Global Temperature Change Prediction

12th December 2017

Ameya Shanbhag
avs431
Vikram Sunil Bajaj
vsb259

MOTIVATION

Global climate change has already had observable effects on the environment. Glaciers have shrunk, ice on rivers and lakes is breaking up earlier, plant and animal ranges have shifted and trees are flowering sooner.

Effects that scientists had predicted in the past that would result from global climate change are now occurring: loss of sea ice, accelerated sea level rise and longer, more intense heat waves and so on.

So far, global warming has been limited to a rise of around 0.75C since the end of the 19th century. This sounds like a small change, but scientific evidence suggests it is already leading to a range of impacts around the world.

More importantly, however, is that the temperature rise observed so far is not the chief cause of concern. More worrying is the significantly larger temperature rise (and associated changes in rainfall, snowfall, sea level and other phenomena) that scientists expect the world to experience in the coming decades and centuries.

Temperature plays an important role in shaping weather patterns, guiding the life cycle of various organisms and maintaining ocean levels. Shifting the temperature a couple of degrees can throw an entire ecosystem into chaos.

We, as Computer Science students, can contribute to climate change and global warming by predicting future temperature changes and making others aware of the consequences they will need to face. We have taken up this topic to dwell upon the rising concern of global warming and to gain insights from data that is available to us in the real world.

TARGET AND PREDICTOR VARIABLES

1. **AverageTemperature**
   Predicted temperature for a given city over a specified time period.
2. **AffectedCities**
   The top 10 cities in the United States that will experience the most temperature change from 2013-2023
3. **Effect of Pollution**
   Analyzing the correlation between pollution levels and temperature
4. **Effect of Greenhouse Gases**
   Predicting future Greenhouse Gas emissions and determining the Greenhouse Gas that has and will have the most correlation with temperature change

PROBLEM STATEMENT

To predict global temperatures by analyzing historical monthly city temperatures between 1750 and 2013.

DATA SOURCES

Temperature Data

Climate Change: Earth Surface Temperature Data from Kaggle

Pollution Data

U.S. Pollution Data from Kaggle

Greenhouse Gases Data

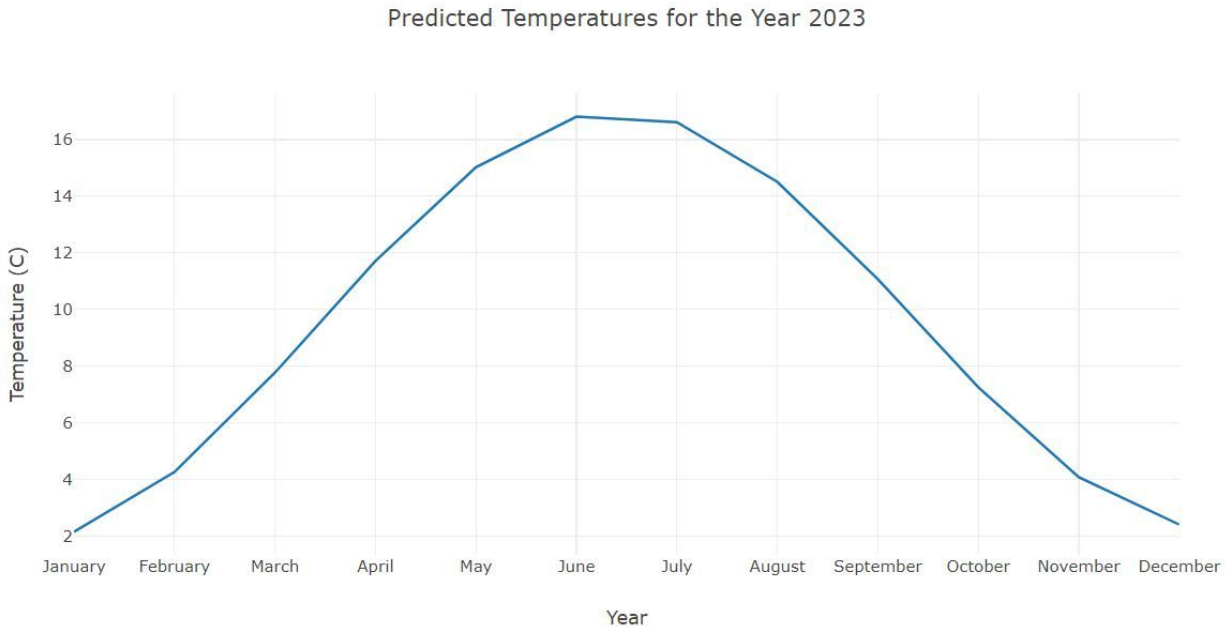
International Greenhouse Gas Emissions from Kaggle

MODEL FOR FORECASTING

Since we have time series data, we have chosen to use an **ARIMA (AutoRegressive Integrated Moving Averages) model** for time series forecasting.

We have automated the process of estimating the best p, q parameters for the ARIMA model using a grid-search technique, for any given time series data. The p,q values were chosen based on the pair that returned the lowest AIC (Akaike Information Criterion) values.

The model was used to successfully forecast the temperatures for a given city across a specified time period.

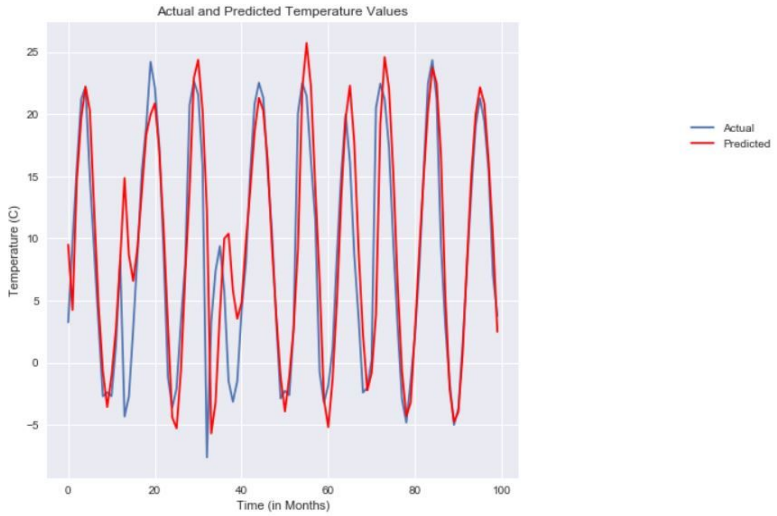


EVALUATION

We have used **Mean Squared Error (MSE)** and **Mean Absolute Error (MAE)** to evaluate the performance of the ARIMA model.

For New York, the ARIMA model achieved:

- Mean Squared Error: 4.67802989468
- Mean Absolute Error: 1.57799576152



ASSUMPTIONS / LIMITATIONS

The city-wise monthly temperature data had several missing values for certain cities. Instead of replacing these NaN values with the mean temperature, we chose to remove rows having missing values.

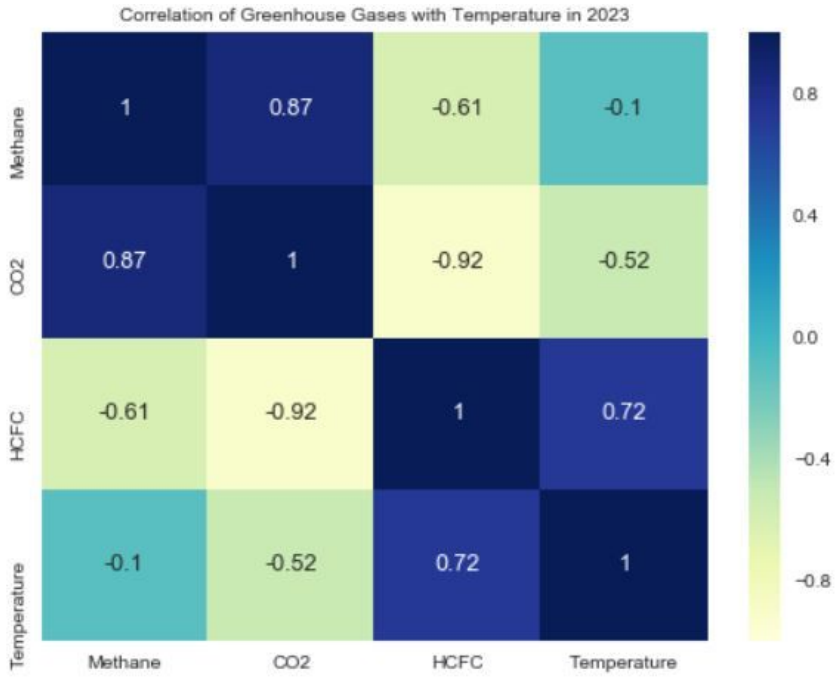
Also, we were not able to fully analyze the effect of Greenhouse gases on temperature change, due to the lack of data for gases like ozone (O3) and water vapor.

PROBLEM IN SCOPE OF CLASS

The crux of our problem lies in **Time Series Forecasting**. This is well within the scope of the class. We have made use of ACF and PACF plots, in combination with AIC values, to estimate the best p, q values for the ARIMA model and have forecasted the temperatures for cities over a given period of time.

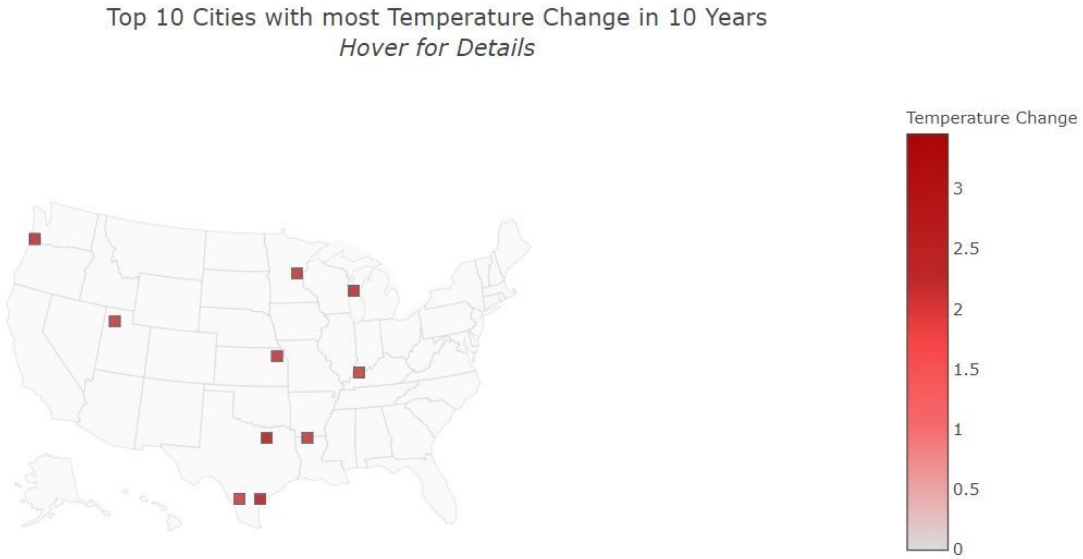
We have used libraries including pandas, numpy, matplotlib, sklearn, seaborn, plotly and statsmodels.

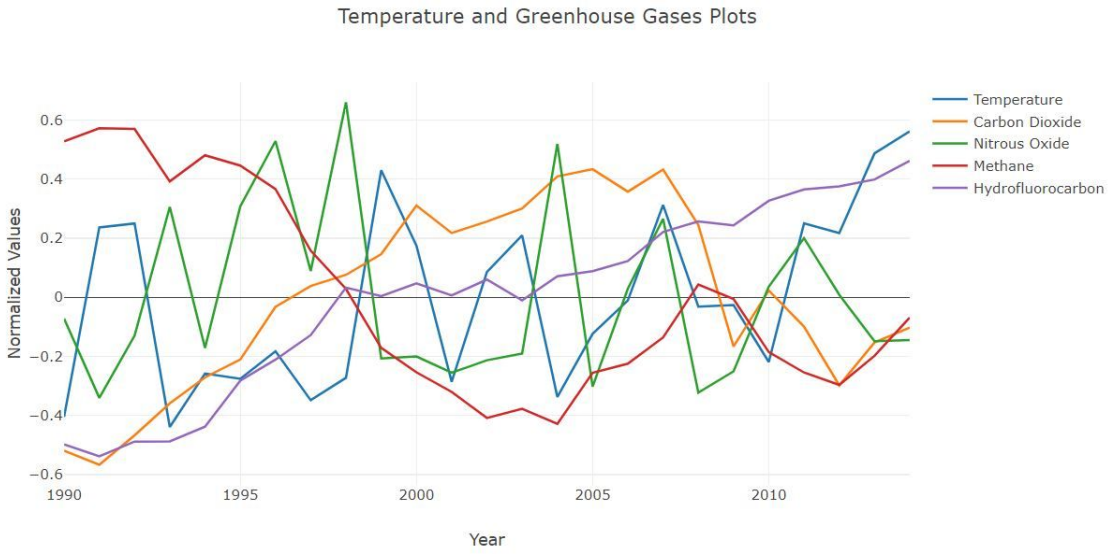
Additionally, we have also used Pearson's Correlation Coefficient to analyze the pairwise correlations between Pollution, Greenhouse Gas emissions and temperature change.

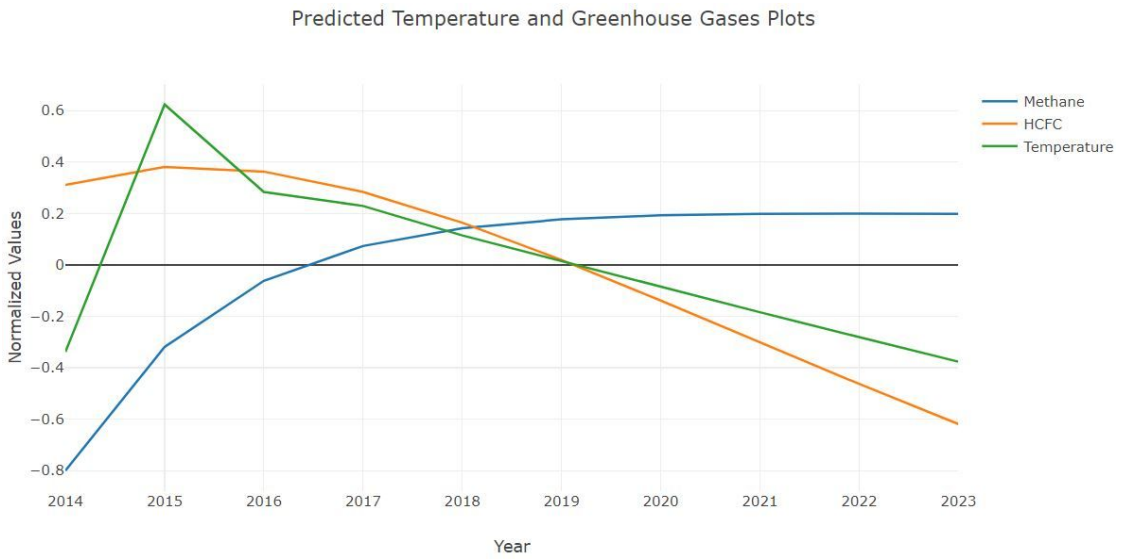


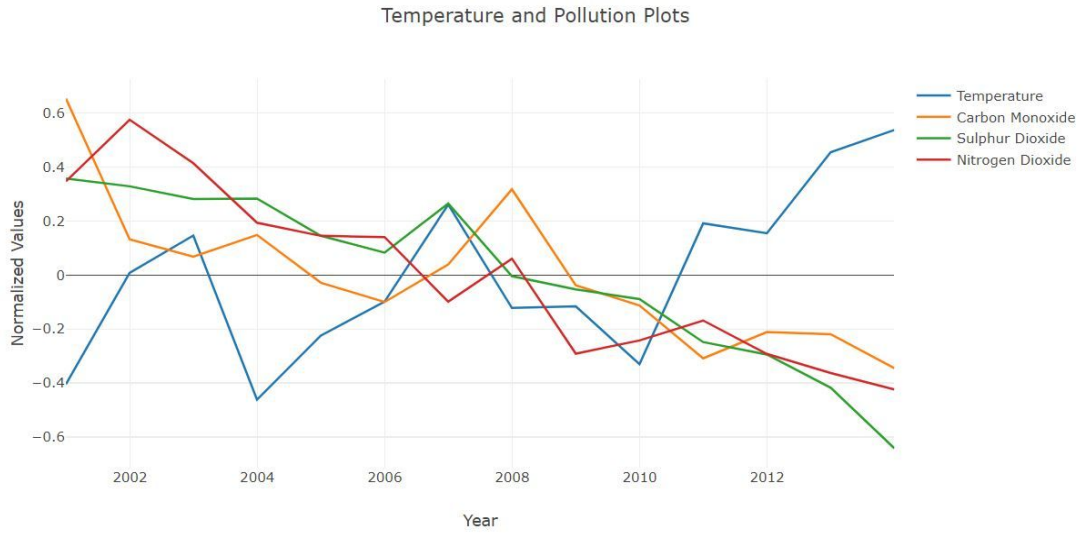
CHANGES MADE IN COMPARISON WITH THE ORIGINAL PROPOSAL

- We extended the scope of our project: instead of just forecasting temperature, we also identified the top-10 cities in the US that are likely to experience the most temperature change. We also analyzed the correlation between pollution levels and temperature, as well as the correlation between Greenhouse gas emissions and temperature, which helped us identify the Greenhouse Gas that has and will have the most impact on temperature change.









- To evaluate the model, we calculated both Mean Squared Error (MSE) as well as Mean Absolute Error (MAE), instead of just calculating MAE.

TEAM MEMBER EVALUATION

Both of us have **equally contributed** to this project. We have decided the following split:

- Ameya Vinod Shanbhag (avs431): 5
- Vikram Sunil Bajaj (vsb259): 5