"Global Warming Time Series Analysis"

Environmental Side Effects and Key Contributors

Business, Economic, and Financial Data Course Final Project

Mohammad Matin Parvanian[†], Bahador Mirzazadeh[‡]

University of Padua

January 2025

[†] mohammadmatin.parvanian@studenti.unipd.it
[‡] bahador.mirzazadeh@studenti.unipd.it

Contents

Introduction

Glaciers are melting, sea levels are rising, precipitation is decreasing, and wildlife is struggling to keep pace. It has become clear that humans have caused most of the past century's warming by releasing heat-trapping gases as we power our modern lives. This global rise in the temperature of the air and oceans is called **global warming**.

In this project, we present a comprehensive time series analysis of global warming, its impact on the environment, and the main countries contributing to this phenomenon. We have considered different datasets: Global CO_2 Concentration, Contributors to Global Temperature Changes, Sea Level, and Precipitation Changes. All of these datasets were downloaded from Our World in Data[1], which provides a variety of datasets for different tasks.

First, we will analyze how one of the most important greenhouse gases, carbon dioxide, has changed over the last centuries. Next, we will examine the country contributors and their share in the rise of greenhouse gases, the sea level trend, and the change in precipitation for contributing countries. For each part, we will present different time series models, choose the best model according to our selected metrics, and provide the relevant results for the dataset. Finally, in the last part, we will present a general conclusion based on all the results from the available datasets.

1 Identifying Seasonality, Trends, and Anomalies in CO_2

1.1 Analysis of CO_2

When people burn fossil fuels like coal, oil, and natural gas, this adds carbon dioxide to the air, since fossil fuels contain lots of carbon and burning means joining most of the atoms in the fuel with oxygen. When people cut down many trees (deforestation), less carbon dioxide is taken out of the atmosphere by those plants.

1.1.1 Explanatory Data Analysis

The amount of CO_2 released from fossil fuels accounts for the largest share of all greenhouse gases. This contributes to global warming, leading to rising temperatures, melting glaciers, rising sea levels, and other environmental changes, such as increasing ocean acidity, which causes species loss and disrupts their natural habitats. The target variable of our dataset is the monthly concentration of atmospheric Carbon Dioxide.

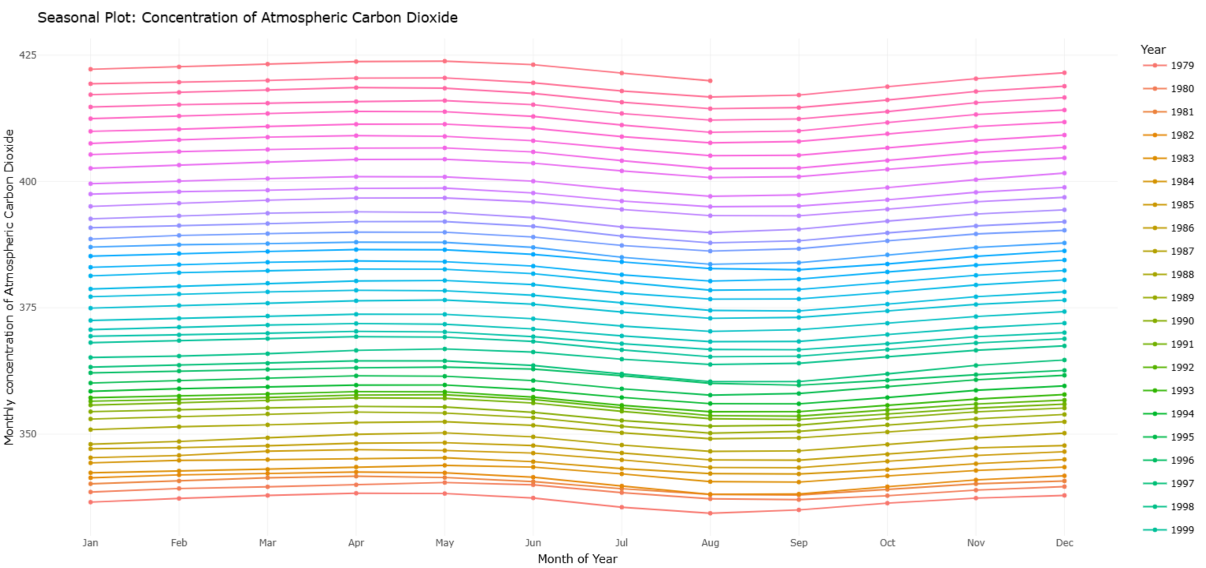

Figure 1: The seasonal plot for the monthly atmospheric carbon dioxide(1980 - 2023). For an interactive version of this plot, **click here**[2].

According to the seasonal plot above, we can draw two important conclusions. First and foremost, the seasonality is evident: the amount of CO_2 is higher during the colder months of the year and lower during the warmer months. This could be attributed to weather inversion. Additionally, a clear upward trend in CO_2 concentration can be observed. From 1979 to 2023, the levels of CO_2 have steadily increased, indicating a positive trend.

[1] `https://ourworldindata.org/`

[2] Using the Plotly library in RStudio and saving the plots in RPubs makes it possible to present interactive plots.

Figure 2: Decomposition of Observed Data: Trend, Seasonality, and Remainder. For an interactive version of this plot, **click here**.

Moreover, we can verify our previous results using the decomposition plot. The plot clearly shows the seasonality and the increasing trend in CO_2 levels over time. Additionally, the remainder of the components are present, It does not show any specific pattern however, the residuals plot will be analyzed for each model afterward.

1.1.2 TSLM Model

For this model, we considered the trend and seasonality. Since our dataset is univariate, we did not include any other explanatory variables.

Set	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	R-squared
Training set	8.27×10^{-16}	2.5274	2.2707	-0.0020	0.6064	1.1984	0.9918	0.9896

Table 1: Accuracy Metrics of the TSLM Model

The model has demonstrated good performance; however, we need to evaluate its ability to predict future behavior and analyze the residuals to ensure reliability.

Based on Figure 3, the TSLM model captures the overall trend in atmospheric CO_2 concentration but fails to account for the seasonal variations observed in the data. The forecast for the next 12 months appears to extend this trend but does not include potential seasonal fluctuations, which may affect its accuracy. Next, we will perform a residual analysis to assess the model's performance in greater detail and identify any systematic patterns that may require further adjustment.

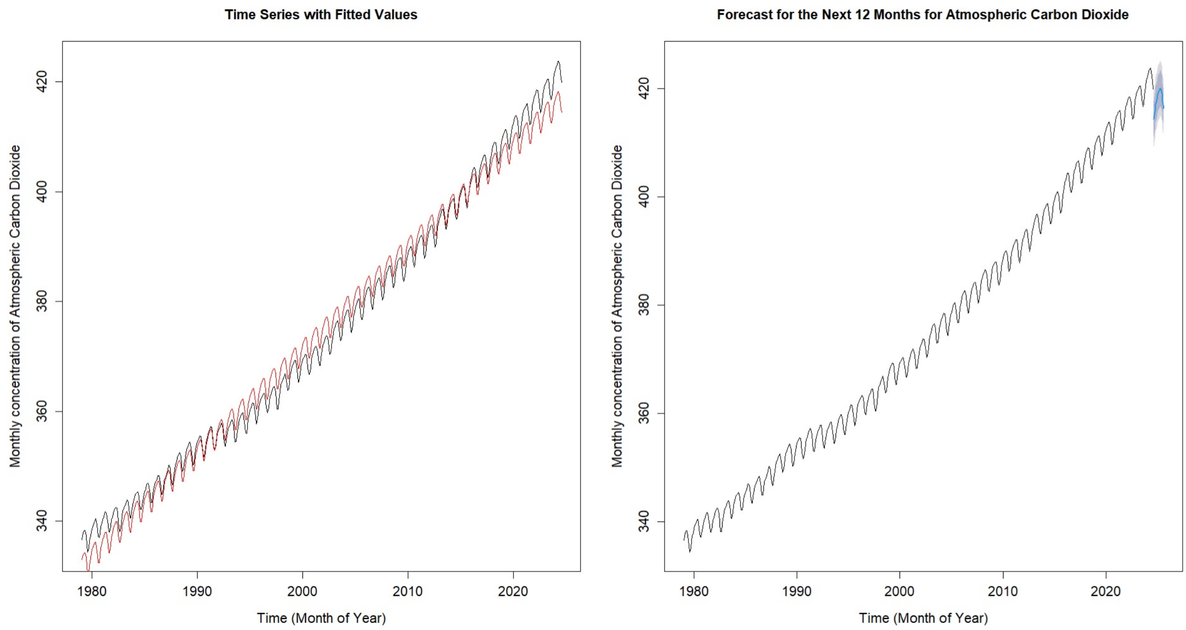

Figure 3: Historical Data with Fitted Values (left) and Future Forecast Based on the TSLM Model (right).

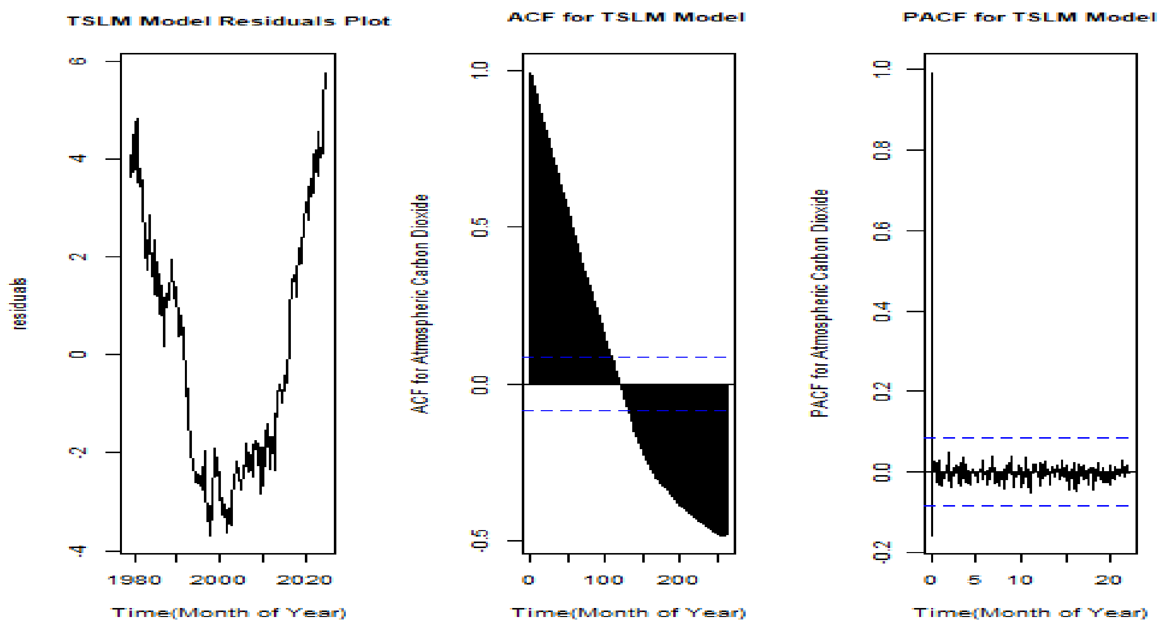

Figure 4: Residual Analysis for the TSLM Model: The Residual Plot (left), ACF Plot (center), and PACF Plot (right).

Test Name	Model Tested	DW Statistic	P-value	Alternative Hypothesis
Durbin-Watson Test	TSLM_model	0.0039946	$< 2.2 \times 10^{-16}$	True autocorrelation is greater than 0

Table 2: Durbin-Watson Test Results for TSLM Model Residuals for CO_2 Concentration

The residuals show a noticeable curved pattern over time, which indicates that the TSLM model isn't fully capturing some important aspects of the data, such as seasonality. Ideally, residuals should be randomly scattered around zero without any obvious patterns, which isn't the case here. So residuals are still dependent on their past values.

Additionally, the ACF plot shows significant autocorrelations at several lags, with values exceeding the blue confidence bands. This means the residuals are not behaving like white noise, suggesting there's still information in the data that the model hasn't accounted for.

On the other hand, the PACF plot doesn't show any significant values outside the confidence bands, which implies that higher-order dependencies might not be as important. That said, the ACF plot highlights that the model struggles with capturing relationships at lower lags, which remains a key issue.

The Durbin-Watson statistic indicates strong positive autocorrelation. The p-value is smaller than 0.05 which confirms that residuals are significantly correlated. This suggests the TSLM model is inadequate in capturing the data's structure. In this situation, we will apply other models that can handle autocorrelation.

1.1.3 Exponential Smoothing: Holt-Winters Method

In this section, we will evaluate two methods: the additive and multiplicative models. Since the trend in our data appears constant, we guess that the additive model will perform better. The multiplicative model is more appropriate when the trend and seasonal variations change proportionally with the level of the series.

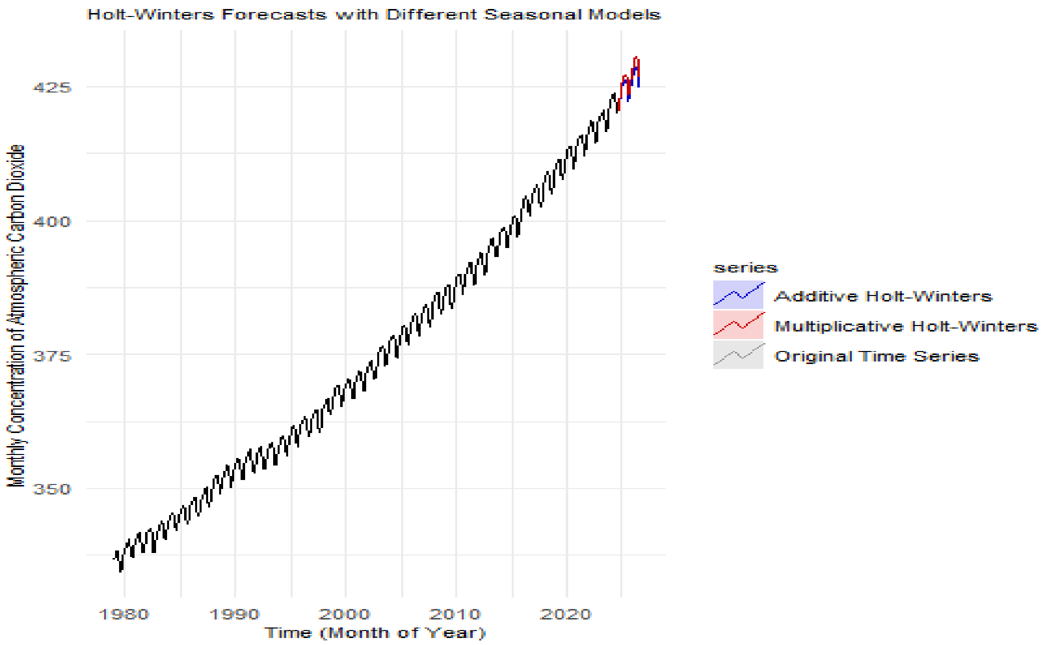

Figure 5: Forecast of CO2 Levels Using Holt-Winters Method.

Both models follow the general trend of the original series, but there are differences in how they capture the seasonal patterns. As we mentioned earlier, the Additive model assumes a constant seasonal variation over time and seems to fit better with the seasonal patterns of the original series but for the multiplicative model, the seasonality is proportional to the trend of the series. This might not be suitable here, as the seasonal variations in the original series appear to remain relatively constant over time.

Model	**ME**	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
Additive	0.0182	0.1666	0.1289	0.0048	0.0349	0.0680	0.4908
Multiplicative	0.0071	0.3013	0.2288	0.0018	0.0621	0.1208	0.6562

Table 3: Accuracy Metrics for Additive and Multiplicative Holt-Winters Models for CO_2 Concentration

We can see that the additive model was generally better than the multiplicative model according to the metric values, as it had lower RMSE, MAE, MAPE, and ACF1 (Autocorrelation of Residuals at Lag 1). So the additive model provides a more accurate fit to the data. Now we can check the residual behavior.

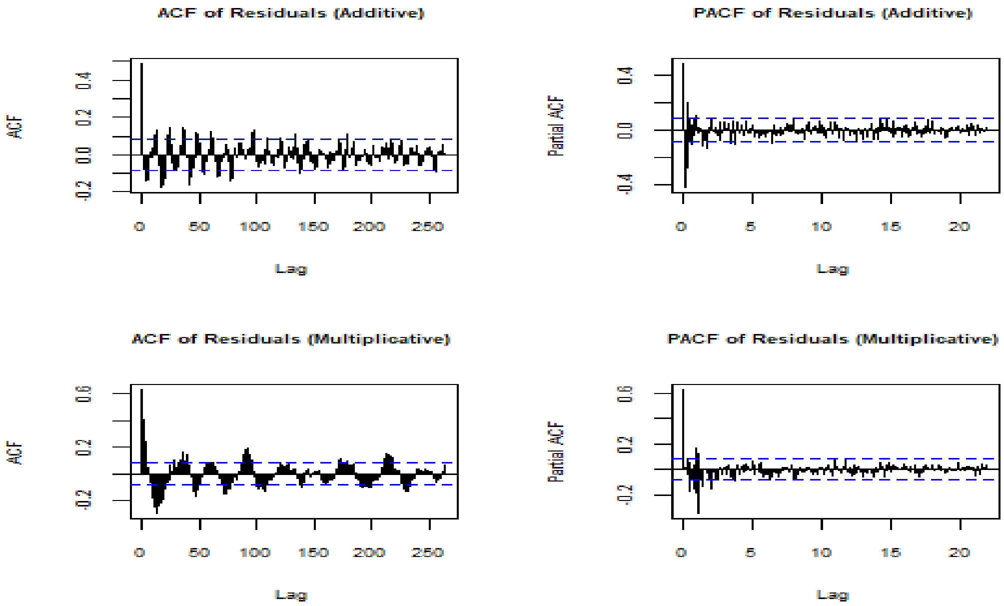

Figure 6: ACF and PACF Plots of Residuals for Additive (top row) and Multiplicative (bottom row) Holt-Winters Models

The plot shows that the Holt-Winters method is better than LSTM in terms of ACF and PACF plots, with Holt-Winters residuals exhibiting fewer significant autocorrelations and closer adherence to white noise assumptions. Between the additive and multiplicative models, the additive model performs better, as it aligns with the constant seasonality in the data, while the multiplicative model struggles due to its assumption of proportional seasonal variation.

1.1.4 SARIMA Model

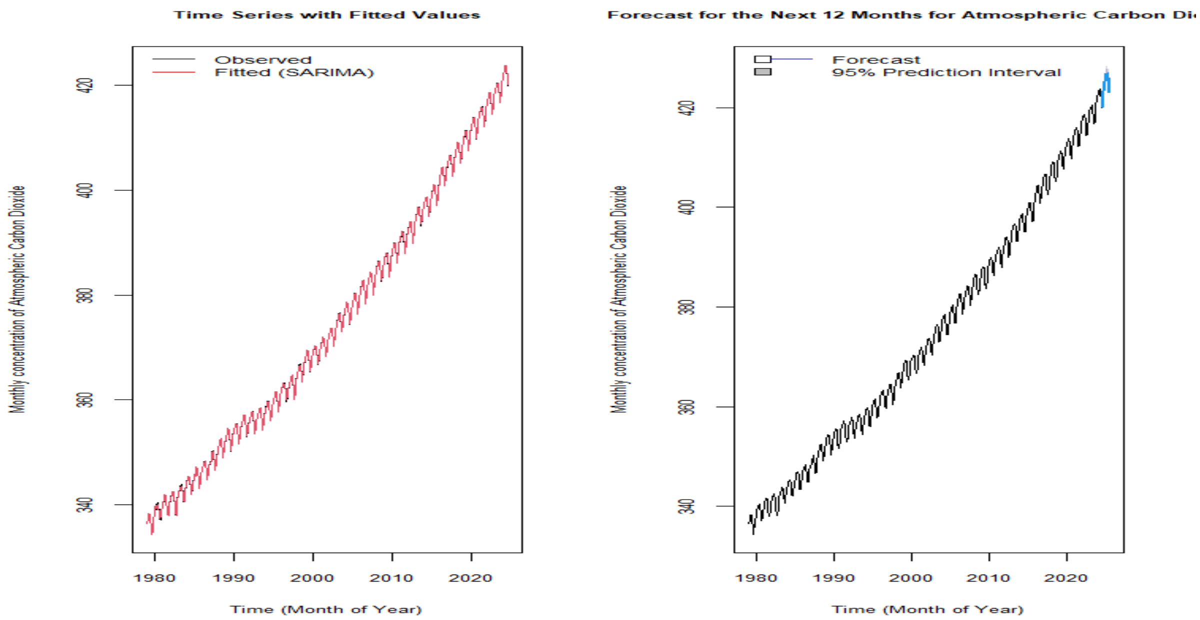

Figure 7: Historical Data with Fitted Values (left) and Future Forecast Based on the Sarima Model (right).

It is obvious that the SARIMA model performed well, providing a good fit to the observed data, as seen in the left plot, and delivering a reliable forecast for the next 12 months with a reasonable 95% prediction interval, as shown in the right plot. This demonstrates the model's ability to capture both the trend and seasonal patterns in atmospheric CO_2 concentrations effectively.

Metric	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training Set	-0.0004	0.1594	0.1180	-0.0002	0.0320	0.0623	0.0140

Table 4: Accuracy Metrics for the SARIMA Model

Low values for RMSE, MAE, and MAPE indicate accurate predictions confirming a good fit to the data.
Moreover, the plot below shows that the ACF plot (left) indicates that most residual autocorrelations fall within the confidence bounds, suggesting minimal autocorrelation and a good fit for the SARIMA model. Similarly, the PACF plot (right) shows no significant spikes beyond the confidence limits, confirming that higher-order dependencies have been effectively captured. Together, these plots indicate that the SARIMA model's residuals resemble white noise, presenting a well-fitted model.

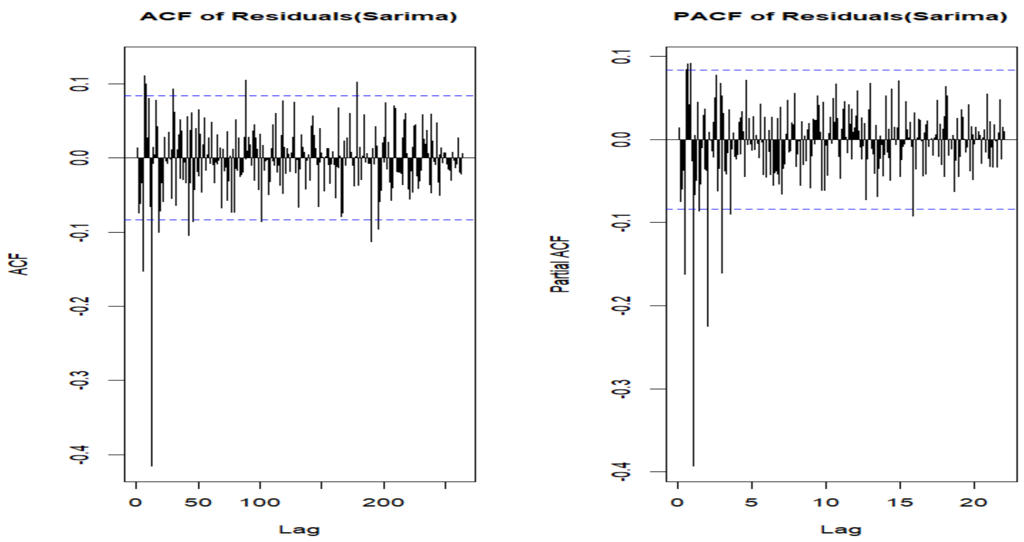

Figure 8: ACF (Left) and PACF Plots of Residuals for SARIMA Model

1.1.5 ARIMA Model

In this section, we present the Auto-ARIMA model as our last model.

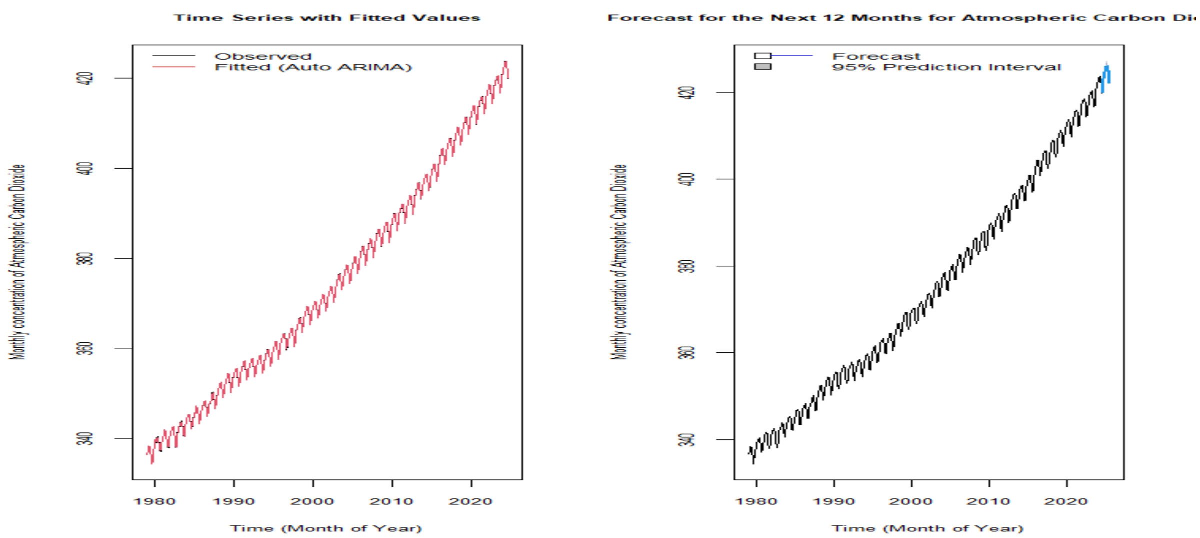

Figure 9: Historical Data with Fitted Values (left) and Future Forecast Based on the Arima Model (right).

The plot shows the **Auto ARIMA model** model could fit the observed data very well. The **forecast for**

the next 12 months follows the upward trend reflecting increasing uncertainty. The model effectively captures the trend and provides a reliable forecast for atmospheric CO_2 concentrations.

Metric	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training Set	0.0043	0.1235	0.0906	0.00099	0.0246	0.0478	0.0295

Table 5: Accuracy Metrics for the Auto ARIMA Model

The Auto ARIMA model shows strong performance, with low errors across multiple metrics and minimal autocorrelation in residuals so the forecasts are reliable.

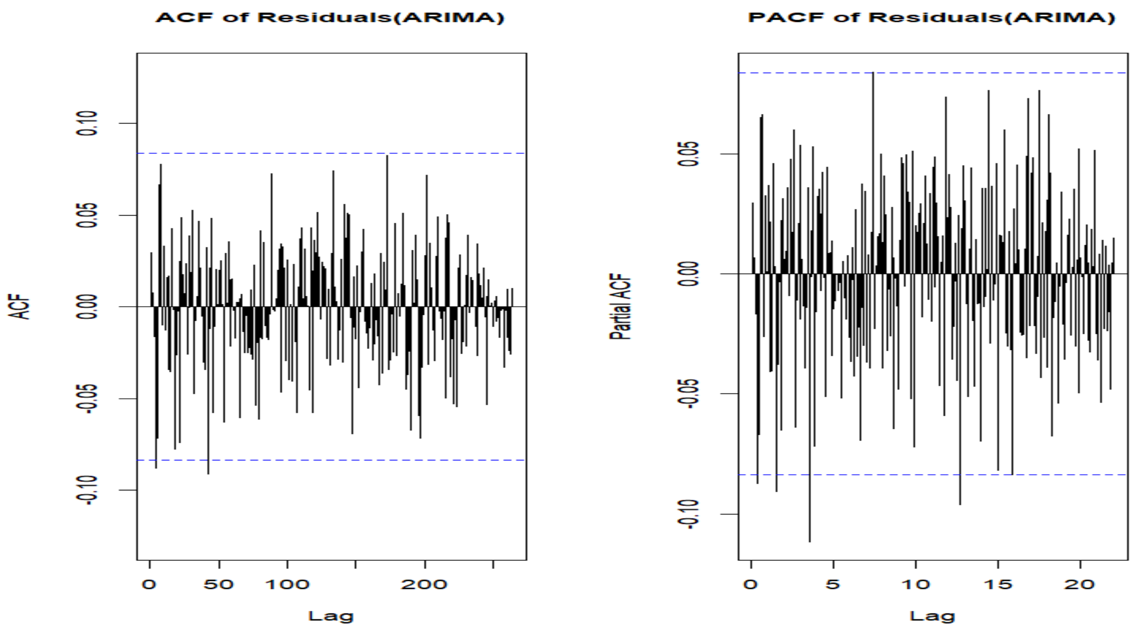

Figure 10: ACF (Left) and PACF Plots of Residuals for ARIMA Model

The ACF plot (left) shows most residual autocorrelations fall within the confidence bounds so, we can say that the predictions are reliable. Similarly, the PACF plot (right) shows no significant spikes beyond the confidence limits. In general, the ARIMA model's residuals present a well-fitted model.

1.1.6 Metrics and Conclusion

Metric	Auto ARIMA	SARIMA	HW Additive	HW Multiplicative	TSLM
RMSE	0.1235	0.1584	0.1666	0.3013	2.5247
MAE	0.0906	0.1181	0.1289	0.2288	2.2707
MAPE	0.0246	0.0320	0.0349	0.0621	0.6064
R-squared	-	-	-	-	0.9896

Table 6: Comparison of Metrics for Different Models

Based on the table, Auto ARIMA appears to be the best model overall. It outperforms the other models in terms of RMSE (0.1255), MAE (0.0906), and MAPE (0.0246), indicating it has the smallest errors and provides the most accurate predictions. SARIMA follows closely, with slightly higher errors. The Holt-Winters models perform reasonably well but still show higher error values compared to Auto ARIMA. Lastly, TSLM shows the highest values indicating it performs the worst among all of the models. In general, the Auto ARIMA model is the most accurate forecasting tool.

2 Rising Sea Levels and Oceanic Changes

In this section, we will examine one of the side effects of global warming: rising sea levels. A warmer planet leads to rising sea levels, primarily due to the melting of polar ice caps and glaciers, as well as the thermal expansion of seawater. Rising sea levels can cause flooding in coastal areas, loss of habitat for wildlife, and displacement of human populations. Additionally, changes in oceanic heat affect marine ecosystems, disrupt weather patterns, and contribute to the increased intensity of hurricanes and typhoons. These changes have profound implications for both the environment and human societies.

2.1 Time Series Analysis of Global Sea Level Rise

As the Earth's surface temperature becomes hotter, the sea level rises. This is partly because water above 4 °C (39 °F) expands when it gets warmer. It is also partly because warm temperatures make glaciers and ice caps melt. The rising sea level causes coastal areas to flood. Weather patterns, including where and how much rain or snow falls, are changing. Deserts will probably become larger. Colder areas will warm up faster than warm areas. Strong storms may become more likely, and farming may not produce as much food.

2.1.1 Explanatory Data Analysis

Figure 11: The seasonal plot for Quarterly Sea Level(1881 - 2009). For an interactive version of this plot, **click here**.

From the seasonal plot above, we can observe a quarterly seasonal pattern for most of the years. This seasonality remains consistent across years, decades, and even centuries. However, there are notable outliers, such as 1882 and 2009. 1881 is among the years with notably low sea levels, potentially due to data anomalies, historical measurement limitations, or unique environmental events that significantly impacted sea levels during that period. 2009, on the other hand, shows unusually high sea levels compared to the long-term trend, possibly due to recent global warming effects, accelerated melting of polar ice, or other climate-related phenomena. Beyond 2009, the dataset contains missing values; if these were available, the analysis could yield more conclusive results. Generally, the sea level tends to be lower during the colder seasons (Q1, Q2, and Q4) and higher during the warmer months (Q3, summer). Additionally, there is a clear positive trend, with sea levels rising significantly from 1881 to 2009.

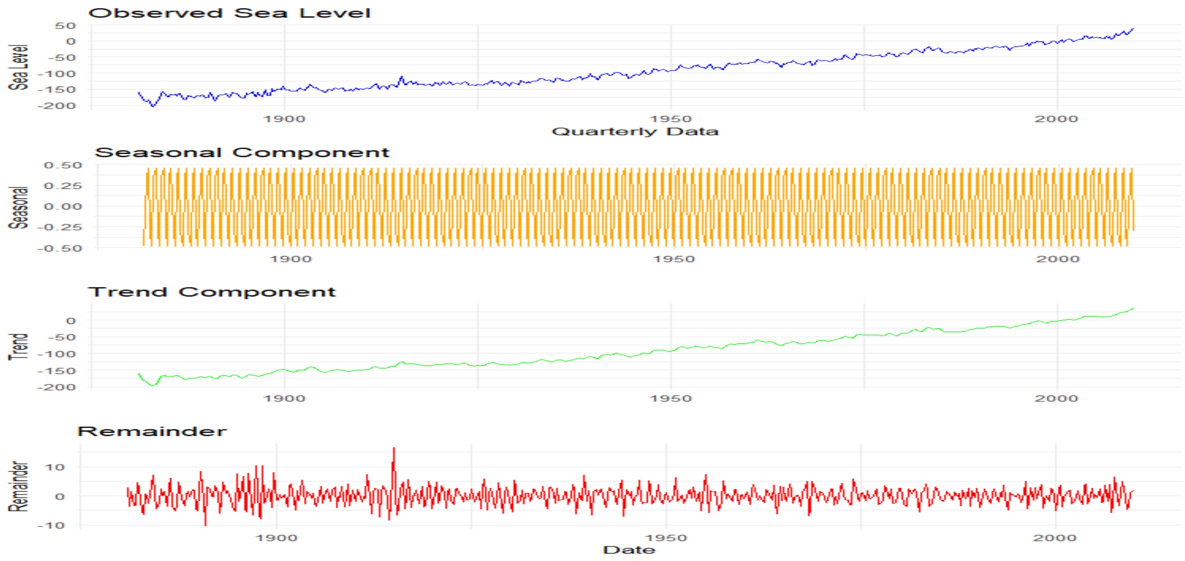

Figure 12: Decomposition of Observed Data: Trend, Seasonality, and Remainder. For an interactive version of this plot, **click here**.

The decomposed sea level plot confirms previous findings. It shows a clear upward trend, indicating a steady sea level rise since the mid-1900s, likely driven by global warming. The seasonal component shows consistent periodic fluctuations with stable seasonal patterns. The observed data integrates these elements, illustrating the overall rise with seasonal variations. The remainder captures random noise without a clear pattern.

2.1.2 TSLM model

For this model, since we have considered frequency equal to 4 (quarterly data) we considered both trend and seasonality.

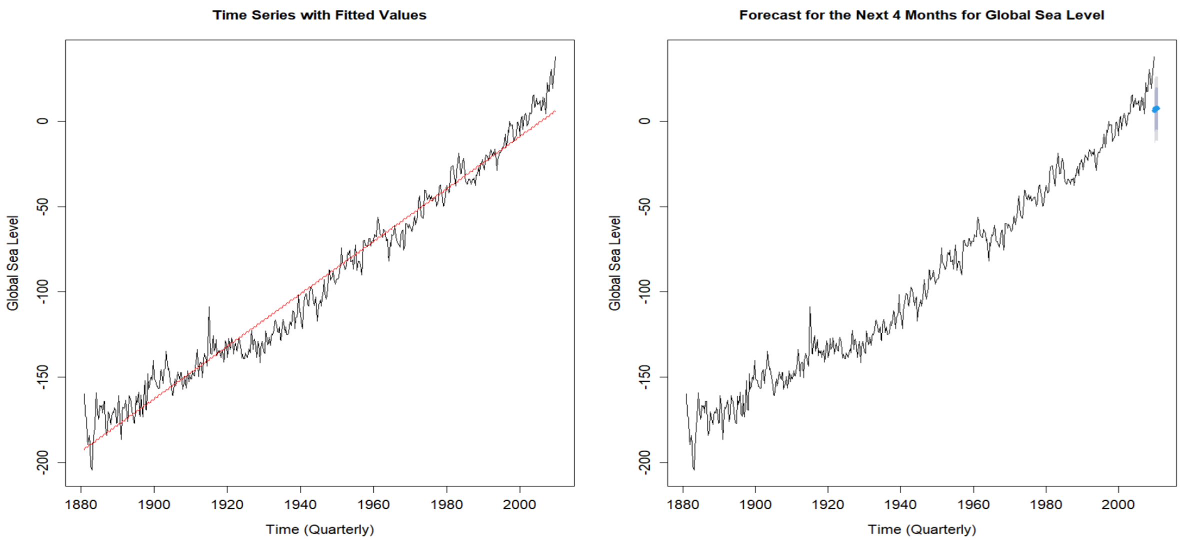

Figure 13: Historical Data with Fitted Values (left) and Future Forecast Based on the TSLM Model (right).

According to this plot, the TSLM model was not able to capture the correct information and the prediction is not accurate. Moreover, the forecast for the 4 months is not very good. The fluctuations are not captured well enough, We have to use another model that we will discuss later.

Set	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	R-squared
Training set	-4.69×10^{-16}	9.4724	7.4330	16.1919	55.6170	1.2036	0.8112	0.9734

Table 7: Accuracy Metrics of the TSLM Model

The performance is fine. We have a high R-squared value and RMSE, and MAPE is low but we need to check the residuals as well.

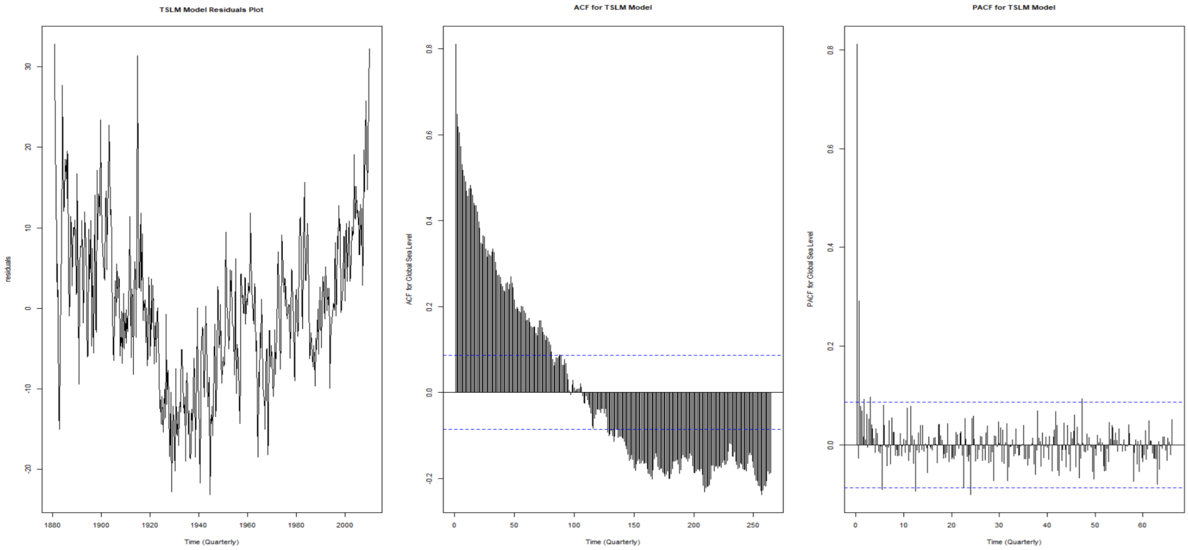

Figure 14: Residual Analysis for the TSLM Model: The Residual Plot (left), ACF Plot (center), and PACF Plot (right).

Test Name	**Model Tested**	**DW Statistic**	**P-value**	**Alternative Hypothesis**
Durbin-Watson Test	TSLM_model	0.332	$< 2.2 \times 10^{-16}$	True autocorrelation is greater than 0

Table 8: Durbin-Watson Test Results for TSLM Model Residuals for Global Sea Level

The residuals show a curved pattern over time, which indicates that the TSLM model isn't fully capturing some important aspects of the data, such as seasonality. Additionally, the ACF plot shows significant autocorrelations at several lags, with values exceeding the blue confidence bands. This means the residuals are not behaving like white noise, suggesting there's still information in the data that the model hasn't accounted for. On the other hand, the PACF plot doesn't show any significant values outside the confidence bands, which implies that higher-order dependencies might not be as important. Besides, the Durbin-Watson statistic indicates strong positive autocorrelation. The p-value is smaller than 0.05 which confirms that residuals are significantly correlated. So, we will apply other models that can handle autocorrelation.

2.1.3 Exponential Smoothing: Holt-Winters Method

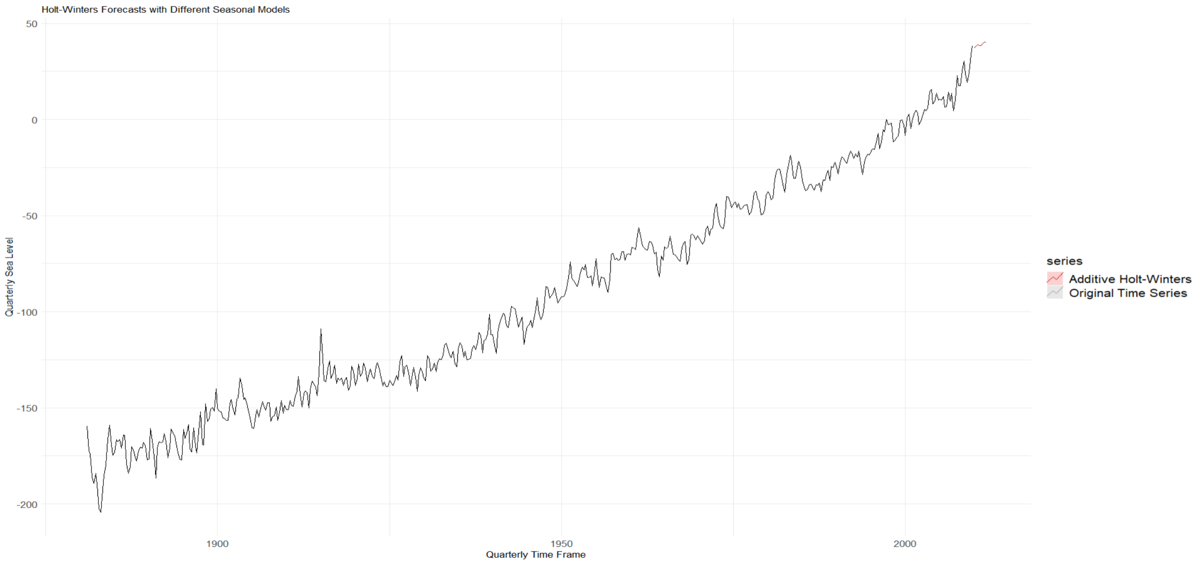

Figure 15: Forecast of CO2 Levels Using Holt-Winters Method.

The original sea level time series and the forecasted values for the next quarter from the additive Holt-Winters model are shown. The model appears to predict logically; however, it is advisable to check the model's accuracy as well. We have to mention that we could consider a multiplicative model as well but since we have negative values it is not possible to work with it.

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Additive	0.0594	5.4446	4.2205	8.2623	23.3949	0.6834	0.0586

Table 9: Accuracy Metrics for Additive Holt-Winters Model for Global Sea Level

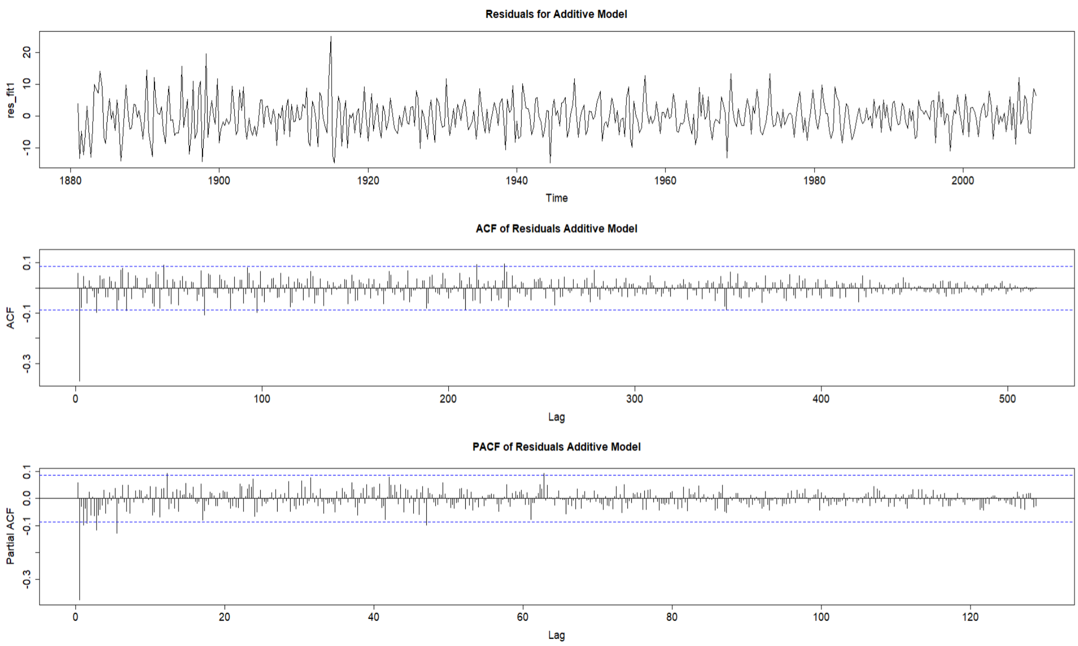

Figure 16: Residual, ACF and PACF Plots for Additive Model for Global Sea Level

The plot shows that the Holt-Winters method is better than LSTM in terms of ACF and PACF plots, with Holt-Winters residuals exhibiting fewer significant autocorrelations and closer adherence to white noise assumptions. Moreover, the plot of residuals is constant and does not have any specific pattern.

2.1.4 SARIMA Model

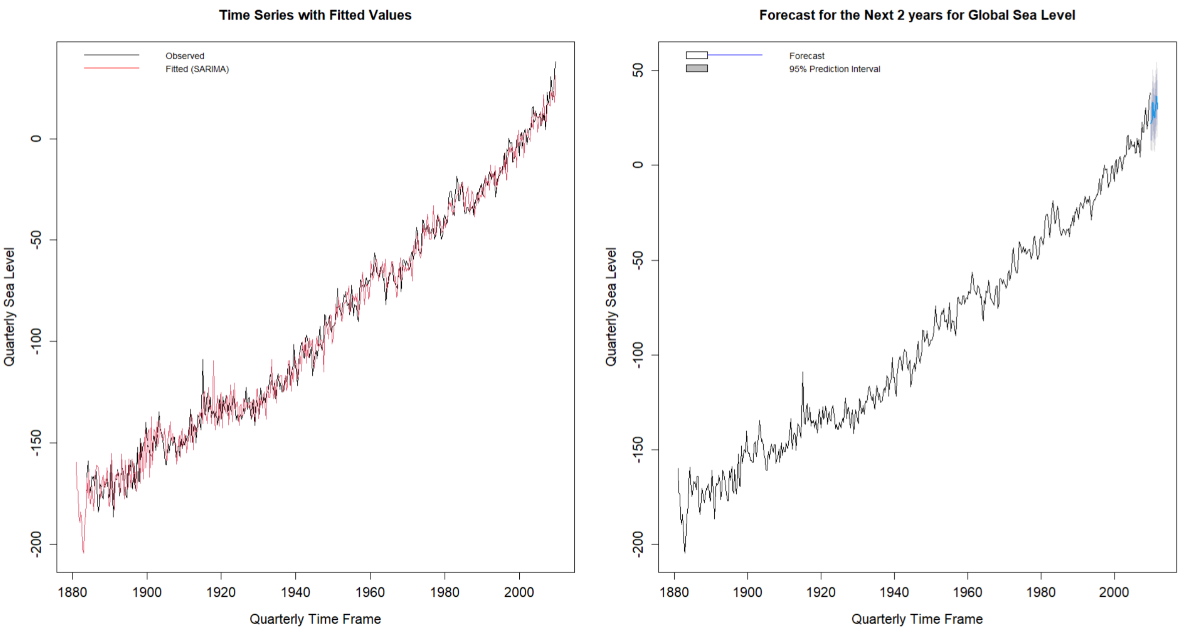

Figure 17: Historical Data with Fitted Values (left) and Future Forecast Based on the SARIMA Model (right).

According to this plot, the SARIMA model was able to capture the correct information and fitted well to the data. Moreover, the forecast looks pretty reasonable.

Set	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.2284	6.9203	5.2635	10.9366	32.3358	0.8523	0.1094

Table 10: Accuracy Metrics for SARIMA Model

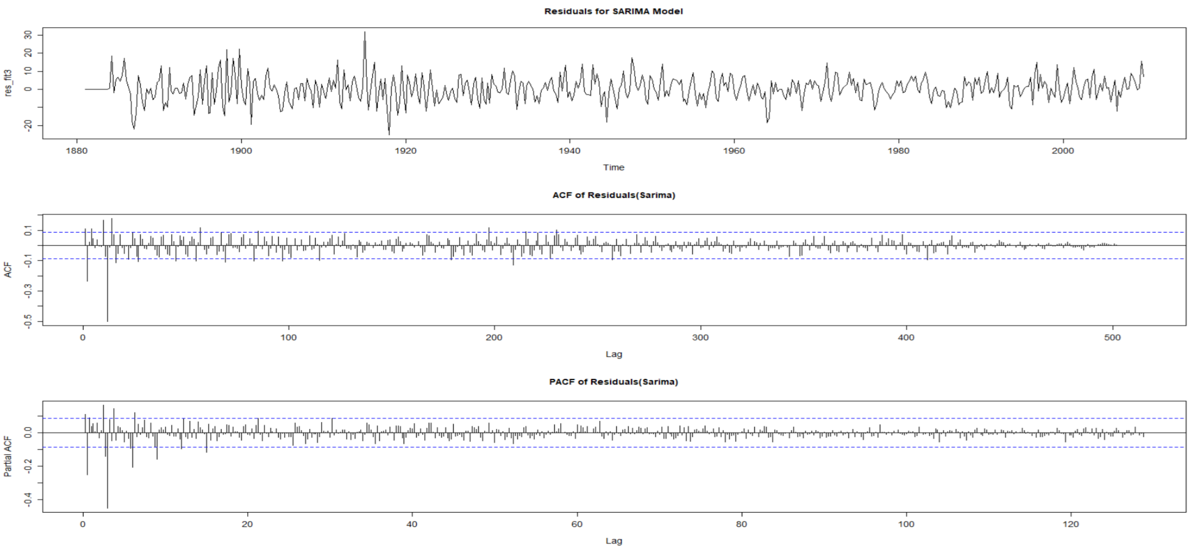

Figure 18: Residual, ACF and PACF Plots for SARIMA Model for Global Sea Level

The residual plot at the top shows the residuals of the SARIMA model, which fluctuate around zero, indicating no obvious pattern and suggesting the model has captured the data trend reasonably well. The ACF plot demonstrates that most autocorrelations are within the confidence limits, suggesting no significant remaining autocorrelation in the residuals. The PACF plot at the bottom also shows that partial autocorrelations diminish rapidly, confirming that the residuals behave like white noise. Overall, the SARIMA model provides a good fit for the data.

2.1.5 ARIMA Model

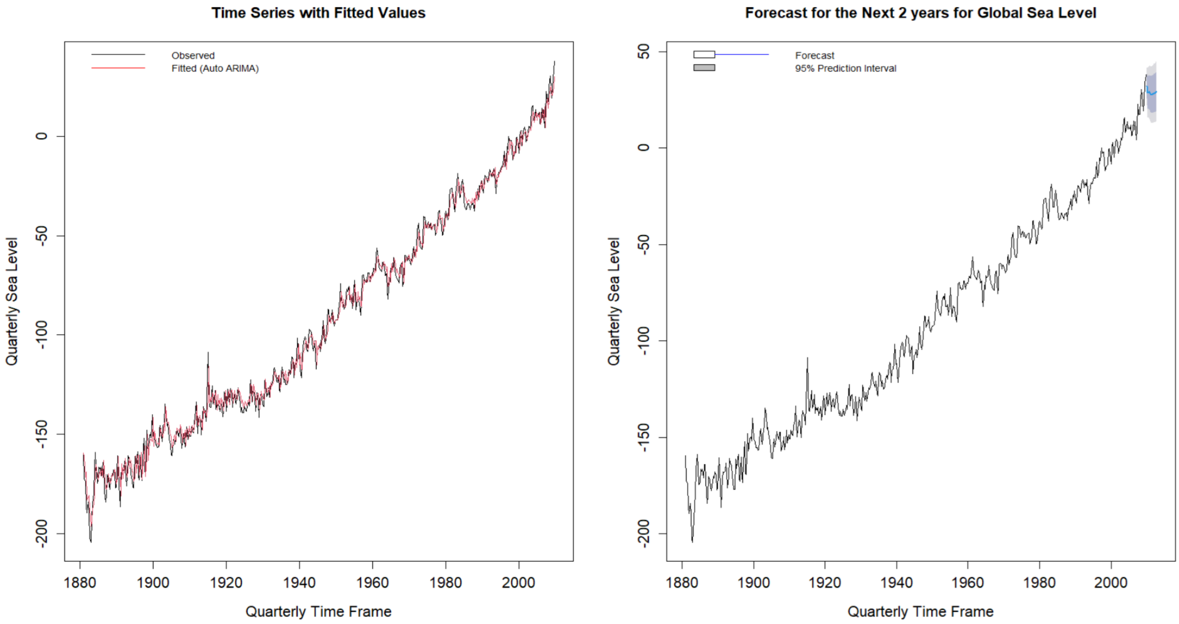

Figure 19: Historical Data with Fitted Values (left) and Future Forecast Based on the ARIMA Model (right).

Just like the SARIMA model, the ARIMA model shows a good fit and effectively captures trends and fluctuations. The forecast for the next two years displays a clear continuation of the upward trend along with 95% prediction intervals. The ARIMA model appears effective for predicting sea level changes.

Set	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	−0.0720	4.9007	3.7593	13.0917	28.9657	0.6087	0.0095

Table 11: Accuracy Metrics for ARIMA Model

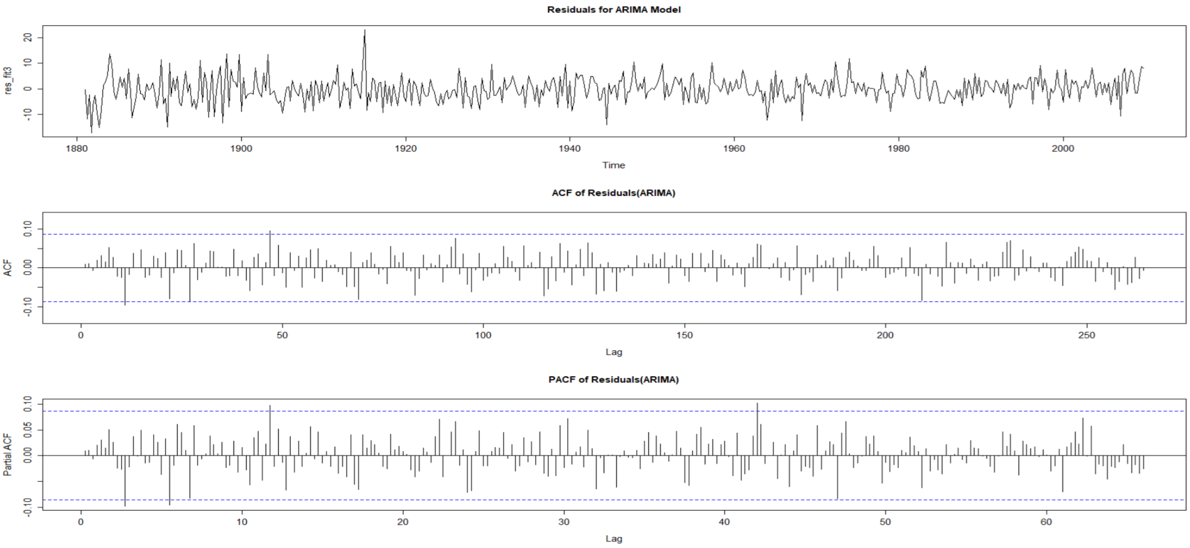

Figure 20: Residual, ACF and PACF Plots for ARIMA Model for Global Sea Level

The residual analysis of the ARIMA model shows that the residuals are random and centered around zero, indicating a good fit for the observed data. The ACF plot shows that most residuals fall within the confidence intervals, suggesting no significant autocorrelation. The PACF plot also supports this observation, with no clear patterns in lag dependencies. So we say that the ARIMA model adequately captured the structure in the sea level time series.

2.1.6 Metrics and Conclusion

Metric	Auto ARIMA	SARIMA	HW Additive	TSLM
RMSE	4.9007	6.9203	5.4446	9.4724
MAE	3.7593	5.2635	4.2205	7.4330
MAPE	28.9657	32.3358	23.3949	55.6170
R-squared	-	-	-	0.9734

Table 12: Comparison of Metrics for Different Models

The ARIMA model appears to be the best choice, as it has the lowest values for RMSE, MAE, and MAPE. These metrics suggest that ARIMA provides the most accurate forecasts with the least error. While TSLM has a high R-squared value, it has significantly higher RMSE, MAE, and MAPE values which is a poorer performance compared to the other models.

3 Main Contributors

In this section, we analyze the main contributors to global warming. Although we had access to a complete dataset for all countries, our focus will be on the primary contributors. For this analysis, we used the ARIMA model, as it provided the best results among the methods we tested. Since the data is annual, we could not use the SARIMA model, which is suited for seasonal data. We also experimented with other models, such as TSLM and Holt's method, but ARIMA performed better and that's why we chose this approach.

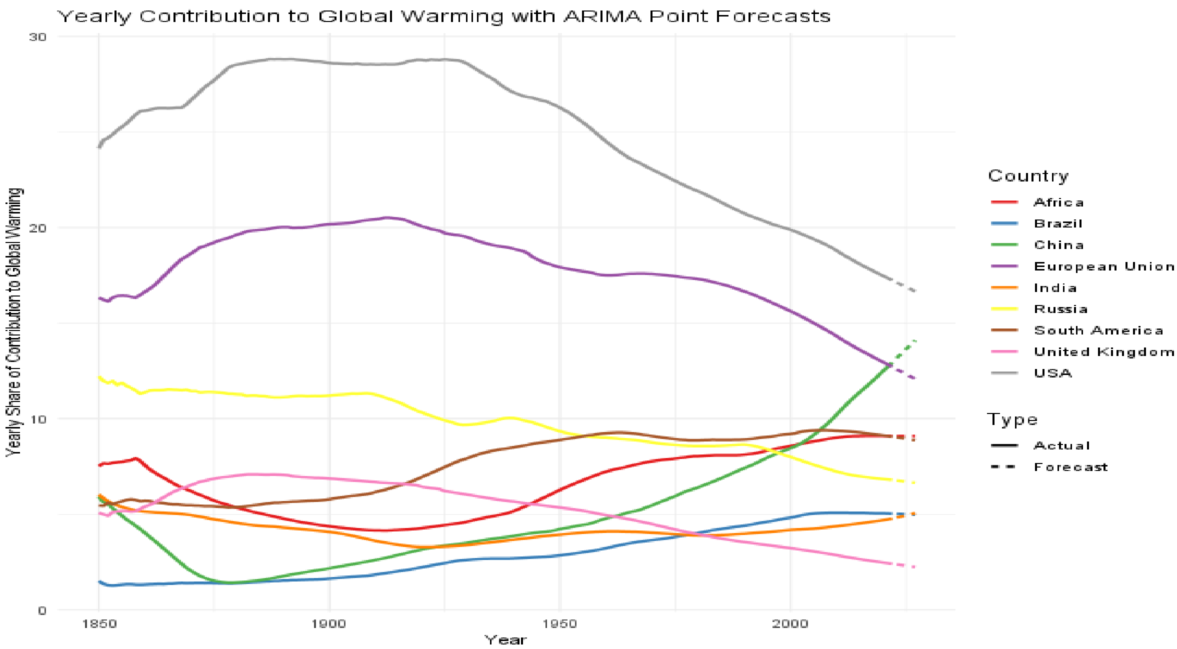

Figure 21: Yearly Contribution to Global Warming Forecasts

Country/Region	X-squared	p-value
USA	3.8410	0.2791
China	6.7888	0.07894
EU	0.3050	0.9591
Russia	0.7008	0.4025
Brazil	6.5040	0.0387
India	2.5289	0.4701
South America	0.1530	0.6957
Africa	1.9305	0.5870
UK	5.0849	0.1657

Table 13: Box-Pierce test results for different countries and regions

The plot displays the yearly contributions to global warming by various countries, with lines representing actual data and dashed lines showing ARIMA model forecasts. The Box-Pierce test results indicate that the ARIMA model fits well for most countries, as their p-values are above 0.05, suggesting no significant autocorrelation in the residuals. However, for Brazil (p-value = 0.0387), the residuals show some autocorrelations.

4 Conclusion

4.1 What Time Series Tell Us About Global Warming

The results from our analysis make one thing clear: global warming is a growing problem that isn't slowing down. Carbon dioxide levels are rising, and sea levels are following the same trend. These two factors alone highlight the severity of the issue and its potential to disrupt ecosystems, weather patterns, and human lives.

Looking at the contributions of different countries, we can see mixed efforts. Some nations have made progress in managing their emissions, but others remain major contributors, showing there's still a long way to go in the fight against global warming.

4.2 Recommendations

For this project, we could have used other models, such as the BASS model, but since our data is focused on global warming and its time-series nature, we decided to stick with models better suited for this type of analysis. Similarly, models like GANs could be explored, but they are not commonly applied to datasets like ours.

There are also other practical datasets that could provide valuable insights into the global warming problem, such as those on renewable energy usage, deforestation rates, air quality, ocean heat, methane gas concentration, and precipitation. However, due to restrictions on the length and scope of this report, we decided to focus on the datasets already discussed.