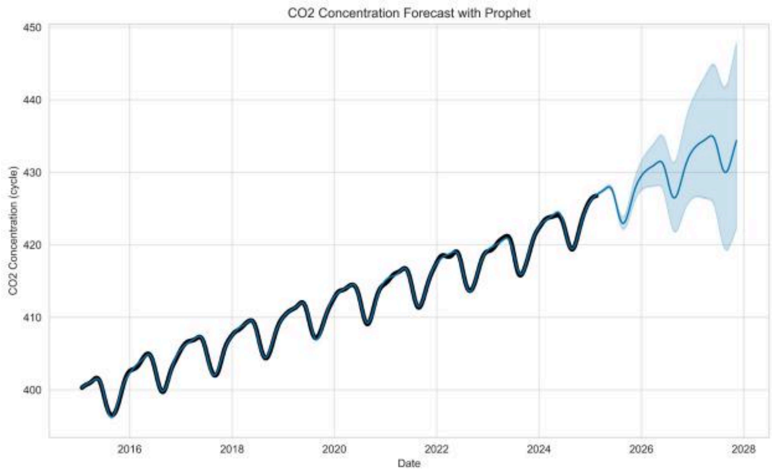

Figure 2. Prophet model fitting curve graph

This study employed the Prophet model to fit and forecast the CO_2 concentration time series during the modelling process. The model was used to extract the data's trend, seasonality, and residual components, which served as the basis for future concentration predictions.

Based on the fitting results of historical data (approximately from 2016 to 2024), the Prophet model effectively captured the overall trend and seasonal variation in the CO_2 concentration time series. This demonstrates the model's strong capability in handling time series data characterized by trend and seasonality.

The analysis reveals a clear upward trend in CO_2 concentration over time, indicating a continuous increase in atmospheric CO_2 levels. In addition, the presence of regular fluctuations in the curve reflects the seasonal nature of CO_2 concentration changes. The shaded blue area in the forecast plot represents the prediction interval (confidence interval). It is evident that as the forecast extends further into the future, the prediction interval becomes wider, suggesting that the model's confidence in long-term forecasts diminishes due to the increasing influence of unknown or unpredictable factors.

3.2. LSTM model

3.2.1. Introduction

This study also utilizes an LSTM neural network to enhance CO_2 forecasting accuracy and model non-linear patterns. RNN-based LSTM variants address vanishing gradient issues [9], facilitating long-sequence data processing and prediction [10].

3.2.2. Application

In applying the LSTM model, this study first normalized the CO_2 concentration time series data and constructed input-output pairs using a sliding window approach. A 10-day time step was employed, utilizing the preceding 10 days' data as input features to enhance the model's capacity to capture complex temporal dependencies. The dataset was partitioned into training and testing sets at an 80:20 ratio for model training and performance assessment.

DOI: 10.54254/2755-2721/2025.BJ25888

The LSTM network was composed of three stacked LSTM layers, each with 50 units, to enhance the model's learning capacity. A dense output layer was incorporated for single-step prediction. The model utilized the Adam optimizer and minimized MSE loss, trained over 100 epochs with batch size 64 to effectively capture sequential data patterns.

After training, the model generated predictions for the training and testing datasets. The outputs were then inverse-transformed to return to the original scale of the CO_2 concentration values. The model prediction graph is shown as follows:

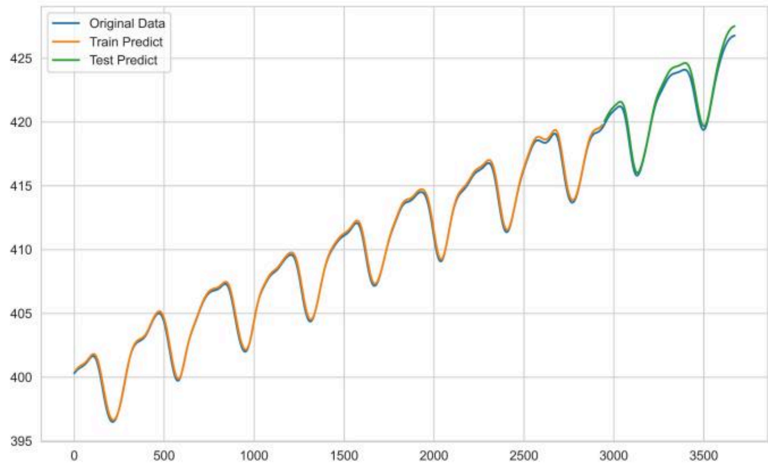

Figure 3. LSTM model fitting curve graph

The visualization shows that all three lines—the original data (blue), the training set predictions (orange), and the test set predictions (green)—exhibit an overall upward trend, indicating that the studied time series demonstrates consistent growth over time. The orange training prediction curve closely follows the trajectory of the original data, suggesting that the LSTM model achieves a good fit on the training set.

The green test prediction curve mirrors the original and training curves initially, but diverges after 2500 data points. While deviation increases, the upward trend is maintained, indicating the model's ability to capture the data's general direction.

The LSTM model's nonlinear modeling and long-term memory improve prediction of complex time series data like CO_2 concentrations. Compared to linear models, it better captures nonlinear dynamics, offering a more accurate tool for atmospheric CO_2 forecasting.

4. Results and discussion

In this study, the Prophet model and the Long Short-Term Memory (LSTM) neural network were employed to model and forecast the time series data of atmospheric CO_2 concentration. The Root Mean Square Error (RMSE) was used as the assessment metric for both the training and testing datasets to evaluate the predictive performance of the two models. The evaluation results are presented in the table below:

Table 2. Comparison of fitting effects of the two models

	Prophet Model	LSTM Model
Training set RMSE	0.148	0.162
Testing set RMSE	1.614	0.139

DOI: 10.54254/2755-2721/2025.BJ25888

From the RMSE values of the training set, the Prophet model demonstrates slightly lower error than the LSTM model, indicating better fitting performance on historical data. This is especially evident in its ability to capture both trend and seasonality, as observed in the visualization from Chapter 3. The Prophet model accurately reflects the upward trend and seasonal fluctuations of CO_2 concentration between 2016 and 2024, and the generated confidence intervals provide insight into the uncertainty of future predictions.

However, in the testing set, the LSTM model significantly outperforms the Prophet model regarding RMSE, suggesting stronger generalization ability when dealing with unseen future data. This advantage stems from LSTM's capability to learn nonlinear patterns and long-term dependencies in time series data. As shown in the prediction plots, the LSTM model closely tracks the original data's trend in the test set's early stages. Although the deviation increases after approximately data points 2500, the predicted curve still follows an overall upward trend, indicating that the model captures the main direction of the data to a considerable extent.

Compared to these two models, it is clear that the Prophet model, with its interpretable components and robustness to missing or noisy data, is suitable for short-term predictions with clear seasonal and trend patterns, especially when the data volume is limited. In contrast, the LSTM model exhibits stronger generalization and nonlinear learning capabilities, making it more appropriate for long-term forecasting tasks with sufficient high-quality data.

In conclusion, the Prophet model is well-suited for time series with strong trends and seasonal components and offers good interpretability and robustness. In contrast, the LSTM model, with its deep learning architecture and ability to model complex nonlinear dynamics, demonstrates superior forecasting accuracy on the test set. Therefore, for long-term forecasting of CO_2 concentrations, a hybrid modeling approach that integrates traditional time series models like Prophet with deep learning models such as LSTM could be considered to improve prediction accuracy and model robustness.

5. Conclusion

This study analyzed a dataset containing daily atmospheric carbon dioxide (CO_2) concentrations from January 25, 2015, to February 13, 2025, by conducting descriptive statistical analysis and applying machine learning-based time series forecasting models. Firstly, the time series exhibits a clear upward trend and seasonal variation, indicating that it is well-suited for time series modeling and forecasting. The Prophet model accurately captures CO_2 concentration trends and seasonality, with a training RMSE of 0.148. However, its test RMSE of 1.614 indicates poor generalization. Conversely, the LSTM model, adept at nonlinear patterns, showed superior performance, achieving training and test RMSEs of 0.162 and 0.139, respectively, suggesting it is more robust for forecasting atmospheric CO2 concentrations.

This study uses Prophet and LSTM models to offer novel insights into atmospheric CO_2 patterns. Precise CO2 forecasts are essential for evaluating climate trends and guiding effective carbon emission policies. However, the Prophet model's suboptimal test set performance reveals generalization and adaptability constraints, necessitating forecasting enhancements. Future studies could explore hybrid models, integrating Prophet and LSTM with ARIMA or Transformer-based methods to boost prediction accuracy and robustness. These ensemble strategies may offer more dependable tools for CO_2 trend analysis and climate decisions.

DOI: 10.54254/2755-2721/2025.BJ25888

References

[1] Cui, Y., Su, W., Xing, Y., Hao, L., Sun, Y., & Cai, Y. (2023). Experimental and simulation evaluation of CO2/CO separation under different component ratios in blast furnace gas on zeolites. Chemical Engineering Journal, 472, 144579.
[2] Xiao, Y., Ma, D., Zhang, F., Zhao, N., Wang, L., Guo, Z., ... & Xiao, Y. (2023). Spatiotemporal differentiation of carbon emission efficiency and influencing factors: From the perspective of 136 countries. Science of the Total Environment, 879, 163032.
[3] Sun, W., & Ren, C. (2021). Short-term prediction of carbon emissions based on the EEMD-PSOBP model. Environmental Science and Pollution Research, 28(40), 56580-56594.
[4] Sun, H., & Park, Y. (2020). CO2 emission calculation method during construction process for developing BIM-based performance evaluation system. Applied Sciences, 10(16), 5587.
[5] Zhong, W., Zhai, D., Xu, W., Gong, W., Yan, C., Zhang, Y., & Qi, L. (2024). Accurate and efficient daily carbon emission forecasting based on improved ARIMA. Applied Energy, 376, 124232.
[6] McMahon, J. (2019). Why agriculture’s greenhouse gas emissions are almost always underestimated. Forbes, December 2.
[7] S. Taylor. (2019, May 14) prophet: Automatic forecasting procedure. [Online]. Available: https: //cran.rproject.org/web/packages/prophet/
[8] Facebook Research. (2017, Feb 23) Prophet: forecasting at scale. [Online]. Available: https: //research.fb.com/prophetforecasting-at-scale/
[9] Chalapathy, R., Khoa, N. L. D., & Sethuvenkatraman, S. (2021). Comparing multi-step ahead building cooling load prediction using shallow machine learning and deep learning models. Sustainable Energy, Grids and Networks, 28, 100543.
[10] Jang, J., Han, J., & Leigh, S. B. (2022). Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks. Energy and Buildings, 255, 111647.