HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 4 (2022)

Exploratory Data Analysis to Understand the Causes of Global Warming and Application of Soft Computing Techniques to Develop Its Forecasting Model

Aarush Mahajan[1] and Reetu Jain[#]

[1] Lotus Valley International School, Noida, India
[#]Advisor

ABSTRACT

Global warming (GW) is one of the major effects of human activity where excessive use of fossil fuels as energy sources has led to an increase in the concentration of greenhouse gases (GHGs), such as CO2, CH4, and water vapour, in the atmosphere one of the main reason to increase the average surface temperature. This study analyzes the time-series data to come to a rational conclusion about the role of GW in increasing sea-water level, the reason for the increase in GHG and the correlation of GHG to GW. In this direction time-series analysis is carried out on four different datasets. The first and second dataset comprises global temperature anomalies data and the cumulative changes in seawater level for the world's oceans since 1880. The third and fourth dataset comprises the records of concentration of GHGs in the atmosphere since 1st AD and the last 4 ice age years respectively. Finally, forecasting models are developed based on Holt's and SARIMA techniques to predict the global temperature anomaly, the concentration of GHGs and their correlation with GW. The developed models showed 74.6%, 94.5% and 95.7% accuracy in predicting temperature anomaly, CO_2, and CH_4 concentration in the atmosphere respectively. The strength of the forecasting model is its ability to compute the critical values of the factors. Therefore, the forecasting models are applied to predict the year in which the critical values of the factors contributing to GW will be attained.

Introduction

"Global warming" (GW) is the result of large-scale deforestation and the burning of fossil fuels such as coal, natural gas and oil, resulting in methane (CH4), carbon dioxide (CO2), nitrous oxide (N2O) and water vapour. Greenhouse gases from layers above the Earth absorb infrared radiation emitted by the Earth's surface to keep it warm. As the population grows, the rate of deforestation and the burning of fossil fuels increases, leading to an increase in global temperature. GW is a major contributor to climate change (1). Frequent heat waves, increased rainfall, and melting glaciers It is related to the consequences of climate change (2). Due to the negative impact of HW on human communities and ecosystems, this is one of the most important environmental problems facing the world. Many modern literary works have blamed GW for increasing global economic inequality (3).

Literature Review

HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 4 (2022)

Although GW is caused due to human activities, it can also be the result of natural events such as volcanic eruptions (VE) and variation in solar radiation on earth. VE significantly affects the temperature of the earth. During VE, many gasses, especially the sulphuric gasses form clouds which can cause a reduction in global temperature up to 3 years (4). Above that, VE emits large amounts of CO_2 and water vapour which are GHG and increase the earth's temperature (5). At an estimation, VE emits at an average of 130 to 230 million tons of CO_2 whereas human activities produce around 26 billion tons (6).

The third reason that is responsible for GW is the Anthropogenic emissions which are caused due to emission of GHG such as water vapour, CH_4, CO_2 and N_2O. Water vapour is the most abundant gas in the atmosphere and it is also responsible for two-thirds of the GW. The atmosphere maintains an equilibrium between temperature and the water vapour concentration due to the short life cycle of water vapour. However, as the temperature continues to increase the balance will be lost and will increase the global warming where water vapour has the ability to double the warming caused by carbon dioxide (7).

With the increase in population, the CO_2 concentration in the atmosphere is increasing. The CO_2 concentration in the atmosphere is mostly due to the burning of fossil fuels. Since 1950, CO_2 concentration has increased in the atmosphere by 30%. With the increased dependency on energy which is mostly generated from fossil fuels, an average of 45% of CO_2 emissions came from coal burning, 35% from oil burning, and 20% from natural gas burning (8). Deforestation is another reason for the increase in CO_2 concentration. Deforestation is ceasing the absorption of carbon by trees which results in a 25 - 30% annual increase in GHG (9). CO_2 needs 5 - 200 years to adjust and achieve the balance and as the percentage of carbon dioxide increases, the balance will happen at higher temperatures and at higher water vapour levels (6). Hence researchers believe that CO_2 is a controlling factor as it has the capability to control the amount of water vapour in the atmosphere (10).

CH4 emission is the second largest anthropogenic contributor to GW which has doubled its concentration in the last 150 years (11). CH4 molecule has the ability to absorb and reradiate the energy 10 times more effective than the carbon dioxide molecule. Out of all the sources of CH4 emissions, almost 60% come from human activities. The primary source of CH4 emissions in the atmosphere is during the extraction, production, transportation, refining, and distribution of natural gas.Also, a significant amount of CH4 is released into the atmosphere from livestock agriculture, human waste, and landfills (12). Figure 1 shows the pie chart of the global CH4 emissions from 2003 to 2013.

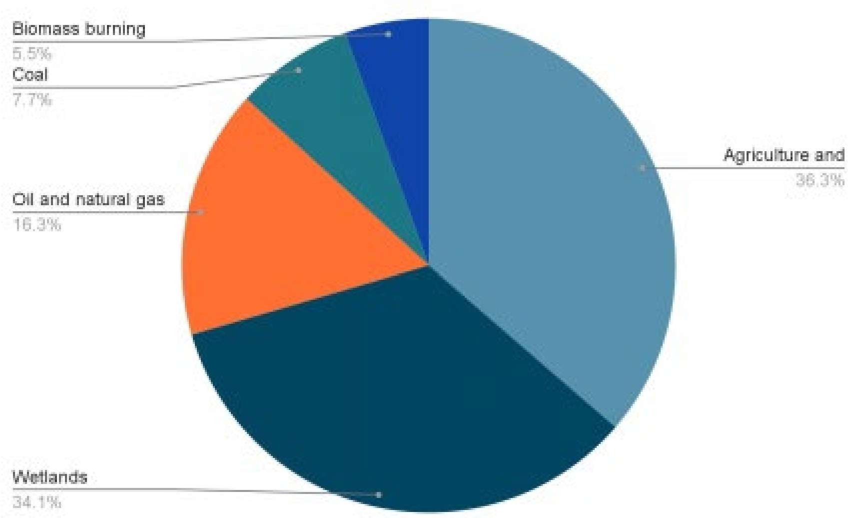

Figure 1. Pie chart of the global CH_4 emissions from 2003 to 2013

Activities such as fossil fuel combustion, agriculture, industrial processes, and wastewater management are responsible for about 40% of the total N_2O emissions (13). Table 1 shows the global N_2O emissions by sector in 2010.

Table 1. Global N_2O emissions by sector in 2010

Sector	**%age of N_2O emissions**	**Sector**	**%age of N_2O emissions**
Agriculture	72.22	Transport	0.52
Residential and commercial	2.98	International bunkers	0.29
Energy	5.75	Land use	2.73
Industry	4.51	Waste	3.89
Other sources	7.09	—	—

According to the current trend in the rate of increase of N_2O per year it is estimated that by the year 2100 the predicted N_2O emissions will be 25.7 Tg (14). N_2O stays approximately 114 years in the atmosphere and it is removed as a part of the nitrogen cycle by certain bacteria or destroyed by chemical reactions or by ultraviolet radiation. N_2O has the ability to warm the atmosphere almost 300 times more than CO_2; however, the concentration of N_2O is much smaller than CO_2 (13).

Because of its negative impacts on human communities and ecosystems, GW is the most important environmental problem in the world faces. Adaptation to the inevitable impacts and mitigation to reduce their magnitude are both necessary. International action is being taken by the world's scientific and political communities. Because of the need for urgent action, the greatest challenge is to move rapidly to increased energy efficiency and to non-fossil-fuel energy sources.

To determine the next steps in climate change mitigation, it is necessary to examine the main drivers of climate change to see how important each factor is. This study will focus on these factors as well as many other factors that can cause differences in global temperature.

Meanwhile, machine learning is increasingly being applied to environmental problems and is showing promising results. As stated in the referred text (15), how to use the dual parallel feed-forward neural networks (NF) to evaluate suspended sediment content that in turn contributes to water resource management. The referred text (16) compared the properties of different NNs in terms of suspended sediments in river systems. Artificial neural networks (ANNs) have also been studied to evaluate the energy consumption and environmental life cycle of incineration and waste disposal and landfill systems (17). In (18), NN is combined with elementary stream segmentation and binary-coded cluster optimization to predict the river quantities. The theory of variable fuzzy sets and fuzzy binary comparison methods have been investigated in assessing water quality (19). Those works demonstrate the applicability of machine learning techniques on environmental issues.

Motivation and Novelties

The primary intention of the present study is to conduct a time-series analysis of the collected data to justify and validate the followings:

I. Polar ice melting is the result of GW.
II. An increase in sea level is due to GW.

III. Develop a time-series forecasting model to predict the impact of different factors on GW

In order to achieve the desired objectives of the paper, the data are analyzed and modelled using time-series forecasting methods namely Holt's and seasonal auto-regressive integrated moving average (SARIMA) techniques. These models are employed to investigate the effects different factors have on global temperature. Then, the plots generated from the algorithms are analyzed to derive a rational conclusion.

The remainder of the paper is organized as follows:

- Section 2 summarizes the methodologies adopted for analyzing and modelling the data.
- Section 3 describes the problem statement and briefly describes the data collected for fulfilling the aims of the paper.
- Section 4 summarizes and briefly describes the result obtained from the analysis which is followed by
- Section 5 which is the conclusion and future scope.

Methodology

This section of the paper briefly describes the preliminary concept required to analyze the data and subsequent development of the forecasting models. The modelling is done using the forecasting techniques namely Holt's model and SARIMA model.

Holt's Forecasting Model

Holt forecasting model, also known as linear exponential smoothing, is a popular smoothing model for predicting trend data. A Holt model consists of three separate equations that work together to produce the final prediction. The first is a basic smoothing expression that directly adjusts the last smoothed value for the last period trend. The trend itself is updated over time using the second equation, where the trend is expressed as the difference between the last two smoothed values. Finally, the third equation is used to generate the final prediction. The Holt model takes two parameters. One is normal smoothing and the other is the trend smoothing equation. This method is also known as double exponential smoothing or trend boosting exponential smoothing.

SARIMA Forecasting Model

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. ARIMA models are applied in some cases where data shows evidence of non-stationarity in the sense of mean (but not variance/autocovariance), where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity of the mean function (i.e., the trend) (20). When the seasonality shows in a time series, the seasonal-differencing (21) could be applied to eliminate the seasonal component. However, in case frequent seasonal effects come into play for time-series analysis then SARIMA model is employed to derive a rational conclusion.

Case Study

In this section of the paper, a brief description of the case study along with the different assumptions and dataset is discussed.

Problem Statement

In this paper, the aim is to validate GW based on the collected data from the NASA repository. Also, the study validates that the polar ice is melting due to GW and it is the reason for increasing sea levels. Moreover, the study analyses the concentration of GHG and temperature of the data for the last 400,000 years and correlates it with GW.

Datasets

Four different datasets were used in this paper to develop the forecasting models. The first dataset comprises global temperature anomalies data since 1880. A temperature anomaly is a deviation from the positive or negative average temperature over a period of the year. The second data set contains the cumulative change in global sea level since 1880, which is based on a combination of long-term tidal measurements and recent Satellite measurements. Shows the average absolute change in sea level, representing sea level height, whether the surrounding land is rising or falling. While satellite data is based only on measured sea level, long-term high tide data contain small correction factors because the size and shape of the oceans change slowly over time.The third dataset comprises the records of CO_2, CH_4 and N_2O concentration since 1 AD. The fourth and final dataset covers the last 4 ice age cycles. Data can be used to study natural atmospheric gas levels, dust levels, and natural variation in temperatures. The natural variations over the last 4 ice ages would be useful for comparison with the anthropogenic climate-change data available for the modern industrial period. The dataset contains information on GT4 ice core chronology (gas and ice chronology), Deuterium and reconstructed temperature, Dust content, Sodium concentrations, CO_2, CH_4, atmospheric oxygen composition.

Results and Discussions

In this section of the paper, the results obtained from analyzing the datasets and a brief discussion is presented.

Validating the GW

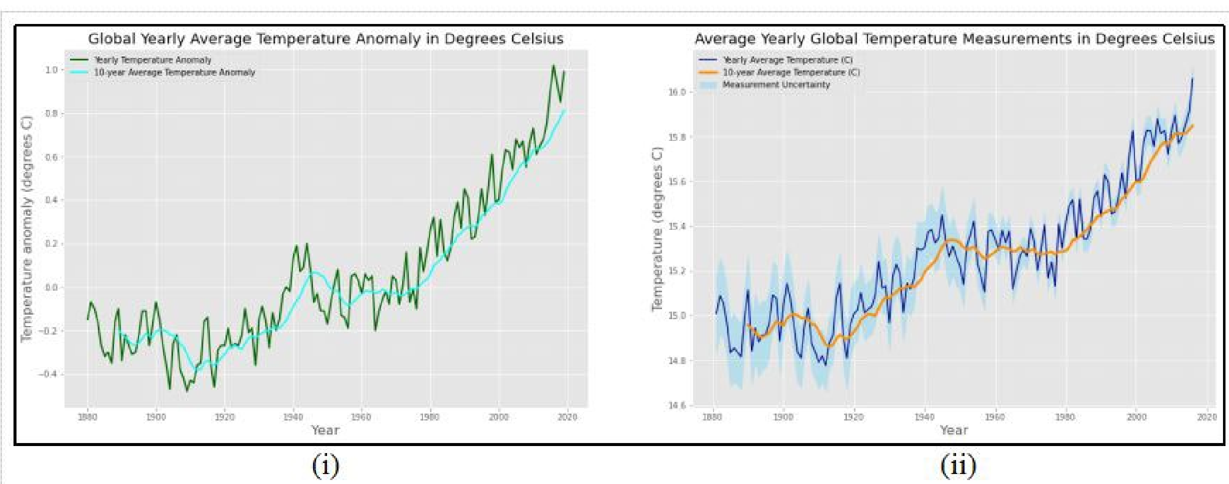

Figure 2. Time-series plot for global average (i) temperature anomaly and (ii) temperature

Time series analysis is carried out for the first dataset. The time series plot for the dataset is shown in figure 2. The global average temperature anomaly [figure 2(i)] displays a steady increasing trend beginning around 1980. The global average temperature anomaly in January of 1980 and 2018 (end of the dataset) is 0.26 C and 0.85 C, respectively. Therefore, the global average temperature anomaly increased by 0.0155 C per year over that span of 38 years. However, the temperature anomaly over the past 18 years (2000 to 2018) has been increasing at a rate of 0.0256 C per year, indicating that the global average temperature anomaly is increasing at a non-linear rate. The figure 2(ii) graph displays the actual global yearly average temperature measurements in degrees Celsius. The upward trend in the temperature anomaly graph [figure 2(i)] beginning around 1980 is seen in the yearly average measurements at the same time. Thus, temperature anomaly is not an arbitrary transformation of temperature data.

Validating The Increase in Sea Level

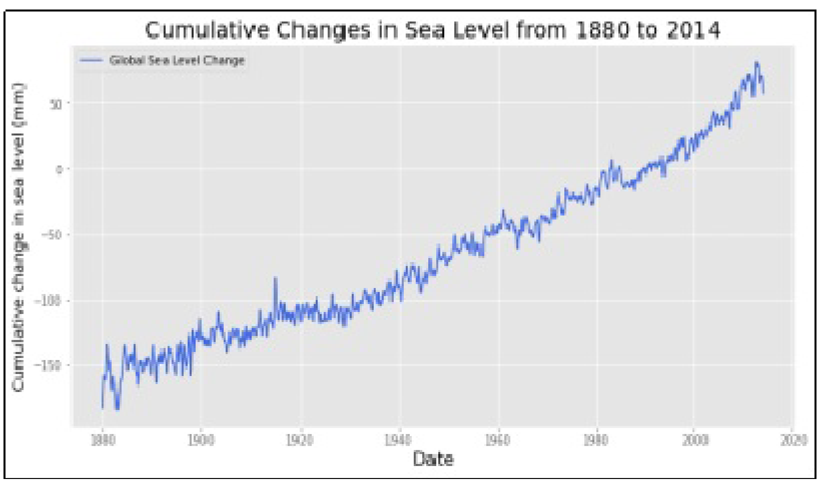

Figure 3. Time-series graph of cumulative change in sea level

There is a clear, upward trend for the data representing the cumulative sea level changes over time. According to this data, the yearly average sea level has risen 1.75 mm per year since 1900.

Analyzing The Time-Series Data of GHG

The values of atmospheric concentrations of carbon dioxide and methane remained relatively stable from 0 CE to 1760 CE (the beginning of the industrial revolution) and then began to increase exponentially. This is not surprising as the production and usage of GHG producing machines rose from that time.

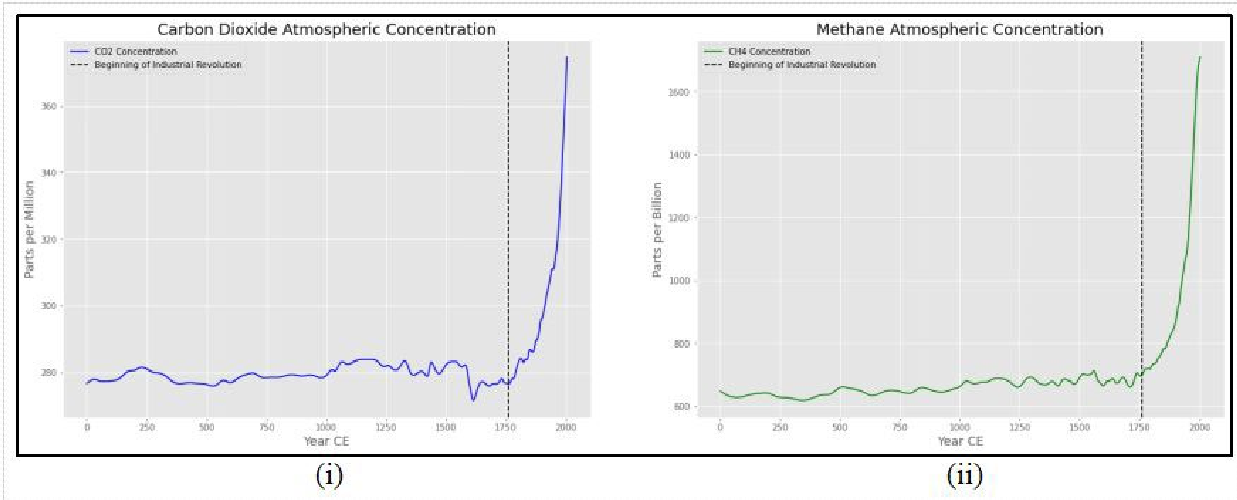

Figure 4. Time-series graph of (i) CO_2 and (ii) CH_4 concentration

Correlation Between Increase in Global Temperature and Greenhouse Gas Concentration

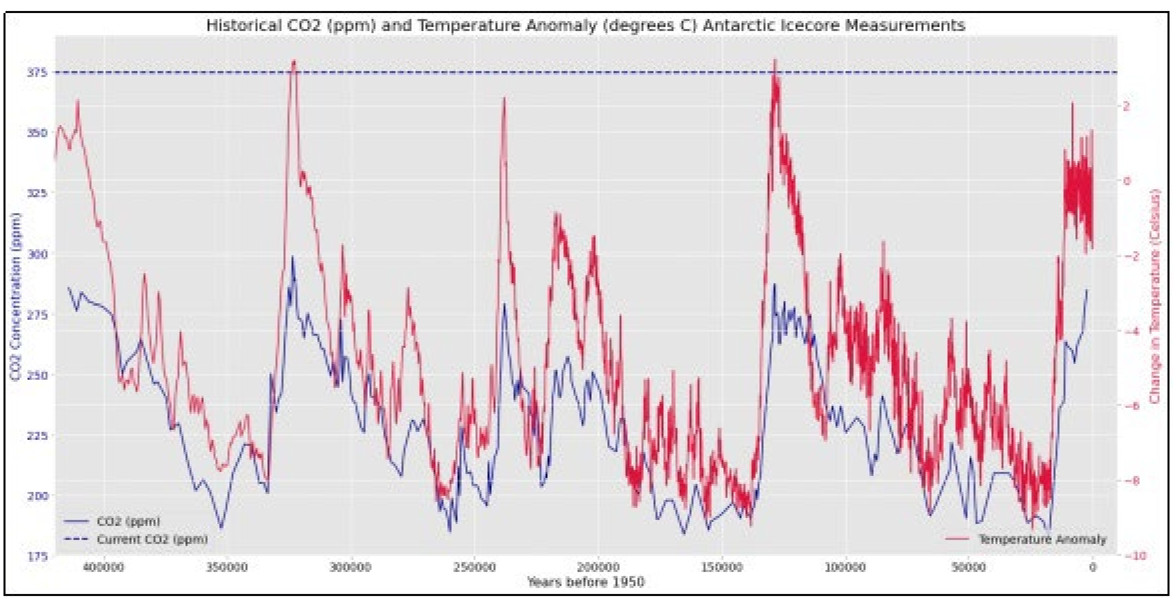

Figure 5. Time series data of CO_2 concentration and temperature anomaly

HIGH SCHOOL EDITION
Journal of Student Research

The temperature anomaly and carbon dioxide measurements seem to correlate over the past 400,000 years, indicating that there is most likely a connection between carbon dioxide and temperature changes. The dashed line at the top of the graph represents the current atmospheric carbon dioxide value and it is almost 75 ppm greater than any value in the past 400,000 years.

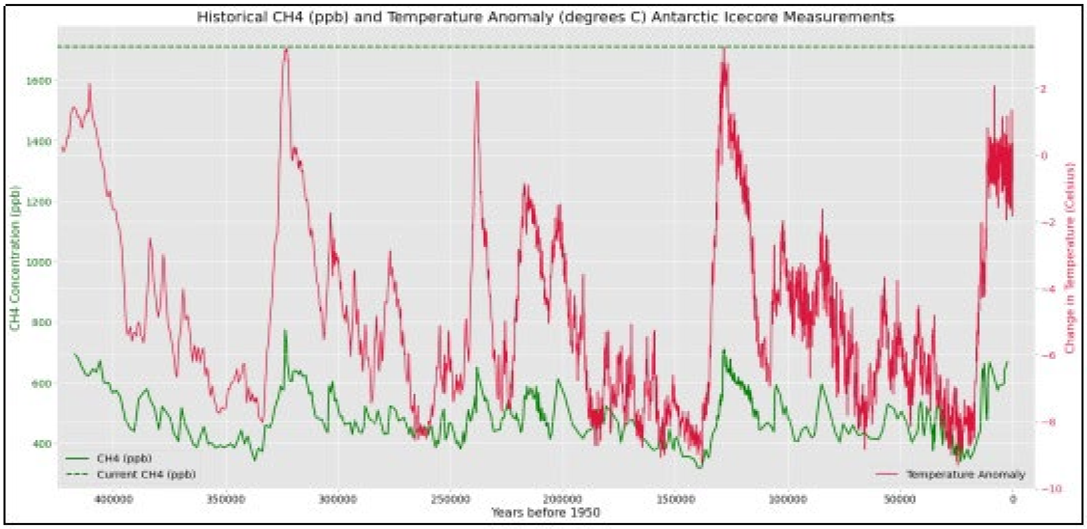

Figure 6. Time series data of CH_4 concentration and temperature anomaly

Temperature anomaly and methane measurements also seem to correlate over the past 400,000 years, indicating that there is most likely also a connection between methane and temperature changes. The dashed line at the top of the graph represents the current atmospheric methane concentration and it is more than 2 times greater than any value in the past 400,000 years.

While the Granger Causality Test did not yield any significant results, the historical correlation of carbon dioxide and methane with temperature changes over the past 400,000 years seems to be very strong. The insignificant results may be due to the bidirectional correlation between GHGs and temperature suggested by the climate science community and the assumptions of the Granger Causality Test. Rising temperatures results in increased ice melt, which results in the release of GHGs stored in the now-melted ice. As one of the assumptions of the Granger Causality Test is that the cause has to precede the effect, increasing temperatures occurring before increasing GHGs from ice melt would violate this assumption since we formatted the data to test if rising GHGs caused a rise in temperature. As most climate scientists agree on the causal role of GHGs in global warming trends, further testing should be performed to illuminate this role.

Developing The Forecasting Models

100 Year Forecast-Ed Temperature Anomaly Model

The temperature anomaly model predicted with 74.6% accuracy on the test set and indicates that the critical value of 2 degrees C will be reached in the year 2081.

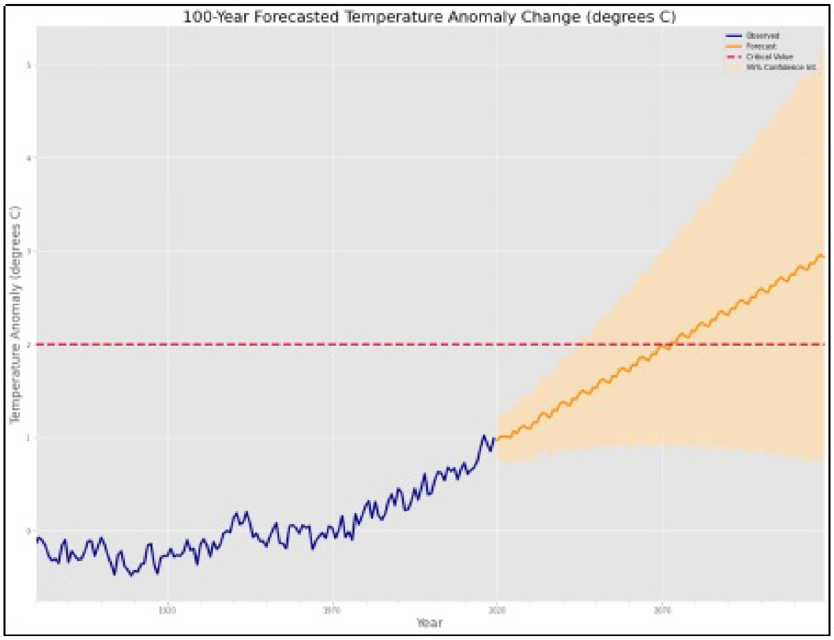

Figure 7. 100 year forecast-ed temperature anomaly

Carbon Dioxide Model

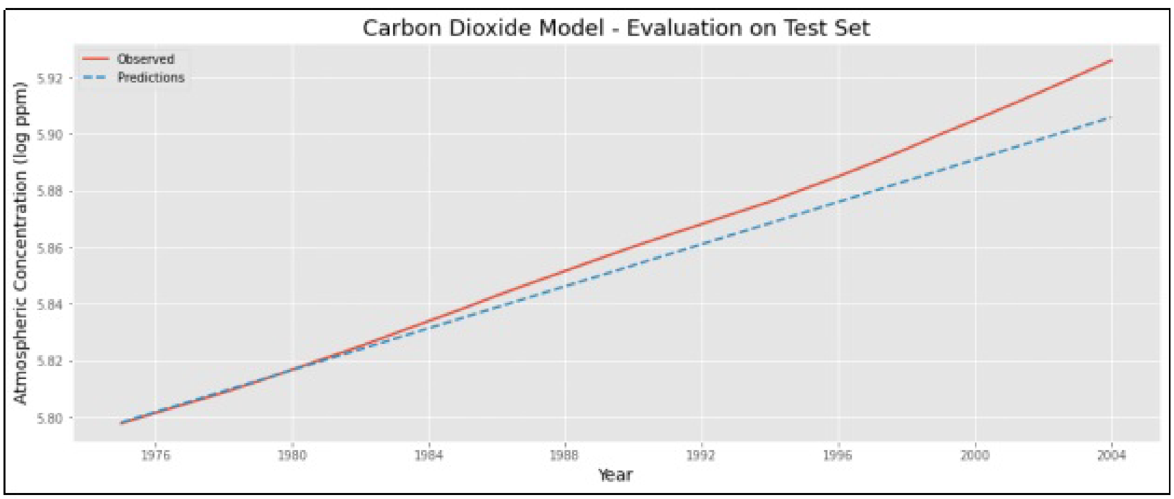

Figure 8. CO_2 model

The Holt model for carbon dioxide concentration had an R-squared score of 0.945 when evaluated on the testing set and the root mean squared error of its predictions were less than the standard deviation of the testing set. These measures indicate that the model performs with roughly 94.5% accuracy and should provide reasonable predictions for future carbon dioxide concentration values. However, based on the model's performance on the testing set, the forecast-ed carbon dioxide concentration values may be lower than the actual values as the model seemed to consistently predict values lower than the testing set.

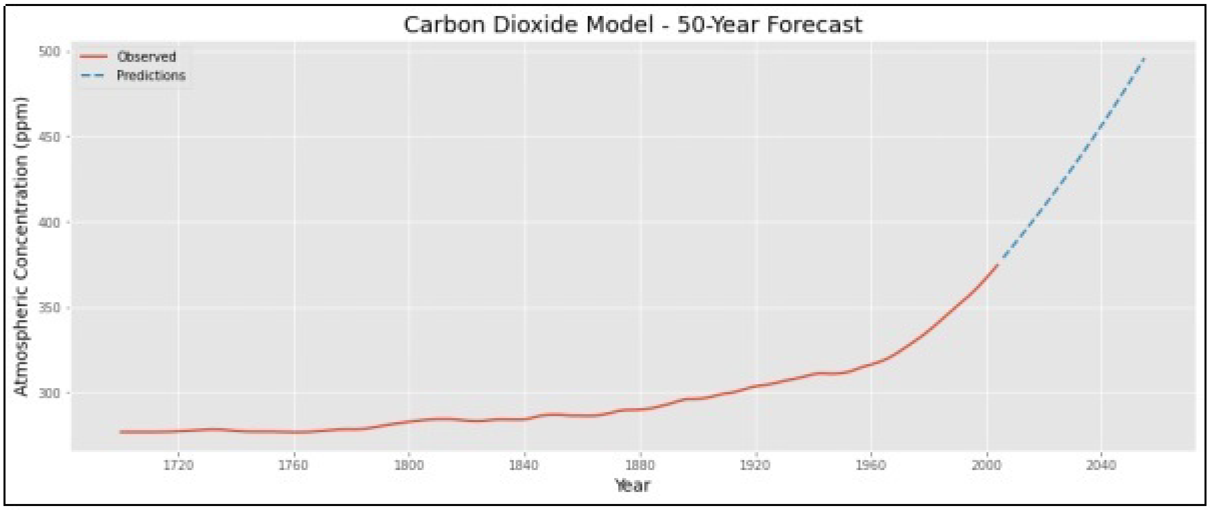

Figure 9. CO_2 forecasting model

The model predicts that the atmospheric carbon dioxide concentration will continue rising and it is expected to reach the critical value of 450 ppm by 2038. However, as the model tended to predict values lower than the observed measures when evaluated on the testing set, the 450 ppm concentration may be reached sooner than 2038.

Methane model

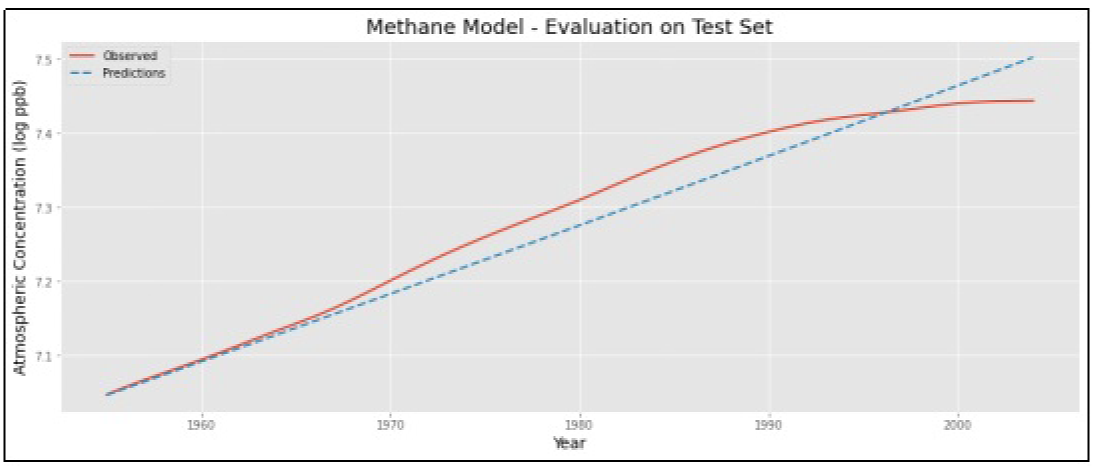

Figure 10. CH_4 model

The Holt model for methane concentration had an R-squared score of 0.957 when evaluated on the testing set and the root mean squared error of its predictions were less than the standard deviation of the testing set. These measures indicate that the model performs with roughly 95.7% accuracy and should provide reasonable predictions for future methane concentration values. However, based on the model's performance on the testing set, the forecast-ed methane concentration values may be higher than the actual values as the model seemed to predict a constant increase in the values, whereas the values of the last 15 years in the testing set seem to show the concentration beginning to level off.

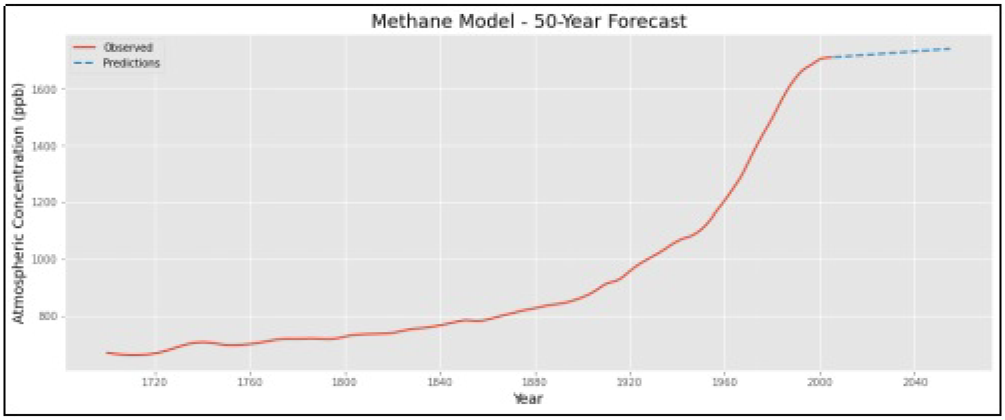

Figure 11. CH_4 forecasting model

The model predicts that the atmospheric concentration of methane will continue to rise, but at a much slower rate than carbon dioxide. However, as the model tended to predict values greater than the observed measures in the testing set, the actual rate of increase may actually be lower than predicted.

Conclusion

This study primarily focuses on developing forecasting methods based on Holt's and SARIMA techniques. Although there are many researches that develop forecasting models to predict the GW, yet there are very few literature's that validates GW and investigate the impact of different factors on it. Above that, this paper proposes a forecasting model to predict the rate of temperature anomaly for 100 years and CO_2 and CH_4 concentration for next 50 years. Moreover the study provides a valid analysis that could answer the questions about the validation of GW and its role in increasing sea-water level, reason for increase in GHG and correlation of GHG on GW. In this process, the study is divided into five phases of analysis.

In order to validate the GW, the time-series global temperature anomalies data since 1880 is analyzed. The result obtained from the analysis concludes that GW is not a myth. It also confirms a steady growth of global temperature anomaly of 0.26^0C till 1980 which has increased to 0.58^0C since then. Since 2000, the average rate of global temperature anomaly is 0.0256^0C per year as compared to 0.0155^0C per year before 2000. This indicates that the rate of growth of GW is increasing at a non-linear rate. The next phase of the paper validates that GW is the major reason for increase of sea-level and melting of polar ice. The third phase analyzes the concentration of GHG in the atmosphere and concludes that with the starting of the industrial revolution since 1760 the concentration of GHG has increased exponentially. The fourth phase analyzes the correlation of the global temperature anomaly with concentration of the GHG. The analysis concludes that there is some correlation between global temperature and concentration of the GHGs, however, on statistical scrutiny it is concluded that the correlation is not significant. The fifth phase of the study develops the forecasting model by Holt's and SARIMA techniques. The conclusion that can be derived for the models is as follows:

a. The temperature anomaly model predicted with 74.6% accuracy that the critical value of 2^0C will be reached in the year 2081.
b. The CO_2 model predicted with 94.5% accuracy that the critical value of 450 ppm will be reached by the year 2038.
c. The CH_4 model predicted with 95.7% accuracy that the CH_4 will continue to rise but at a slower rate than CO_2.

From the overall discussion and the result obtained from the analysis it can be concluded that the study has successfully been able to fulfill the aims and objectives of the research with which it was adopted.

Acknowledgment

We would like to express my heartfelt gratitude to all the tutors and mentors of On My Own Technology pvt. ltd. for extending their help in carrying out the particular project. It is because of their help that we are able to conduct the research. We shall remain ever grateful for their help and generosity.

Conflict of interest

The authors would like to declare that no funding in any form is received for carrying out this research work.

References

1. Nema, P., Nema, S., & Roy, P. (2012). An overview of global climate changing in current scenario and mitigation action. Renewable and Sustainable Energy Reviews, 16(4), 2329-2336.

2. Singh, B. R., & Singh, O. (2012). Study of impacts of global warming on climate change: rise in sea level and disaster frequency. Global warming—impacts and future perspective.

3. Diffenbaugh, N. S., & Burke, M. (2019). Global warming has increased global economic inequality. Proceedings of the National Academy of Sciences, 116(20), 9808-9813.

4. The Effect of Volcanoes on the Earth's Temperature; http://www.climatedata.info/forcing/volcanoes/ (assessed on 1st April, 2022)

5. Man, W., Zhou, T., & Jungclaus, J. H. (2014). Effects of large volcanic eruptions on global summer climate and East Asian monsoon changes during the last millennium: Analysis of MPI-ESM simulations. Journal of Climate, 27(19), 7394-7409.

6. Houghton, J. T., Ding, Y. D. J. G., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., ... & Johnson, C. A. (Eds.). (2001). Climate change 2001: the scientific basis: contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge university press.

7. Riebeek, H. (2010). Global warming: Feature articles.

8. IEA, P. (2016). CO2 Emissions from fuel combustion 2016. Paris: IEA.

9. Martin, A. R., & Thomas, S. C. (2011). A reassessment of carbon content in tropical trees. PloS one, 6(8), e23533.

10. Anderson, T. R., Hawkins, E., & Jones, P. D. (2016). CO2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models. Endeavour, 40(3), 178-187.

11. Solomon, S., Manning, M., Marquis, M., & Qin, D. (2007). Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC (Vol. 4). Cambridge university press.

12. Environmental Protection Agency (EPA). Overview of Greenhouse Gases n.d. https://www.epa.gov/ghgemissions/overview-greenhouse-gases#methane (assessed on 1st April, 2022).

13. Anderson, B., Bartlett, K. B., Frolking, S., Hayhoe, K., Jenkins, J. C., & Salas, W. A. (2010). Methane and nitrous oxide emissions from natural sources.

14. Kroeze, C. (1994). Nitrous oxide and global warming. Science of the total environment, 143(2-3), 193-209.

15. Chen, X. Y., & Chau, K. W. (2016). A hybrid double feedforward neural network for suspended sediment load estimation. Water Resources Management, 30(7), 2179-2194.

16. Olyaie, E., Banejad, H., Chau, K. W., & Melesse, A. M. (2015). A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. Environmental monitoring and assessment, 187(4), 1-22.

17. Nabavi-Pelesaraei, A., Bayat, R., Hosseinzadeh-Bandbafha, H., Afrasyabi, H., & Chau, K. W. (2017). Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management-A case study in Tehran Metropolis of Iran. Journal of cleaner production, 148, 427-440.

18. Taormina, R., Chau, K. W., & Sivakumar, B. (2015). Neural network river forecasting through baseflow separation and binary-coded swarm optimization. Journal of Hydrology, 529, 1788-1797.

19. Wang, W. C., Xu, D. M., Chau, K. W., & Lei, G. J. (2014). Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. Water resources management, 28(12), 4183-4200.

20. For further information on Stationarity and Differencing see https://www.otexts.org/fpp/8/1

21. Hyndman, R. J., & Athanasopoulos, G. (2015). 8.9 Seasonal ARIMA models. Forecasting: principles and practice. oTexts. Retrieved, 19.