BIA - 600

Final Project Report

Spring - 22

Climate Change Gas Emissions Prediction

Team 5

Anh Le

Asad Shaik

Malarvizhi S Pandian

Manobala Purushothaman

Michael Dooley

**Table of Contents**

[Introduction: 3](#_Toc101212421)

[Problem Statement: 4](#_Toc101212422)

[Data Description: 5](#_Toc101212423)

[Time Series Forecasting 6](#_Toc101212424)

[ARIMA method: 6](#_Toc101212425)

[ETS method: 9](#_Toc101212426)

[Evaluation Metrics 10](#_Toc101212427)

[Evaluation graphs: 12](#_Toc101212428)

[Evaluation CO2 gas Prediction using ARIMA Method: 12](#_Toc101212429)

[Evaluation CO2 gas Prediction using ETS method: 13](#_Toc101212430)

[Evaluation Methane gas prediction using ARIMA: 13](#_Toc101212431)

[Evaluation Methane gas Prediction using ETS method: 14](#_Toc101212432)

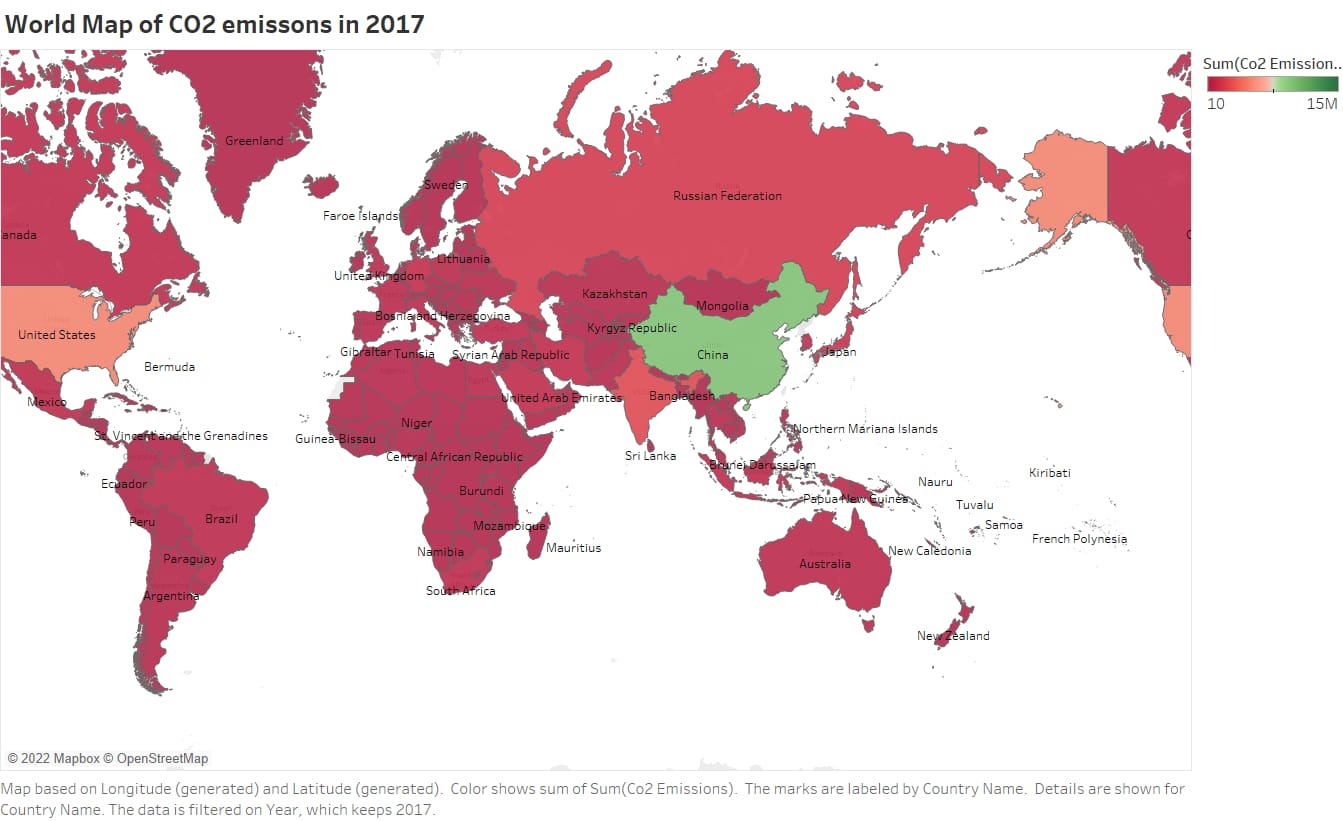
[Results: 15](#_Toc101212433)

[Recommendation: 15](#_Toc101212434)

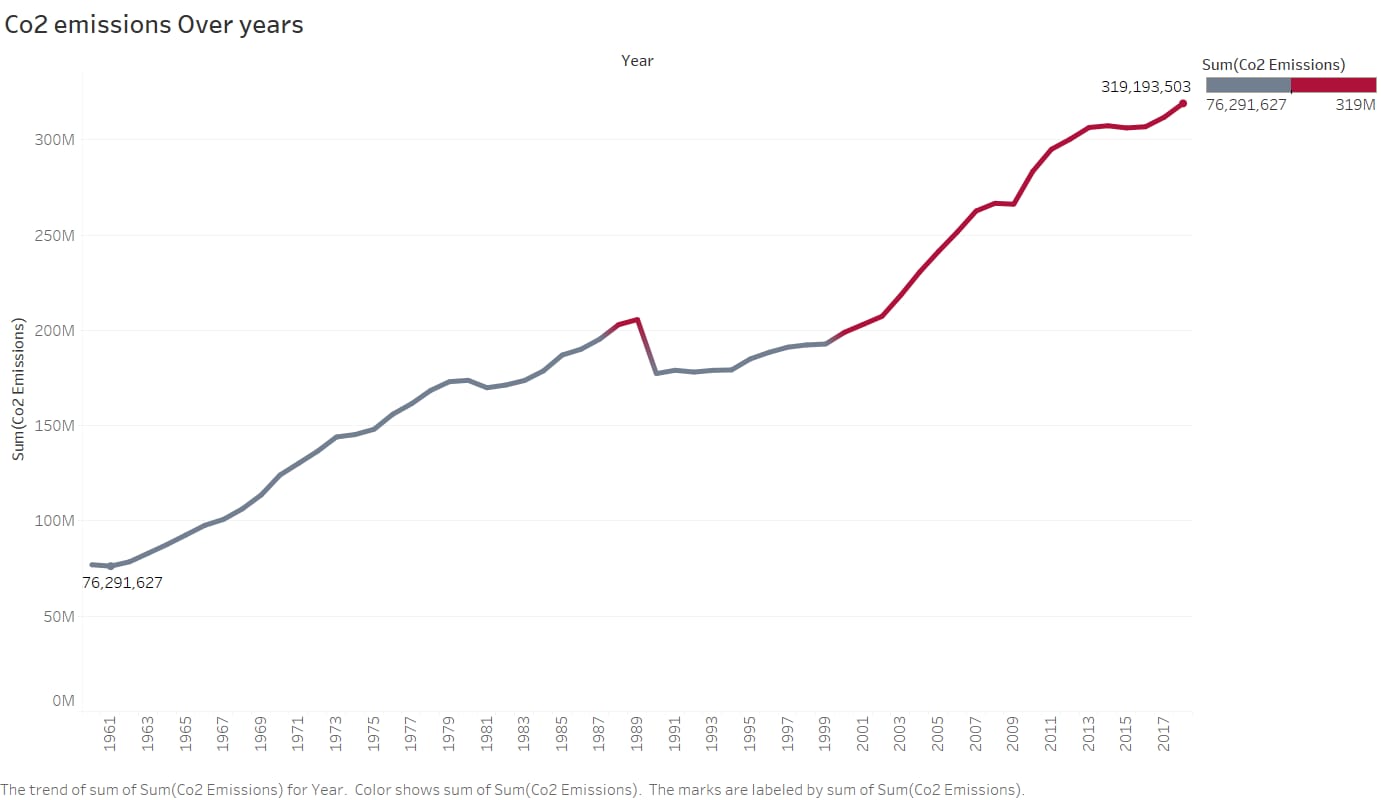
[References: 16](#_Toc101212435)

# **Introduction:**

Climate change is a long-term change in the average weather patterns that have come to define Earth’s local, regional, and global climates. These changes have a broad range of observed effects like global warming, heatwaves, drought, strong hurricanes, sea-level rise, and more. Most of it is caused due to emission of greenhouse gases which is a result of more fossil consumption by humankind. The impact of 1.5 degree Celsius of global warming is above pre-industrial level and related global greenhouse gas emission. What is 1.5 degree Celsius? It is likely 70 to 90% of coral reefs will die off worldwide. At 2 degrees Celsius of warming, 99% are lost. Human induced warming reached approximately 1-degree Celsius (between 0.8 degree and 1.2 degree) above pre-industrial level in 2017, increasing at 0.2-degree (between 0.1 degree and 0.3 degree) pre decade.

In this project, only two gases such as CO2 – carbon dioxide and CH4 – Methane is taken into consideration. Based on previous historical data on each of these gas emissions to the atmosphere, a time series forecasting is conducted to predict the emissions amount till 2035 from the present day.

We can see from the color fluctuation in the heat map of CO2 emissions around the world in 2017 that most CO2 emissions in 2017 came from China and the United States of America. Apart from those countries, we can observe that Russia and India have significant CO2 emissions that must be regarded in the context of climate change. The rest of the world is affected by climate change as well. Preventive actions should be implemented as soon as possible in nations with high levels of CO2 emissions to limit global warming to a manageable level.



The above line graph depicts the global average of CO2 emissions and the increase in the value.

# **Problem Statement:**

Few questions are expected to be addressed by doing this predictive analysis and those are stated below.

* Given the current rate of increase in greenhouse gas emissions, by what percentage would CO2 and Methane gas emissions increase by the year 2035?
* Does the predicted value go beyond the unacceptable level of CO2 (1000-2000ppm) and Methane (above 5000 ppm) gas amount in the atmosphere?

# **Data Description:**

There are two datasets that we retrieved from the global monitoring laboratory’s website.

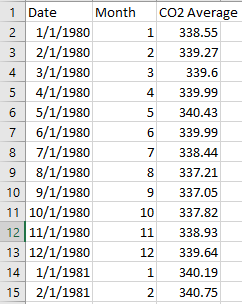
**Data Source:** https://gml.noaa.gov/ccgg/trends/data.html

The first one is Trends in CO2 that show the average carbon dioxide daily from 1/1/1980 until 1/1/2022. This consists of 3 columns: Date, Month, CO2 average in unit of part per million(ppm).

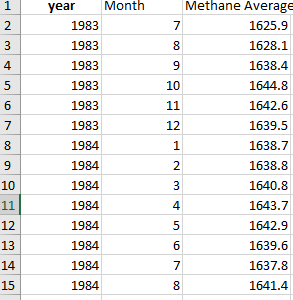
The second dataset is Trends in CH4 which show the average methane from 1983 until 2021. This dataset also consists of 3 columns: Date, Month, Methane Average in unit of part per million (ppm).

A snippet of the data that we use from the data sources:

Dataset 1 – Trends in CO2 from 1980 to 2022



Dataset 2 – Trends in CH4 from 1983 to 2021



# **Time Series Forecasting**

Here, in this project based on historical data of CO2 and Methane emissions in the atmosphere, a time series forecasting is conducted to predict the future emission values. This is quantitative forecasting of climate data.

Time Series Forecasting is the method used in machine learning to predict a particular data, where we are not aware of actual data for a future period. Prediction is performed based on previous values recorded by learning the pattern and trend in the data. There are different components in time series data and each one is explained below.

* Level: Any time series will have a base line. To this base line we add different components to form a complete time series. This base line is known as level.
* Trend: It defines whether, over a period, time series increases or decreases. That is, it has an upward (increasing) trend or downward (decreasing) trend. For e.g. The above time series has an increasing trend.
* Seasonality: It defines a pattern that repeats over a period. This pattern which repeats periodically is called seasonality. In the above graph, we can clearly see the seasonality component present.
* Cyclicity: Cyclicity is also a pattern in the time series data, but it repeats aperiodically, meaning it doesn’t repeat after fixed intervals.
* Noise: After we extract level, trend, seasonality/cyclicity, what is left is noise. Noise is a completely random fluctuation in the data.

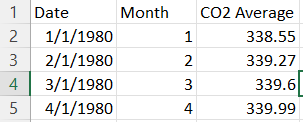
**Different time series methods:**

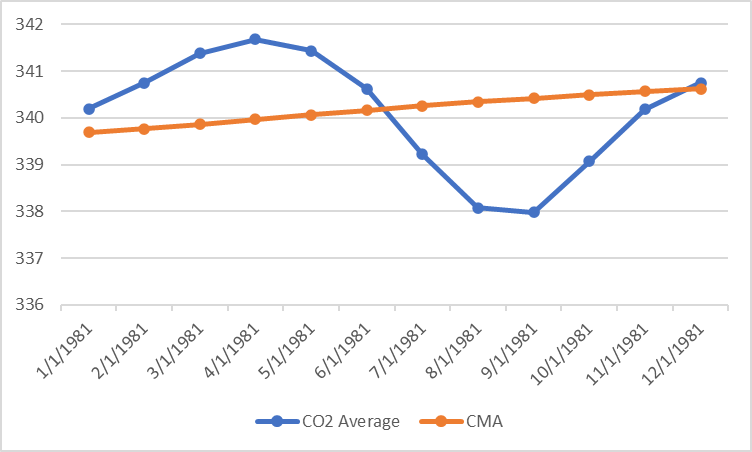
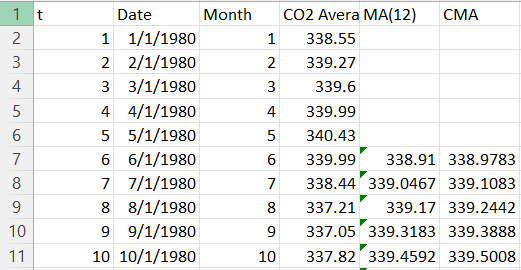
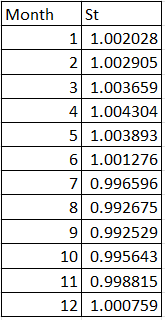
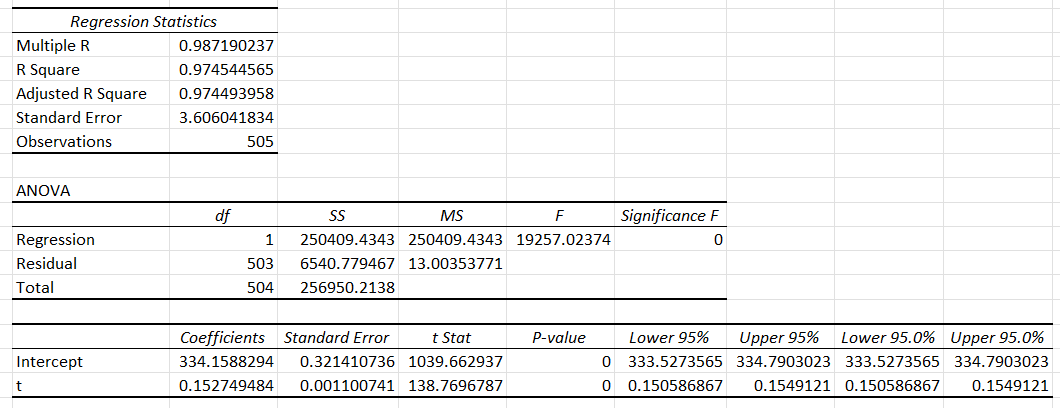
For this project, we used two different kinds of methods to predict the CO2 and methane value in the atmosphere. The methods used are ARIMA and ETS, and each one is explained in detail below,

## **ARIMA method:**

Auto Regressive Integrated Moving Average – ARIMA method is a stationary technique and does not use seasonality in prediction. This non-seasonal model is a combination of autoregression and a moving average model.

Execution steps in excel for ARIMA:



1. Added a time component ‘t’, a new column and filled with increasing number starting from one.
2. The moving average for seasonality of 12 is calculated and filled in a new column called ‘MA’. Seasonality is 12 as per the number of months in a year. This is calculated by taking an average of 12 CO2/ Methane values.
3. Since the moving average is done for an even seasonality value, another new column is added to calculate centered moving average. This is named ‘CMA’ and calculated by taking an average of two values in the ‘MA’ column.
4. The next step is aimed at calculating the seasonality value further, where we are calculating the seasonality along with the irregularity value is calculated. This is named ‘Y/CMA’ and as the name suggests the value is calculated by diving CO2/Methane value by the CMA value. How to interpret this value is
   1. If the value is greater than one i.e., for example 1.08, then it means the actual observed value at that position is 0.08 above the CMA value.
   2. If the value is less than one i.e., for example 0.94, then it means the actual observed value at that position is 0.06 below the CMA value.
5. Determine the seasonality value by calculating the average of ‘Y/CMA’ value for each of 12 seasons we have in our data. That value is updated according to the season (in our case it is the month) in a column named ‘St’.
6. Once the seasonality is defined then this step is performed to deseasonalize the data. This is calculated by dividing the actual CO2/ methane value by seasonality value and named as ‘Deseasonalize’.
7. The trend value is calculated using Simple Linear Regression method in Excel Data Analysis tool. For regression, the independent value Y and dependent value X are deseasonalized data and t value respectively. The co-efficient of CO2/methane emission ‘Intercept’ value(c0) and co-efficient of ‘t’ value(c1) is used in the formula – {c0 + c1 \* t} to obtain trend value. The column is named ‘Trend’

Application, table

Description automatically generated

1. Finally, forecast values are calculated by multiplying the trend value and seasonality (St) value for each time period. This is added in the column named ‘Forecast’.

## 

## **ETS method:**

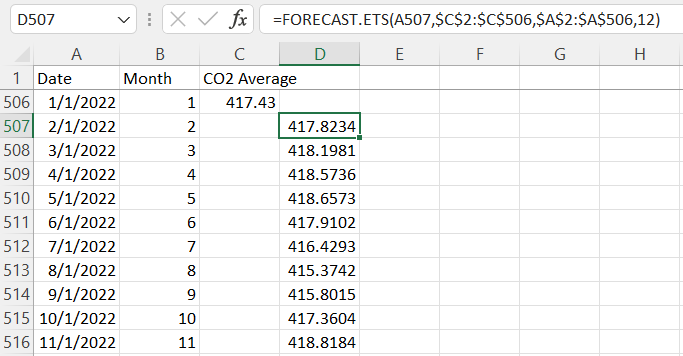
Exponential Triple Smoothing – ETS method is a non-stationary technique which considers the error, trend and seasonality components of data. The past observations are weighted equally, in the simple exponential smoothing technique, where the recent observations are given more weightage than the past observations.

**Execution steps for ETS method:**

For ETS forecasting method, in-built excel formula is used on the original data to predict the values.

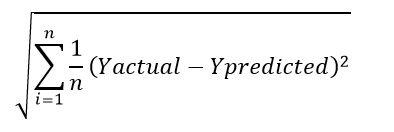
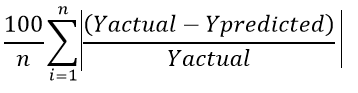
*FORECAST.ETS(target\_date, values, timeline, [seasonality], [data\_completion], [aggregation])*

In the above formula is explained as follows:

1. Target\_date – is the date we are trying to predict the value for.
2. Values – the range of CO2/methane emission values in the data
3. Timeline – date column in the data
4. Seasonality – 12, as we have monthly average data for every year
5. Data\_completion and aggregation – these does not apply for our problem and left blank

# **Evaluation Metrics**

There are many methods to evaluate the predicted values compared to the actual values. In this project we are going to use only two metrics and those are – RMSE and MAPE. Both these metrics should be lower to conclude that our model predicted closer to actual values.

1. *Root Mean Squared Error* (RMSE) is the square root of Mean Squared Error (MSE). MSE is nothing but a representation of how forecasted values differ from actual or true ones. It is represented by the following formula:
2. *Mean Absolute Percentage Error* (MAPE) is the measure of how accurate a forecast system is. It measures this accuracy as a percentage and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values. It is represented by the following formula:

In both formulas, Y actual is the true value and Y predicted is the predicted value at that particular time. n is the number of observations.

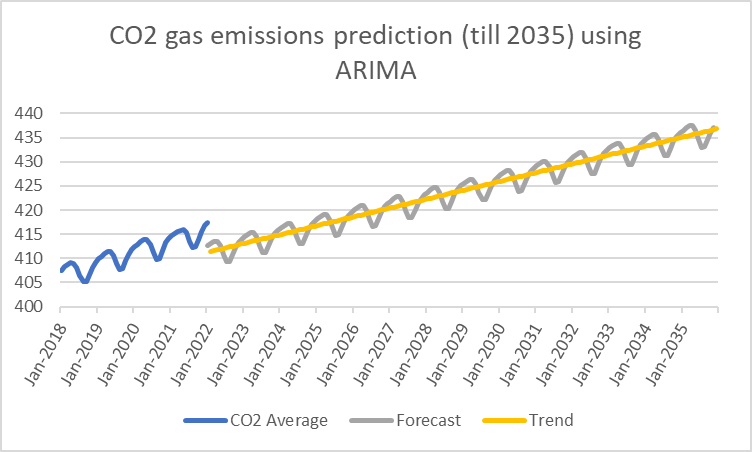
|  |  |  |
| --- | --- | --- |
| **Gas / Metric** | **RMSE** | **MAPE** |
| CO2 - ARIMA | 0.028184408 | 0.525000476 |
| CO2 - ETS | 0.21 | 4.60 |
| Methane - ARIMA | 0.075771161 | 0.538953342 |
| Methane - ETS | 2.47 | 1.926414077 |

Based on the metrics RMSE and MAPE, it seems like for our data the ARIMA method works well compared to the ETS method. The variance between the actual value and the predicted value for CO2 gas using RMSE and MAPE are 2.8% and 52.5% respectively using ARIMA. The variance between the actual value and the predicted value for methane gas using RMSE and MAPE are 7.5% and 53.8% respectively using ARIMA.

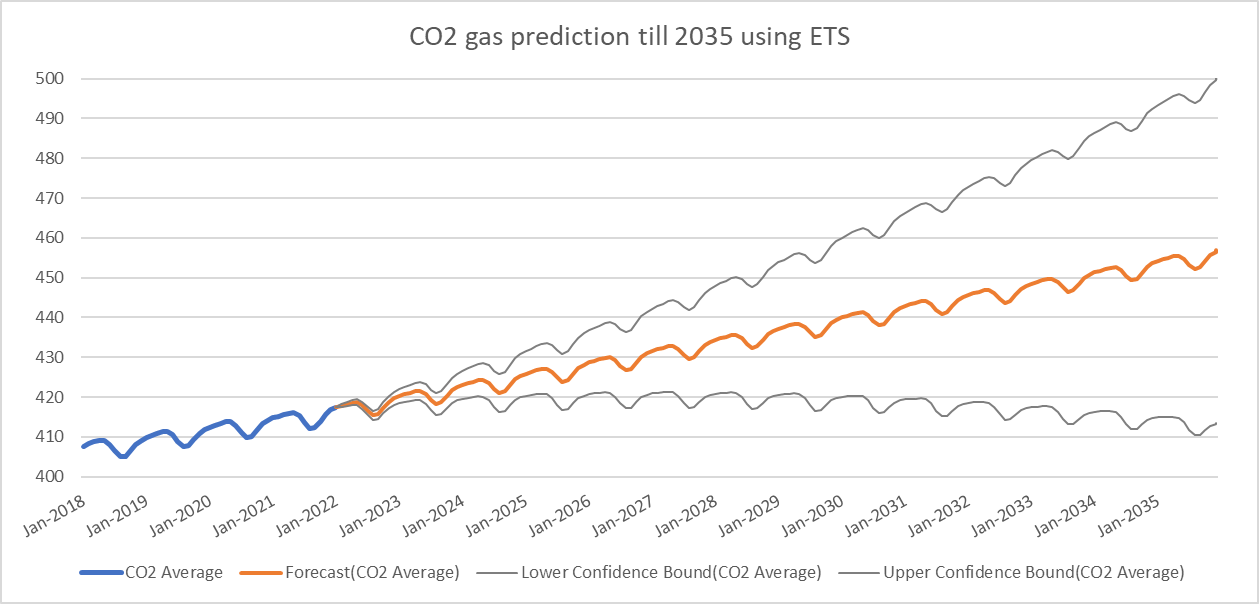
# **Evaluation graphs:**

## **Evaluation CO2 gas Prediction using ARIMA Method:**

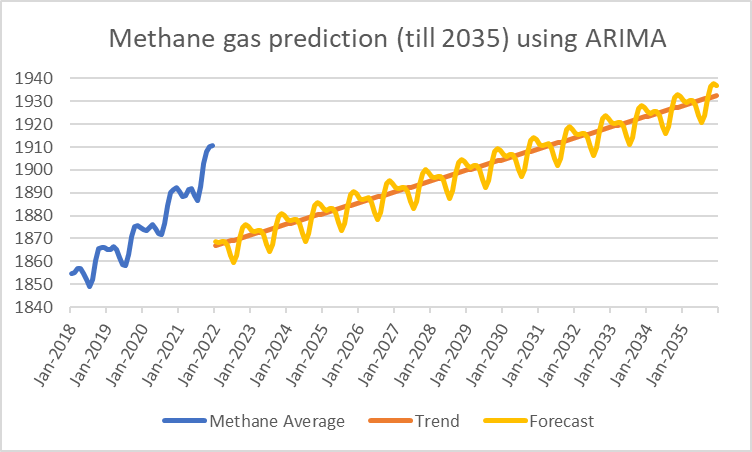
The graph shows globally-averaged, monthly mean atmospheric carbon dioxide. The first graph shows yearly means for the last four years plus the predicted year, and the grey curved line shows the NOAA time-series starting in 2022, when we have confidence in the data. values for the last year are preliminary pending recalibrations of standard gases and other quality control steps. Blue curved line indicate Average of CO2 emission, when the data is de-seasonalize with respective to CO2 average and regression, we get yellow straight line. The trend value is calculated using Simple Linear Regression Model, by using co-efficient of intercept and time value in regression and assumption time value. The trend line is used to smoothen the curve by using the de-seasonalize method. The curved grey graph is known as Forecast and is calculated until 2035, Forecast value calculated by multiplying seasonalize value and Trend Value.



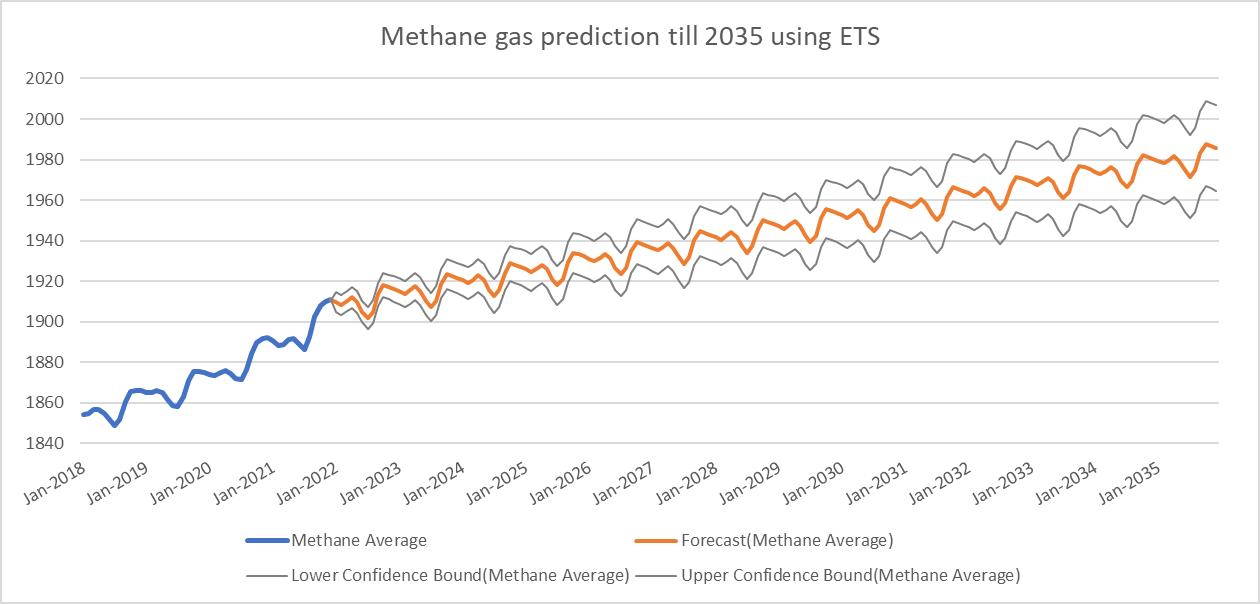
## **Evaluation CO2 gas Prediction using ETS method:**

The ETS value is calculated in Excel by using FORECAST-EST formula. In ETS method Forecast of CO2 lies in-between Lower Confidence Bound and Upper Confidence Bound. The forecast value increases and reaches an average value of 450.

## **Evaluation Methane gas prediction using ARIMA:**

NOAA’s preliminary analysis showed the annual increase in atmospheric methane during 2021 was 17 parts per billion (ppb), the largest annual increase recorded since systematic measurements began in 1983. Atmospheric methane levels averaged 1,895.7 ppb during 2021, or around 162% greater than pre-industrial levels. From NOAA’s observations, scientists estimate global methane emissions in 2021 will be 15% higher than the 1984-2006 period. In graph it shows us average methane. Reducing methane emissions is an important tool we can use right now to lessen the impacts of climate change in the near term, and rapidly reduce the rate of warming.

## **Evaluation Methane gas Prediction using ETS method:**

The ETS value is calculated in Excel by using FORECAST.EST formula. In this graph there is a slight difference between forecast value and upper and lower bound values when compared to CO2.

The graph depicts the average amount of methane emitted in different countries over time, as well as the estimated amount of methane gas that could be emitted until 2035, so that we can take the necessary steps to avoid global warming, which causes climate change and, in turn, has an impact on Earth's living species.

# **Results:**

A few insights based on all the metrics and model evaluation obtained are listed below separately for CO2 and Methane gas.

For CO2 gas:

1. By 2035, the CO2 gas emission is predicted to reach the value of 437 ppm given the rate of increase is the same as now.
2. According to the prediction, the CO2 presence in atmosphere increases by 4.7% I.e., from 417ppm to 437ppm
3. At this rate of increase, the CO2 level will reach the value of 500 ppm in next 38 years which is 2073

For Methane gas:

1. By 2035, the Methane gas emission is predicted to reach the value of 1936 ppm given the rate of increase is the same as now.
2. According to the prediction, the Methane presence in atmosphere increases by 1.3% I.e., from 1910ppm to 1936ppm
3. The predicted value is only a few percentages more because of sudden hike in values after 2018, but the model follows the overall average trend and moves at lower pace.
4. At this rate of increase, the Methane level will reach the value of 2000 ppm in next 33 years which is 2068

# **Recommendation:**

Globally, 2021 was the fifth warmest year on record, but it was only marginally warmer than 2015 and 2018.

Some of ideas that could be worked on in future and alternative ways to predict more accurate results are suggested.

* Try and perform time series forecasting using many other evolved methods. One example is to use SARIMA - Seasonal Autoregressive Integrated Moving Average method where seasonality is considered unlike ARIMA, a non-seasonal method. Taking seasonality into account will make the model predict values closer to the actual values
* Consider predicting more greenhouse gases similar to our project here. That will make us reach conclusions about global temperature predictions as well. Since greenhouse gases are the main reason for global warming.
* During the times of Covid, there was a notable reduction in pollution gases in the atmosphere due to reduced usage of vehicles. Other factors that increase the emission are industrial pollution, cement emissions, electricity waste and many more. Accounting all factors like this might even have less emission values in future.

# **References:**

1. <https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f>
2. <https://machinelearningmastery.com/time-series-forecasting/>
3. <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/>
4. <https://www.youtube.com/watch?v=5C012eMSeIU&t=0s&ab_channel=JalayerAcademy>
5. <https://datahub.io/core/co2-fossil-by-nation>
6. <https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide>
7. <https://www.sciencedirect.com/science/article/pii/S1876610217310056>
8. <https://www.nature.com/articles/s41598-021-86183-5#Tab8>