CS-541 OPTIONAL PROJECT REPORT
SURFACE TEMPERATURE PREDICTION

Wyatt Blair
wblair@stevens.edu

Cindy Tran
ctran2@stevens.edu

Shubhankar Deol
sdeol@stevens.edu

Introduction

Climate change has far-reaching consequences that will eventually affect everyone. Therefore, it is crucial to convince legislators to implement policies which can limit the effects humans have on their environment. One metric, which captures a large portion of the issue, is carbon emissions. Our team has found two datasets consisting of historical data on global temperatures and global carbon emissions. Equipped with these datasets, we implemented an algorithm to predict future global temperatures given future carbon emissions. In doing so, we could predict the outcomes of not lessening carbon emissions.

Background

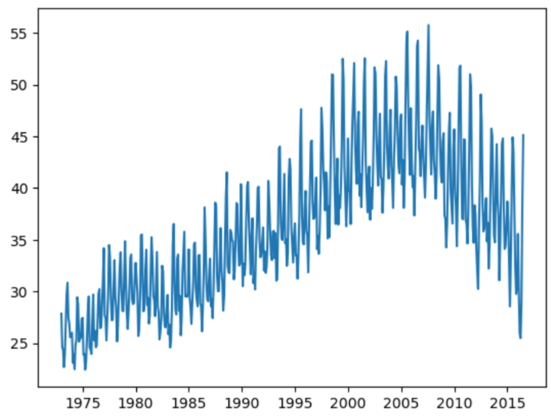
Because of global warming and climate change, there has been plenty of research into global temperature changes. In a research conducted by James Hansen, documented procedures for data over land, satellite measurements of sea surface temperatures, and ship-based analysis were used to graph temperature changes and also predict temperature change trends. Based on the results, warmth was found to be nearly ubiquitous with larger temperatures being reported over land compared to ocean and also the largest numbers being from high latitudes in the Northern Hemisphere. These results make sense as there is a lot of more human traffic that can contribute to rising temperatures on land compared to water. There are also a lot of developed countries in the Northern Hemisphere and these countries are more likely to contribute to the rising global temperatures. Research into temperature changes and carbon emissions are significant because they provide information on the changes that happen to Earth because of our actions. Based on the graphs and predictions, scientists can create corresponding efforts in order to reduce global warming.

Datasets Description

Our group will make use of two different datasets. The first is the carbon emissions dataset from Kaggle, which provides historical data on global carbon emissions broken down by region.

MSN	YYYY MM	Value	Column _Order	Descrip tion	Unit
CLEIE US	197301	72.076	1	Coal Electric Power Sectio CO2 Emissions	Million Metric Tons of Carbon Dioxide
CLEIE US	197302	64.442	1	Coal Electric Power Sectio CO2 Emissions	Million Metric Tons of Carbon Dioxide
CLEIE US	197303	64.084	1	Coal Electric Power Sectio CO2 Emissions	Million Metric Tons of Carbon Dioxide
CLEIE US	197304	60.842	1	Coal Electric Power Sectio CO2 Emissions	Million Metric Tons of Carbon Dioxide
CLEIE US	197305	61.798	1	Coal Electric Power Sectio CO2 Emissions	Million Metric Tons of Carbon Dioxide

As you can see from the first five dataset values, it provides us the MSN value, the year and the month in the format YYYYMM, the value of the amount of global carbon emissions, the column order, a description about where the carbon emission is coming from, and the unit used to measure the CO_2.

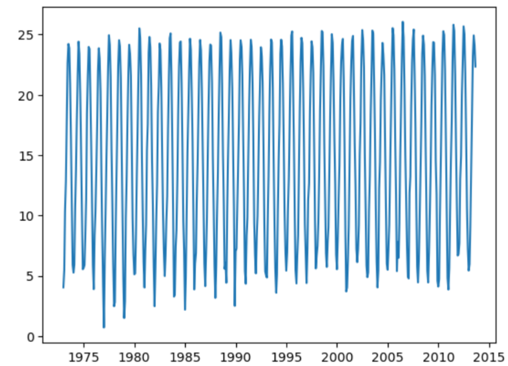


Along with this visual graph, you can see that over time, the global carbon emissions have steadily increased over time until it reached its peak in the early 2000s and then went down most likely due to the global warming conservation efforts.

The second dataset is the Earth surface temperature dataset from Kaggle, which provides historical global surface temperatures broken down by region.

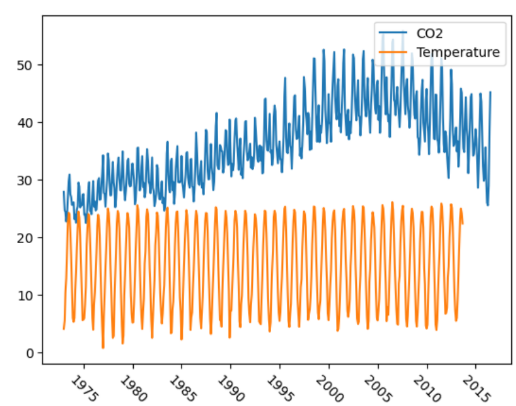
dt	AverageTemperature	AverageTemperatureUncertainty	City	Country	Latitude	Longitude
1743-11-01	6.068	1.737	Arhus	Denmark	57.05 N	10.33 E
1743-12-01	NaN	NaN	Arhus	Denmark	57.05 N	10.33 E
1744-01-01	NaN	NaN	Arhus	Denmark	57.05 N	10.33 E
1744-02-01	NaN	NaN	Arhus	Denmark	57.05 N	10.33 E
1744-03-01	NaN	NaN	Arhus	Denmark	57.05 N	10.33 E

As you can see from the first five dataset values, this dataset tells you the date, average temperature, the uncertainty of the average temperature, the city and country along with its latitude and longitude.



The visual graph for the surface temperatures shows that the temperatures fluctuate up and down between 5 to 25.

Putting the two datasets together gives us the graph that is shown below. By merging these two datasets together, we are able to create one master dataset that can extract any relationships between carbon emissions and global surface temperature. For the master dataset, we combined the two smaller datasets based on their timestamp.



The Model

The models that we decided to use on the dataset are the ARIMA and VAR models. The ARIMA or AutoRegressive Integrated Moving Average is a time series model that is used for forecasting

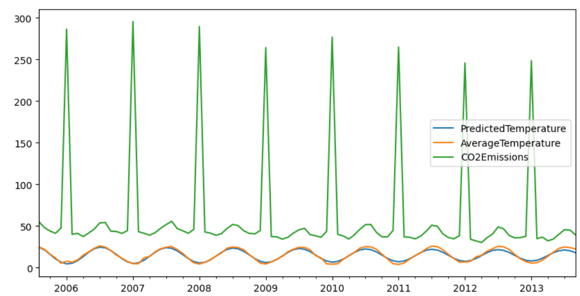
future values based on past observed values. It takes in account the correlation between successive observations and then uses the information to make predictions. The ARIMA model has three main components. These components are the autoregression (AR), differencing (I), and moving average (MA). The autoregression uses past values to predict future values. Differencing is used to remove the trend of seasonal components from the data. Moving average is used to model the errors or residuals between the predicted and observed values.
We also used the VAR (Vector AutoRegression) model for multivariate computation to generate predictions using the combined temperatures and CO_2 emissions data. The VAR model is a time series model that analyzes the relationship between multiple time series variables. The VAR model is represented as a system of k variables where each variable is modeled as a system of its own past variables and other past variables in the system. This model allows us the estimation of the dynamic interactions between the variables.

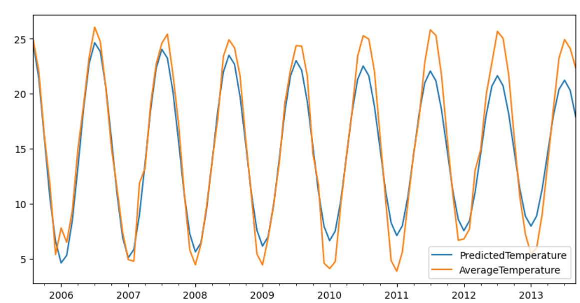
$$Y_{1,t} = \alpha_1 + \beta_{11,1} Y_{1,t-1} + \beta_{12,1} Y_{2,t-1} + \epsilon_{1,t}$$
$$Y_{2,t} = \alpha_2 + \beta_{21,1} Y_{1,t-1} + \beta_{22,1} Y_{2,t-1} + \epsilon_{2,t}$$

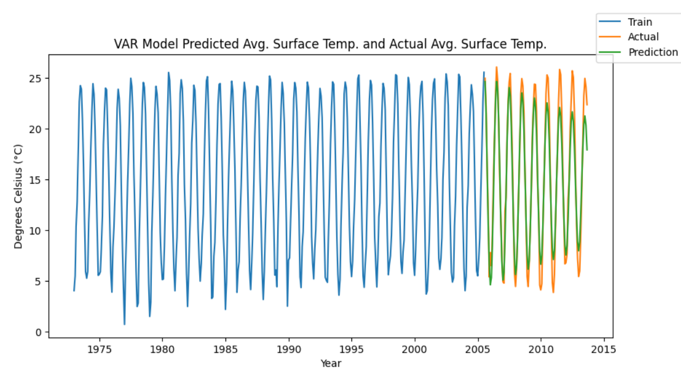
Each variable is assumed to be a linear function of past lags of other variables and itself. The covariance matrix is computed using the outer product of gradients.

Results

Both the ARIMA and the VAR model produced less than desirable results. The team believes that the root cause of the model's performance issues is the dataset itself. Both models predicted, essentially, a lower amplitude in seasonal temperatures over time, meaning the winters would be warmer and the summers would be colder. The figure below demonstrates the predicted values of average surface temperature from the VAR model as well as the actual data from those periods of time. As one can observe, the predicted values do not match the expected values closely, and therefore the forecasting predictions of this model cannot be trusted.







The VAR model had an RMSE value of 1.916 when the predicted Average Surface Temperature values were compared to the actual Average Surface Temperature values from the dataset. This portion of the data was withheld from the model during training.

Conclusion

In conclusion, the implementation of the VAR time series model in this experiment did not produce satisfactory results in predicting global surface temperature. The dataset used for CO2 emissions, which served as a key input for the model, appeared to be insufficient for capturing the complex dynamics of the relationship between CO2 emissions and global surface temperature. As a result, the predicted data did not align with the actual data. However, our team remains optimistic that with a more comprehensive and robust dataset, the VAR time series model could yield more accurate and reliable predictions. The limitations of the current dataset underscore the need for further research and data collection efforts to better understand the intricate factors influencing global surface temperature. Future studies should focus on acquiring more extensive and diverse datasets to enhance the model's predictive capabilities and shed light on the intricate dynamics of the climate system.

Future Work

Based on the experiments, it is found that multiple different variables can have significant correlations in periodically repeating trends. Global warming is not limited to CO_2 emissions and average surface temperatures. Using additional data including renewable energy usage trends we can create a state based agent that suggests optimal policies (e.g. recommended increase in usage of renewable energy sources/strategic continuous retreat from non-renewable energy sources) based on target temperature and current trend of temperature and resource usage.