MLR Analysis: Climate Change

Carly Lundgreen

Problem Background

Climate scientists have long predicted that increasing levels of carbon dioxide, and other greenhouse gases such as methane and nitrous oxide, in the atmosphere would increase average global temperature because of the greenhouse effects. Carbon dioxide, as well as other greenhouse gases, is transparent to visible light–which heats the earth. The warming earth emits thermal radiation back toward space but greenhouse gases, unlike oxygen, absorb soeme of that thermal radiation, which warms the earth and the lower portion of its atmosphere. Over 100 years ago, the Nobel prize winning chemist Svante Arrhenius wrote that 'doubling the percentage of carbon dioxide in the air would raise the temperature of the earth's surface by 4 degrees celsius.'

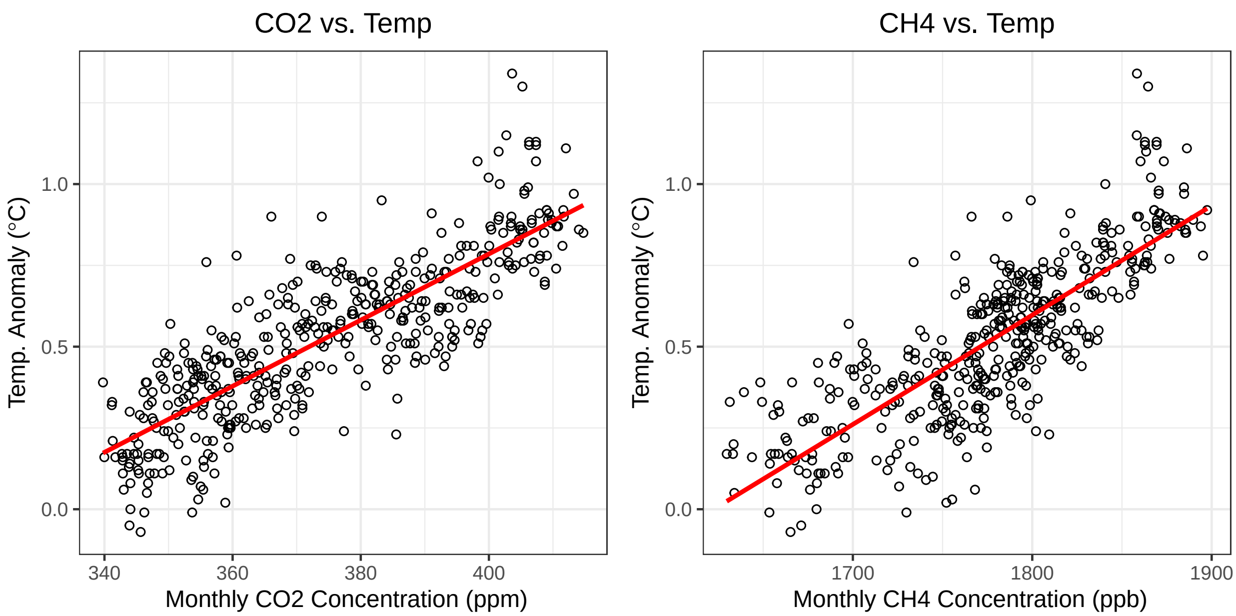
Data

These data include temperature information gathered from the Global Land-Ocean Temperature Index from the Goddard Institute of Space Studies (GISTEMP). The temperature (`temp`) is reported in units of 1/100 degree celsius increase above the 1950-1980 average. We have converted this to degrees celsius in our dataset. This measurement is often referred to as the global surface tempearture anomaly. The CO2 and CH4 (carbon dioxide and methane) measurements were retrieved from The Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (NOAA). These data are a record of monthly mean atmospheric CO2 and CH4 concentration at Mauna Loa Observatory in Hawaii. This location was selected because it is relatively unaffected by any local emissions, and as such is more representative of the global concentration of well-mixed gases. It should also be mentioned that these observations were started by C. David Keeling in March of 1958 at a NOAA facility and are often referred to as the Keeling Curve. The monthly atmospheric CO2 concentration (`co2`) is expressed in units of micromol per mole and abbreviated as ppm. The CO2 measurements began in 1969, and the CH4 measurements began in 1983. The monthly atmospheric CH4 concentration (`methane`) is reported in nanomoles per mole and is abbreviated as ppb.

Exploratory Data Analysis

The correlation coefficient of temp and co2 is found to be 0.829, which is reflective of a strong, positive, linear relationship between co2 and our response variable (temp). This strong relationship is confirmed by the scatterplot shown below. There is a distinct upward, linear trend seen in the plot–indicative of a correlation coefficient equal to 0.829.

The correlation between temp and methane shows a similar pattern. With a correlation coefficient of 0.807, we observe another strong, positive, linear relationship between monthly levels of an atomospheric gas (CH4) and global temperature anomaly. Again, we confirm this with a scatterplot, shown below. There is a distinctly positive trend seen in this plot, and we can see that, on average, as CH4 increases, so does the global temperature anomaly.



Analysis

We will use following multiple linear regression model for the response variable, the global surface temperature anomaly (temp), and our explanatory variables (monthly carbon dioxide and methane concentrations in ppm and ppb, respectively):

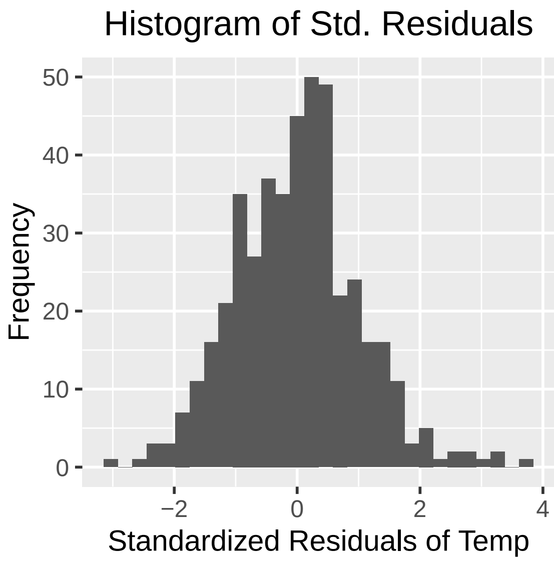
$$temp = \beta_0 + \beta_1 CO2 + \beta_2 methane + \epsilon,\ \ \epsilon \sim\ \ N(0, \sigma^2)$$

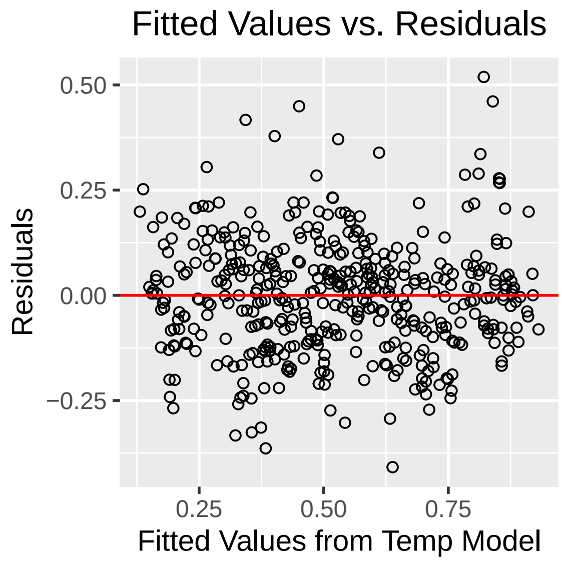
Assumptions of Model

In fitting the above model, a few assumptions are made. First, we assume a linear relationship between our response and explanatory variables, confirmed in the Exploratory Data Analysis section above. Secondly, we assume independence between temperature measurements. This assumption may be violated, as it is likely that global surface temperature anomaly measurements are correlated with previous monthly measurements. However, we will move forward with this analysis for the moment and perhaps revisit the potential correlation at a later time. The final two assumptions made with this analysis are that the residuals are normally distributed and that they vary equally about the regression line.

After fitting the above model, we checked the final two assumptions. The plots below show that these assumptions are met–the histogram of standardized residuals appears normal and there are no concerning issues with unequal variance in a scatterplot of fitted values vs. residuals.

	Estimate	Std. Error
(Intercept)	-4.036	0.264
co2	0.007	0.001
methane	0.001	0.000





Model Fit

A table of β estimates and associated standard errors is below.

The model predicts that on average, for every 1-unit increase in CO2 levels, global surface temperature anomaly will increase by 0.007degrees celsius assuming all else is held constant.

Similarly, the model predicts that, on average, for every 1-unit increase in methane levels, temp will increase by 0.001 degrees celsius assuming all else is held constant.

Inference

We also want to analyze whether there is a statistically significant greenhouse effect (i.e. whether increasing CO2 or methane levels leads to an increase in global surface temperature. We will utilize an ANOVA test, and the hypotheses are as follows:

$$H_0 : \beta_1 \ = \ \beta_2 = \ 0$$
$$H_A : \textit{At least one } \beta_i, \ \neq \ 0, \ (i = 1, 2)$$

The F statistic was found to be 505.6 and the p-value was essentially 0. Because this p-value is much smaller than our significance level of 0.05, we reject the null hypothesis that neither gas has an effect on global surface temperature. There is significant evidence of an greenhouse effect.

In other words, the effect of CO2 and CH4 levels on the temperature anomaly was too strong, essentially, to have happened by chance, which is why we can say that there is statistically significant evidence of a greenhouse effect.

Because we concluded that there is a statistically significant greenhouse effect, we want to analyze whether the effect is due to CO2, CH4, or both CO2 and CH4. To test this, we calculate confidence intervals for β_1 (effect of CO2) and β_2 (effect of methane).

	Lower Bound	Upper Bound
CO2	0.0056	0.0092
CH4	0.0004	0.0016

Min	Mean	Median	SD	Max
0.001	0.063	0.049	0.065	0.2

The 95% confidence interval for β_1, the effect of CO2, is found to be (0.0056,0.0092) degrees Celsius. On average, for a 1 ppm increase in monthly CO2 concentration, the global surface temperature anomaly increases between 0.0056 and 0.0092 degrees Celsius, assuming CH4 concentration stays constant, with 95% confidence. This means that (0.0056,0.0092) is a likely range of values for the true effect of CO2 on surface temperature, and because the entire interval is greater than zero, we know that CO2 has a significant effect on surface temperature.

A similar result is found for the effect of methane on surface temperature. The 95% confidence interval for β_2 is found to be (0.0004,0.0016) degrees Celsius. For a 1 ppb increase in monthly methane concentration in the atmosphere, the global surface temperature anomaly increases, on average, between 0.0004 and 0.0016 degrees Celsius, assuming CO2 levels are held constant. Again, because this likely range of values for the true effect of methane on surface temperature is entirely greater than zero, we know that methane also has a significant effect on global surface temperature.

Because of these intervals, we conclude that both CO2 and CH4 contribute to the significant greenhouse gas effect observed in the atmosphere.

Prediction Accuracy

In order to assess the predictive accuracy of the model, we decided to perform a cross-validation of the model using a test set that included the data from only 2019, and a training set (that we used to fit a new model estimated from the climate data from all other years in the dataset besides 2019). We predicted the `temp` for the test set and calculated the absolute prediction error for each of our predicted values. There were 7 temperature measurements from 2019 due to a few missing values in the dataset, so we have 7 predictions and thus 7 measurements of prediction error to summarize. The summary is included in the table below. Note that we calculated absolute prediction error by finding the absolute value of the difference between the `temp` values from our test set (2019 data) and the predicted values calculated by fitting the training model to that test set.

Note that the average absolute prediction error was 0.063 with a standard deviation of 0.065. The range of values for our temperature anomaly measurements from 2019 was found to be 0.26, which is about 4.15 times higher than our average prediction error. This indicates that our predictions were not incredibly far off from the actual temperature measurements, relative to the range of actual observations, so model prediction for the coming year will likely be quite accurate.

Conclusion

While we have not yet doubled the amount of CO2 in the atmosphere since 1983, which is the first year in the data gathered from NOAA, there has has been an increase in CO2 levels from about 340 ppm to about 415 ppm in 2019. According to our model, if we hold methane values constant, we predict an average increase in surface temperature anomaly of 0.007 degrees Celsius for every one ppm increase in C02 concentration. Given that CO2 levels have risen by almost 50 ppm since only 1983, that average surface temperature increase will only keep adding up until we inevitably reach the prediction made by Svante Arrhenius over a century ago. All in all, there is a significant greenhouse gas effect on the warming of the earth's surface, and rising CO2 and CH4 levels in the atmosphere both play a part in that global warming.