Statistical Analysis of Climate Change Variables

We obtained climate data from Mauna Loa and the National Oceanic and Atmospheric Administration and performed several statistical tests on that data, for the purpose collecting evidence to reject claims that climate change is not occuring. This included CO_2, precipitation, and temperature measurements, and many other long-term climate conditions and short-term weather conditions. Most of our data was quantitative, but we also have categorical data for seasons and decades. We performed a hypothesis test on difference in mean levels of CO_2 in the ‘80s compared to the ‘00s. We also constructed several regression models, performed multiple ANOVA tests, and constructed confidence intervals for CO_2 levels. Most of our testing produced statistically significant results in support of the existence of climate change.

Background and Significance

In recent years the validity of climate change research has been contested by policymakers, scientists, and members of the general public alike. Through this analysis of climate and weather data for the Mauna Loa observatory in Hawai'i, the change in carbon dioxide (CO_2) levels within Earth's atmosphere are related to changes in climate behavior over time. This report details the relationship between variations in CO_2 concentration (ppm) within the atmosphere, and variables such as temperature, precipitation, wind speed, relative humidity, volcanic activity, seasonal change, and solar radiation levels. The main objective of this project is to find statistical evidence to either reject or fail to reject claims that climate variables are not changing over time. CO_2 is known as a greenhouse gas, meaning that it traps heat. By this logic, as CO_2 concentration increases in the atmosphere, so should the temperature. Assessing the statistical validity of this change and changes in other climate variables will contribute depth to the climate change argument.

Methods

In order to combine the datasets from the Mauna Loa observatory and National Centers for Environmental Prediction Climate Forecast System Reanalysis into one dataset, the daily data from the Mauna Loa observatory was averaged into weekly data. The two data sets were combined so that the weeks matched up, and selecting one week would display all of the data that corresponded to it. This larger dataset of weekly averages for each of the variables from 1979 to 2014 served as the population for the rest of the project. Next, a random sample of 100 weeks from the 1826 weeks ranging from January 1979 to July 2014 was created. Months were grouped into four seasons starting with winter from December to February. Once the sample of 100 weeks was formed, sample statistics and graphs for variables were created. Boxplots and scatterplots were generated first. Then a 3D scatterplot was used to demonstrate the "primary variables" in our dataset which were determined to be carbon dioxide levels, temperature, and time. The other variables were investigated to discover if they were confounding or contributing variables to climate change.

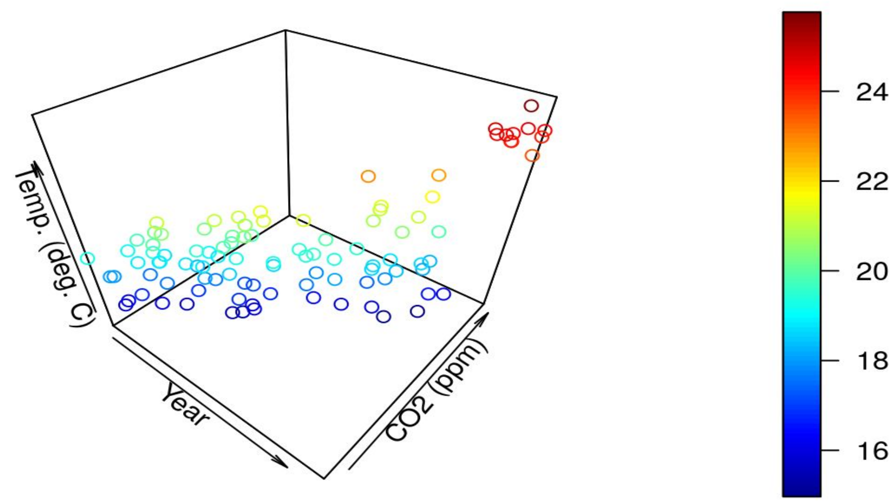
The hypothesis test analyzed the difference between mean CO_2 concentration (ppm) in the 1980s and mean CO_2 concentration in the 2000s. Both the 1980s and 2000s decade groups were taken from our original random sample of 100 cases. Although the 2000's group only had 26 cases, conditions for this test were met because the distributions for both the 1980's and 2000's were normal.

For the confidence interval, two 100 week samples from 1980s and 2000s were created respectively. Then, confidence intervals were calculated for CO_2 emissions at 95% confidence level for these two decades by looking at 2.5th and 97.5th quantiles of the sample.

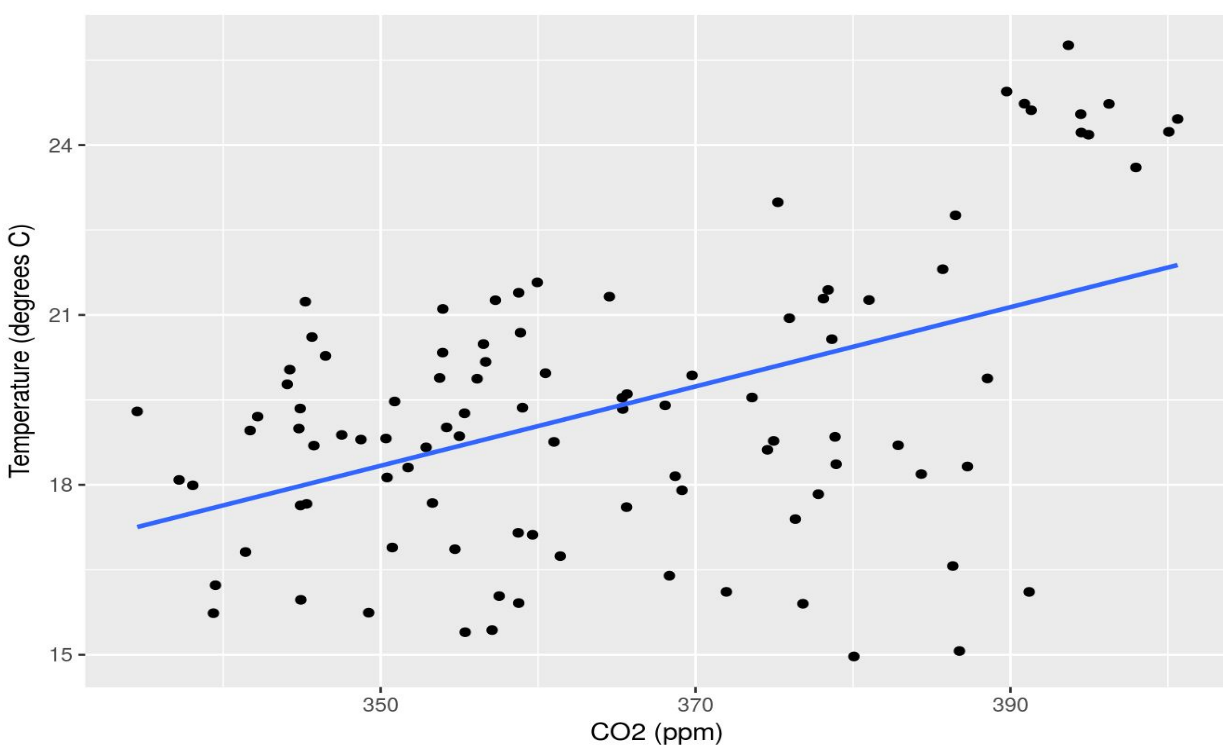
A regression model to predict temperature was developed using CO_2, solar energy, relative humidity, wind, and precipitation as explanatory variables. Conditions for this test were verified by ensuring that the residuals density plot was roughly normal, the Normal Q-Q Plot was linear, and the residuals vs. fitted plot has even variability.

For our Analysis of Variance test, the difference in means for the average CO_2 for each decade was tested, omitting the two data points for 1979 as our data did not go before that year, making a 70s category incredibly small. Conditions were checked to ensure the data was normal distributed and that the variance of each group did not exceed more than twice the value of any other group. Since some of the categories had less than 30 cases, the data was graphed to ensure they were all normal. Since they were, all conditions were met and the tests were performed. This was also done for the difference in weekly maximum temperature for each decade, and weekly maximum temperature for each season. We also tested for differences in mean temperature for the seasons.

Results



The 3D graph models increase in temperature in relation to time and increase in CO_2.



The simple regression model indicates that CO_2 is a predictor of temperature with a p-value of 4.708e-07. The coefficient for CO_2 is 0.07002, indicating that for each increase of 1 ppm of CO_2 in the atmosphere, our predicted temperature is 0.07002 degrees Celsius higher. Using this simple regression model we found that CO_2 concentration was a significant predictor of temperature. The multiple regression model showed that CO_2 was a less significant predictor of temperature than humidity and solar exposure. The model is not the most parsimonious, but indicates which climate variables are most influential over temperature.

The hypothesis test demonstrated that the difference in mean CO_2 concentration (ppm) in the atmosphere for the 1980s and 2000s is significant at the alpha = 0.05 level. Therefore we reject our null hypothesis that there is no change in mean CO_2 levels from 1980 to 2000.

For the confidence interval, two 100 week samples from 1980s and 2000s were created respectively. Then, confidence intervals were calculated for CO_2 emissions at 95% confidence level for these two decades by looking at 2.5th and 97.5th quantiles of the sample. The hypothesis test demonstrated that the difference in mean CO_2 concentration (ppm) in the atmosphere for the 1980s and 2000s is significant at the alpha = 0.05 level. Therefore we reject our null hypothesis that there is no change in mean CO_2 levels from 1980 to 2000.

All ANOVA tests found significant results. We reject the null hypothesis that there has not been a significant change in mean CO_2 levels from decade to decade. We also reject the null hypothesis that there has not been a significant change in mean maximum temperature from decade to decade. Additionally, we reject the null hypothesis that there has not been a change in mean maximum temperature from decade to decade for spring. Finally, we reject the null hypothesis that there has not been a change in mean maximum temperature from season to season. For all of these tests, we have evidence at the 0.05 alpha level that there has been a change of some sort.

Conclusion

The main objective of the study was to see if there is statistical evidence to reject or fail to reject claims that climate variables are not changing over time was achieved. From hypothesis testing, it is possible to ascertain that CO_2 in the atmosphere is changing, and from the regression model it is also significant predictor of temperature. Assuming that temperature is an indicator of climate change, there is evidence to support the claim that CO_2 is connected to this increase. We acknowledge the relationship between increasing CO_2 levels and heightened temperature and the correlation this has with an increase in extreme weather. Although this is not an experimental study with the capability to prove causation, it is clear there is a strong correlation between the variables we have explored in this report. The next step in our analysis would be to analyze a different area of the world that is already feeling the impact of climate change, perhaps the Arctic Circle or the Great Barrier Reef, or try finding out the confounding variable in the jump that is observed in our 3D graph.

References

Global Greenhouse Gas Reference Network. (n.d.). Retrieved from Earth System Research Laboratory Global Monitoring Division website: https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html

Global Weather Data for SWAT. (n.d.). Retrieved from Global Weather Data for SWAT website: https://globalweather.tamu.edu/

Keeling, R. (2008). Recording Earth's Vital Signs. Science Magazine, 319. Retrieved from https://www.esrl.noaa.gov/gmd/co2conference/Reporters/ keeling.pdf

The Significant Volcanic Eruption Database. (n.d.). Retrieved from https://www.ngdc.noaa.gov/nndc/servlet/ShowDatasets?dataset=102557&search_look=50&display_look=50