Climate Change:
Predicting Earth Surface Temperature

Mike Wu

Abstract:

This report is about predicting the Earth surface temperature using machine learning techniques, specifically using ARIMA (Auto Regressive Integrated Moving Average) model to a time series data. Through using the dataset from past years, from 1750 to 2015, to predict the temperature afterwards. This is significant as it could provide us insights about the rapid pace the earth is facing global warming and remind us to take actions to mitigate this issue. However, the ultimate results may not perfectly fit in nowadays' data since there are lots of external factors including changing climate policies, model limitations.

Background information:

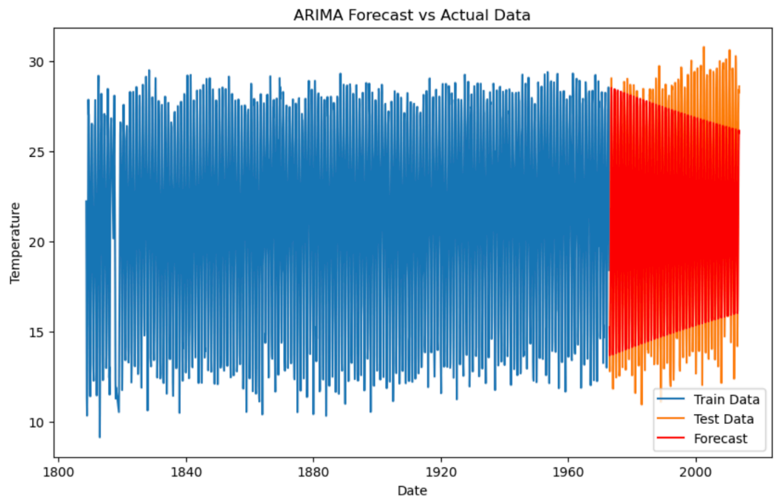
Although planet Earth has undergone several major changes in its history, the era with the most changes began with the appearance of human beings. We are the one who have brought a great imbalance caused by wars, resources exploitation, contamination. All these activities have caused serious environmental problems, one of them being global warming, which leads to climate change. Climate change (United Nations, n.d.), according to United Nation, is the long-term changes in temperatures and weather patterns in the earth. Meanwhile, according to the World Health Organization (WHO, 2023.10.12), climate change is impacting health in a myriad of ways, including by leading to death and illness from increasingly frequent extreme weather events, such as heatwaves, storms and floods, the disruption of food systems, increases in zoonoses and vector-born diseases, and mental health issues. Consequently, we need to take it seriously.

Methods:

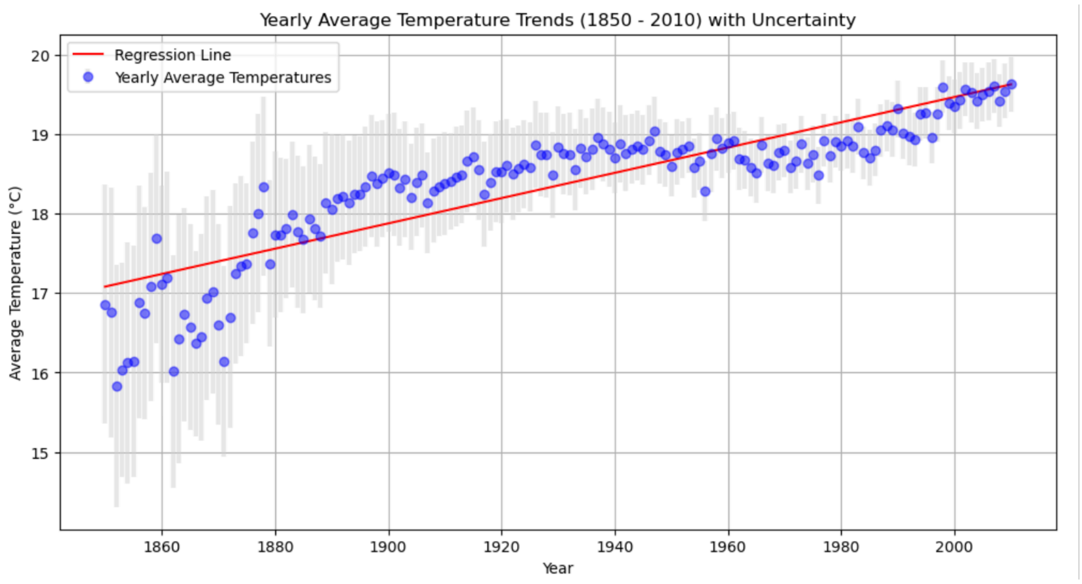
I used the dataset from one of the previous Kaggle competition about earth surface temperature. The dataset is from Berkeley Earth, which combines 1.6 billion temperature reports from 16 pre-existing archives. The datasets contained temperature data from 1750 to 2015 monthly. After the prediction made by the model, we can compare the temperature from the model and real-life to check the correctness of the ARIMA model. Apart from using typical Python-related tools such as numpy, pandas and visualization tools such as matplot, I also used ARIMA (Autoregressive integrated moving average), which is commonly used in statistics and econometrics in time series analysis. To better comprehend the data or forecast upcoming series points, this model is fitted to times series data, which is the case for this project of predicting the earth surface temperature. An ARIMA model is a form of regression analysis (Haynes, n.d.) that gauges the strength of one dependent variable relative to other changing variables.

Results:

After the exploratory data analysis on the datasets, we could observe the trends and characteristics of the temperature data across times. Meanwhile, when I applied the ARIMA model to the times series data and split the time series into training and testing, I got a graph demonstrating the ARIMA forecast data VS actual data.



From the above graph, the blue part represents the historical temperature data that was used to train the ARIMA model. The orange part represents the test data that was used to evaluate the model's predictive performance. The red part is the forecasted data generated by the ARIMA model based on the patterns learned from the training data. Theoretically, for the forecast and actual part, the closer they are, the better the model is at making predictions. However, after analyzing the graph, I figured out that there are minor discrepancies between the test data and forecast data. Also, more significantly, for the forecast data, the highest temperature is showing a trend of decreasing which does not align with the reality. Consequently, I decided to use linear regression to predict the results again. After setting up the parameters and fit the model, I got the coefficient of the linear model as 0.015 and intercept as -10.629. The coefficient value represents the estimated change in average temperature for each additional year in terms of degrees Celsius. The positive coefficient suggests a long-term trend of rising temperatures, which is consistent with what nowadays is going on.



This is the graph of the linear regression method used to predict the future earth surface temperature.

Discussion:

First of all, I think there are overall three reasons why at first the ARIMA model could not capture the real trends of temperature rise. It could be due to the forecast horizon that ARIMA model is more accurate for short-term forecasts. As the forecasting horizon extends, the cumulative error can increase, leading to a divergence from actual values. Meanwhile, certain external factors could affect also. Some sudden climatic events or policy change such as the shift to renewable energy could lead to changes in temperature trends and the model could not anticipate this change. The result of the linear regression gives us a sense of how rapid the earth is experiencing temperature rise that we need learn more about this issue and act.

If I want to continue working on this project, I think we should train and run the model with input that reflects different hypothetical emission scenarios. According to Caltech research scientist Anna Jaruga (Anna Jaruga, n.d.), the scenarios estimate emissions of greenhouse gases and smog-forming air pollution that people will generate in various areas on Earth. This approach helps identify the range of potential future climate conditions and better inform our decisions and preparations.

References:

United Nations. (n.d.). *Climate change*. United Nations. https://www.un.org/en/global-issues/climate-change

World Health Organization. (n.d.-a). Climate change. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health

Hayes, A. (n.d.). Autoregressive integrated moving average (ARIMA) prediction model. Investopedia. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

How do we predict climate change?. Caltech Science Exchange. (n.d.-b). https://scienceexchange.caltech.edu/topics/sustainability/climate-change-predictions