Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 9, September 2024

Forecasting Future CO2 Levels (Ppm) using the SARIMAX Model

Trivikram Sai Krovi
Department of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

Ashika S S
Department of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

Jenifer Shanmugasundaram
Department of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

Sushmithaasri K N
Department of Computer Science and Engineering
Vellore Institute of Technology
Vellore, India

***Abstract*—Climate change poses one of the most significant challenges of our time, affecting ecosystems, human health, and economies globally. The increasing concentration of greenhouse gases, particularly carbon dioxide (CO2), has led to unprecedented global warming and climate disruptions. To combat these effects, it is imperative to develop innovative strategies that not only reduce emissions but also enhance our ability to adapt to changing climate conditions.**

Artificial intelligence (AI) has emerged as a powerful tool in this endeavor, offering advanced capabilities in data analysis, predictive modeling, and real-time monitoring. This study presents a comprehensive analysis of historical carbon dioxide (CO2) levels using a dataset comprising monthly average CO2 mole fractions from March 1958 to the present. A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) model was employed to forecast future CO2 levels. The SARIMAX model's suitability for capturing seasonal variations and trends in time series data was leveraged to make accurate predictions. This research highlights the importance of historical data analysis in understanding and predicting CO2 trends, contributing valuable insights for climate change studies and policy-making.

***Keywords*—Artificial Intelligence, SARIMAX model, CO2 levels, climate change, predictive modeling, real-time monitoring**

I. INTRODUCTION

The ever-increasing climate change crisis is one of the most daunting challenges of our time, requiring immediate attention and innovative solutions. Climate change, mostly caused by human-generated greenhouse gas emissions, manifests as more frequent and severe weather events, disrupting ecosystems and having serious effects for human health and infrastructure. In order to prevent catastrophic results, the Intergovernmental Panel on Climate Change (IPCC) has emphasized the significance of significant global efforts to reduce emissions and ameliorate the effects of climate change.

AI is a powerful ally in the battle against climate change. We can enhance our ability to analyze large datasets, optimize resource utilization and develop predictive models to aid in policy and decision-making processes by harnessing AI technologies. AI's potential applications in combating climate change are wide-ranging, including but not limited to improving energy efficiency, integrating renewable energy, expanding climate modeling, and environmental monitoring.

The incorporation of AI into climate change mitigationefforts offers both significant prospects and challenges. On the one hand, AI can drive advances in energy-efficient technologies, improve industrial processes, and enable precision agriculture, cutting emissions across industries. On the other hand, AI deployment must be carefully managed to minimize its environmental impact and ethical problems.

Forecasting CO2 levels is a critical component in the fight against climate change. Accurate predictions of CO2 levels enable policymakers, researchers, and environmental organizations to make informed decisions regarding emissions reduction strategies and climate change mitigation efforts. Understanding future CO2 trends is essential for developing effective policies, setting realistic targets, and implementing timely interventions to curb greenhouse gas emissions.

The Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) model plays a pivotal role in this endeavor. SARIMAX is a sophisticated statistical method designed to handle time series data, especially those exhibiting seasonal patterns and trends. It is defined by parameters (p, d, q) for non-seasonal autoregressive, differencing, and moving average orders, and (P, D, Q, s) for seasonal counterparts and the seasonal period. The model captures complex patterns by considering both short-term and long-term trends along with seasonality. By leveraging the SARIMAX model, we can capture and analyze the seasonal fluctuations and long-term trends in CO2 levels, leading to more accurate and reliable forecasts.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 9, September 2024

The SARIMAX model's ability to decompose time series data into seasonal, trend, and noise components allows us to gain deeper insights into the underlying factors driving CO2 variations. This decomposition is crucial for identifying the seasonal peaks and troughs in CO2 emissions, which can be linked to specific human activities or natural processes. By understanding these patterns, we can better anticipate periods of high emissions and take preemptive measures to mitigate their impact.

Furthermore, the SARIMAX model aids in the evaluation of historical data, providing a solid foundation for projecting future CO2 levels. This historical analysis is essential for recognizing past trends, assessing the effectiveness of previous policies, and identifying areas where further action is needed. By combining historical data with advanced forecasting techniques, the SARIMAX model helps us build a comprehensive picture of future CO2 scenarios, informing long-term climate strategies. Thus, contributing to a more sustainable and resilient future.

Another statistical method we have used for the time series forecasting is the ARIMA model which stands for Autoregressive Integrated Moving Average and is useful due to its efficient handling of different standard temporal structures present in time series data.

II. LITERATURE SURVEY

Climate change is a multifaceted problem requiring diverse and innovative approaches for effective mitigation. This Literature Review comprises a different range of papers based on Artificial Intelligence in the mitigation of climate change. In "Artificial Intelligence (AI) and the Prediction of Climate Change Impacts" (Mankala Satish et al.,2023) a simple climate model linear equation was formulated to predict temperature changes using AI and ML algoritms which was then concluded to be foundational and could not be employed for the several implications and complexities of real-world climate changes. In "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations' '(Josh Cowls et al.,2021) mentions the possible challenges of AI deployment exacerbating existing social and ethical issues. The carbon footprint of research in AI influences GHG emissions estimated by tools like experiment-impact-tracker and ML Emissions Calculator. In "AI-Based Campus Energy Use Prediction for Assessing the Effects of Climate Change " (Soheil Fathi et al.,2020) a campus-scale energy use prediction tool was developed for prediction of long-term climate changes in the campus using AI techniques which had four steps. The research could be improved if it could obtain more building data and improve accuracy to provide building energy use prediction for various climate scenarios. The "Artificial neural networks in drought prediction in the 21st century–A scientometric analysis" (Abhirup Dikshit et al.,2021) used artificial neural networks for drought prediction which could be improved by using deep and graph neural networks, and deep learning for interpretability. The "Deploying Artificial Intelligence for Optimized Flood Forecasting and Mitigation" (Mohammad Algarni et al.,,2023) discusses predicting and managing floods with the help of satellite imagery and IoT sensors to enhance the accuracy of AI based predictions. Its limitations included Data Integrity, Computational Intensity, and Integration with Existing Systems. In "Human-AI Symbiosis: Decode Climate Change to Prevent Heat-Related Mortalities and to Protect Our Most Vulnerable Population" (Anitha Ilapakurti et al.,2019) utilizes Electronic Health Records (EHR) data to identify senior citizens who are susceptible to heat waves to prevent medical complexities and death. Every research paper here provides significant information regarding AI in climate change mitigation.

III. PROPOSED METHODOLOGY

A. Dataset

The dataset utilized in this study reports the dry air mole fraction of carbon dioxide (CO2) in parts per million (ppm). This fraction is calculated as the number of CO2 molecules divided by the total number of molecules in air, including CO2 itself, after water vapor has been removed. For example, a mole fraction of 0.000400 is expressed as 400 ppm.

The data includes several attributes: the Date, representing the month and year when the measurement was taken, and the Decimal Date, which is the date in decimal form for computational convenience. The Average column contains the average CO2 mole fraction for each month, expressed in ppm. The Interpolated column provides interpolated values to fill gaps where data might be missing, while the Trend column shows the long-term trend component of the CO2 data. The Number of Days column indicates the number of days in the month for which data was available. Additional parsed_extra column includes parsed information that was not used in the analysis.

To ensure the quality and reliability of the data for modeling, several preprocessing steps were undertaken. Any null values in the dataset were identified and removed to avoid potential biases and errors in the analysis. The Average column, representing the average CO2 mole fraction for each month, was selected for model training due to its completeness and relevance to predicting future CO2 levels.

The prepared dataset contains a total of 792 rows after preprocessing. This historical CO2 data was used to train a Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) model. The SARIMAX algorithm was chosen for its ability to handle seasonal variations and trends in time series data, making it suitable for predicting future CO2 levels.

The SARIMAX model was trained using the Average column from the dataset, which provided a robust basis for understanding and forecasting the future levels of CO2. This approach allowed us to capture the seasonal patterns and trends inherent in the historical data, thus enabling more accurate predictions.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 9, September 2024

B. Methodology

We first Visualize the data as follows:

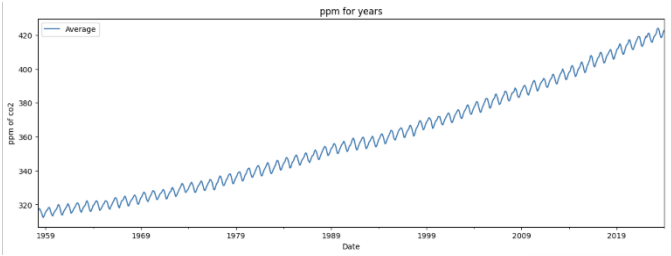

Figure 1: plot of the average levels of ppm levels for co2 from 1959 to 2004

As we have observed from the dataset, we need a methodology to predict the future levels of co2 in the atmosphere. In order to do this, we make use of the time series concept in machine learning. The main objective of time seriesanalysis is predicting the future while training it on the presentdata on hand, which the predicted data can be used by various industries to help predict their future outcomes or to prevent future collapses based on the future predictions.

We are predicting thefuture CO2 levels of the atmosphere.

Two models related to time series forecasting in this paper namely, “ARIMA” and the “SARIMAX” model. The Arima model needs nonstationary data, while the SARIMAX model uses the Seasonal data.

We can check the Seasonality of the data using the stationarity test, here the ADF (Augmented Dicky - Fuller) test.When we apply for this test, we get the values of p as 1.0. So, this isnon-Stationary as for it to be stationary the p value should be less than 0.05. So, we apply the stationarity tests for it but shifting the order by a value of 12 (this 12 indicates the data has a 12-month duration in a year) and when we apply the test on the new data again, it indeed shows the data is stationary now.

But the ARIMA model doesn't give a proper prediction to the data on hand and gives an inaccurate prediction.

So, we turned to the SARIMAX model. The main advantageof the SARIMAX model is that it doesn't depend on the Seasonality of the data as mentioned before. That means we don't need to convert to stationary data. We can just give the data to the model, and it will be able to give us the required output.

An important tool we used here is auto Arima from the pmdarima library. With the Advancements in the Machine Learning process, instead of manually testing the various p q and d values, we can use the auto_arima code to run and generate all the various combinations in the SARIMAX model. It then gives us the most accurate model parameters to run in the model. We get the parameters for this, using the summary to get the information .

SARIMAX Results

Dep. Variable:	Average	No. Observations:	632
Model:	SARIMAX(2, 1, 1)x(1, 0, 1, 12)	Log Likelihood	-166.909
Date:	Sun, 21 Jul 2024	AIC	345.818
Time:	16:22:39	BIC	372.502
Sample:	03-01-1958 - 10-01-2010	HQIC	356.182
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4003	0.040	10.050	0.000	0.322	0.478
ar.L2	0.0991	0.038	2.587	0.010	0.024	0.174
ma.L1	-0.7332	0.038	-19.497	0.000	-0.807	-0.660
ar.S.L12	0.9997	0.000	3112.585	0.000	0.999	1.000
ma.S.L12	-0.8753	0.022	-38.918	0.000	-0.919	-0.831
sigma2	0.0911	0.005	18.306	0.000	0.081	0.101

Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	0.58
Prob(Q):	0.79	Prob(JB):	0.75
Heteroskedasticity (H):	1.05	Skew:	-0.04
Prob(H) (two-sided):	0.73	Kurtosis:	3.13

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

Figure 2: Result obtained after executing the sarimax model on the chosen dataset

We get the parameter values from this dataset namely

- p = 2 (Auto Regressive Component)
- q = 1 (Moving Average Component)
- d = 1 (Integrated Component)
- P = 1 (Seasonal Auto Regressive Component)
- D = 0 (Seasonal Integrated Component)
- Q = 1 (Seasonal Moving Average Component)
- s = 12 (Seasonal Period)

Now we split the data into train and test where 80% is for train and 20% is for test. We then train the train part using he model we made and then we apply this model to the testdataset. We then plot the test dataset and the actual dataset which when plotted we get a pretty accurate plot

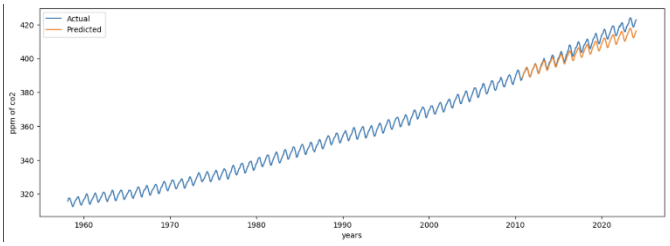

Figure 3: plot of the taken dataset with the predicted data using the sarimax model

As now we get a proper prediction, we then make the future predictions using the model on hand. We make the a few future dates and then assign then the values according to the time series forecast of the SARIMAX model.

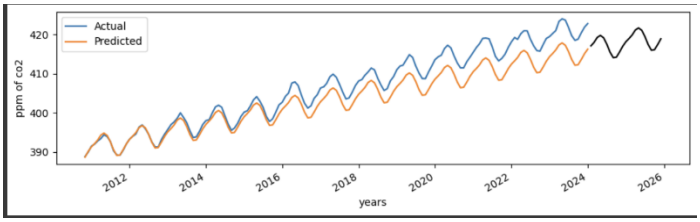

Figure 4: plot of the chosen dataset and the predicted data from sarimax model with the future predictions of the same predicted data

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 9, September 2024

IV. RESULTS AND DISCUSSION

We tried applying both SARIMAX and ARIMA for the CO2 forecasting. From our experiments, we found that ARIMA doesn't give accurate predictions, with SARIMAX proving to be notably better.The application of the SARIMAX(Seasonal Autoregressive Integrated Moving Average with Exogenous Factors) model for forecasting has demonstrated its effectiveness in capturing the underlying patterns and seasonality in the historical data. Through careful preprocessing and differencing, the data was made stationary, which is a prerequisite for effective time series modeling.

The SARIMAX model with parameters (2,1,1)x(1,0,1,12) was fitted to the training dataset and subsequently validated using a test dataset. The model's predictions showed a good fitwith the actual observed values, indicating that it successfully captured the seasonal trends and cyclical behavior of CO2 levels over time. Additionally, the model successfully forecasted future CO2 levels up to December 2025, providing valuable insights into potential future trends.

V. CONCLUSION AND FUTURE PROSPECTS

In our study on CO2 forecasting using the SARIMAX and ARIMA models, we showed that SARIMAX and ARIMA models can capture historical CO2 data trends and seasonality in our CO2 forecasting investigation. Both models successfully predicted future CO2 levels after thorough adjustment and verification. Through this we have achieved the main objective of our research.

In the future, we plan to improve the accuracy of SARIMAX and ARIMA models by experimenting with different seasonal and non-seasonal parameters. We will consider incorporating additional exogenous variables, such as economic indicators, industrial activity data, and policy changes, to improve predictive power. Extending the forecast horizon beyond 2025 will provide long-term insights crucial for policy planning and climate action strategi

To capture subtle data patterns and provide more reliable predictions, we plan to integrate these models with Prophet, LSTM, or other deep learning methods. Our objective is to distribute the findings of our CO2 prediction to the general public in order to increase understanding of the current patterns in greenhouse gas discharges and the critical need to tackle climate change. Our model's projections can be utilized by educational campaigns to demonstrate hypothetical future scenarios and underscore the significance of sustainable behaviors.

REFERENCES

[1] M. Satish, Prakash, S. M. Babu, P. P. Kumar, S. Devi and K. P. Reddy, "Artificial Intelligence (AI) and the Prediction of Climate Change Impacts," 2023 IEEE 5th International Conference on Cybernetics, Cognition andMachine Learning Applications (ICCCMLA), 2023, pp. 660- 664.

[2] J. Cowls, A. Tsamados, M. Taddeo and L. Floridi, "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations.", Ai& Society, 2021, pp. 1-25.

[3] S. Fathi, R. S. Srinivasan, C. J. Kibert, R. L. Steiner and E. Demirezen, "AI-based campus energy use prediction for assessing the effects of climate change", Sustainability, vol.12,no.8. 2020, pp. 3223.

[4] A. Dikshit, B. Pradhan and M. Santosh, "Artificialneural networks in drought prediction in the 21st century–A scientometric analysis." Applied Soft Computing, vol.114, 2022, pp. 108080.

[5] M. Algarni, "Deploying Artificial Intelligence for Optimized Flood Forecasting and Mitigation", 2023 20th ACS/IEEE International Conference on Computer Systems andApplications (AICCSA), IEEE, 2023, pp.1-6.

[6] A. Ilapakurti, S. Kedari, R. Vuppalapati, S. Kedari, J. S.Vuppalapati and C. Vuppalapati, "Human-AI Symbiosis: Decode Climate Change to Prevent Heat-Related Mortalities and to Protect Our Most Vulnerable Population", 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), IEEE, 2019, pp.331-338.