# The Greenhouse Gas Effect: Diagnostics

Mitchell Pudil

February 3, 2019

## 1 Introduction

Is there evidence that increasing man-caused greenhouse gases results in higher global average temperatures?

The Greenhouse gas effect states that this may just be the case. In fact, scientists have predicted for years that increasing levels of carbon dioxide or methane would increase the global temperature. Over a hundred years ago, one scientist, Svante Arrhenius, wrote that 'doubling of the percentage of carbon dioxide in the air would raise the temperature of the earth's surface by 4°C.' Today, we will continue our analysis to find out if this is the case, and if so, approximate to what extent these gases raise the global temperature.

## 2 Data

Data for this project comes from various sources. Temperaure data comes from the Global Institute of Space Studies (GISTEMP), and is reported in 1/100 of a degree centigrade incrase above the 1950-1980 mean. The $CO_2$ and $CH_4$ data come from The Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (NOAA). These data are a record of monthly mean atmospheric $CO_2$ and $CH_4$ concentration at Mauna Loa Observatory, Hawaii. This remote island is a great place for our data to be gathered because it isn't affected much by local emissions. Additionally, we will be adding data of Monthly Atmospheric Hydrochloroflurocarbon (HCFC) Concentration in ppt, and Monthly Atmospheric Sulfur Hexaflouride ($SF_6$) Concentration in ppt.

## 3 EDA

We begin our analysis by determining the correlation between temperature and each of the gases. The following table shows this correlation, with the variable that is being correlated with temperature on the left, and the value of the correlation, $\rho$, on the right:

| Gas | Correlation with Temperature |
|---|---|
| Carbon Dioxide ($CO_2$) | 0.8415 |
| Methane ($CH_4$) | 0.7752 |
| Nitrogen Oxide ($N_2O$) | 0.6616 |
| HCFC | 0.5652 |
| Sulfur Hexaflouride ($SF_6$) | 0.6556 |

## 4 Analysis

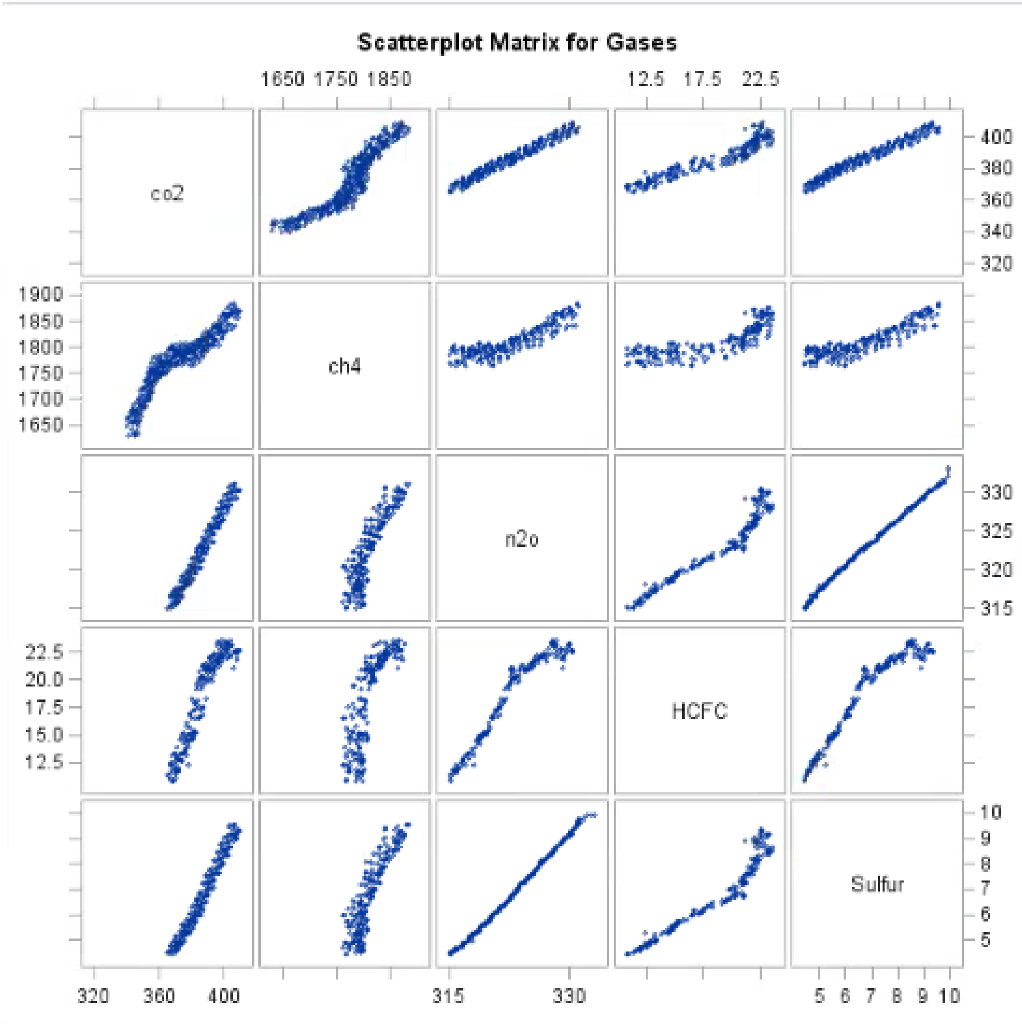
The response variable for this experiment is the average global temperature for a specific month and year, reported in °C. The explanatory variables are the corresponding carbon dioxide, in micromoles per mole (ppm), methane levels in nanomoles per mole (ppb), nitrogen oxide in ppb, HCFC in parts per thousand (ppt), and sulfur hexaflouride in ppt. Our model is:

$$\widehat{Temp} = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{CO}_2 + \hat{\beta}_2 \mathrm{methane} + \hat{\beta}_3 \mathrm{N}_2\mathrm{O} + \hat{\beta}_4 \mathrm{HCFC} + \hat{\beta}_5 \mathrm{SF}_6 + \epsilon, \epsilon \sim N(0, \sigma^2)$$

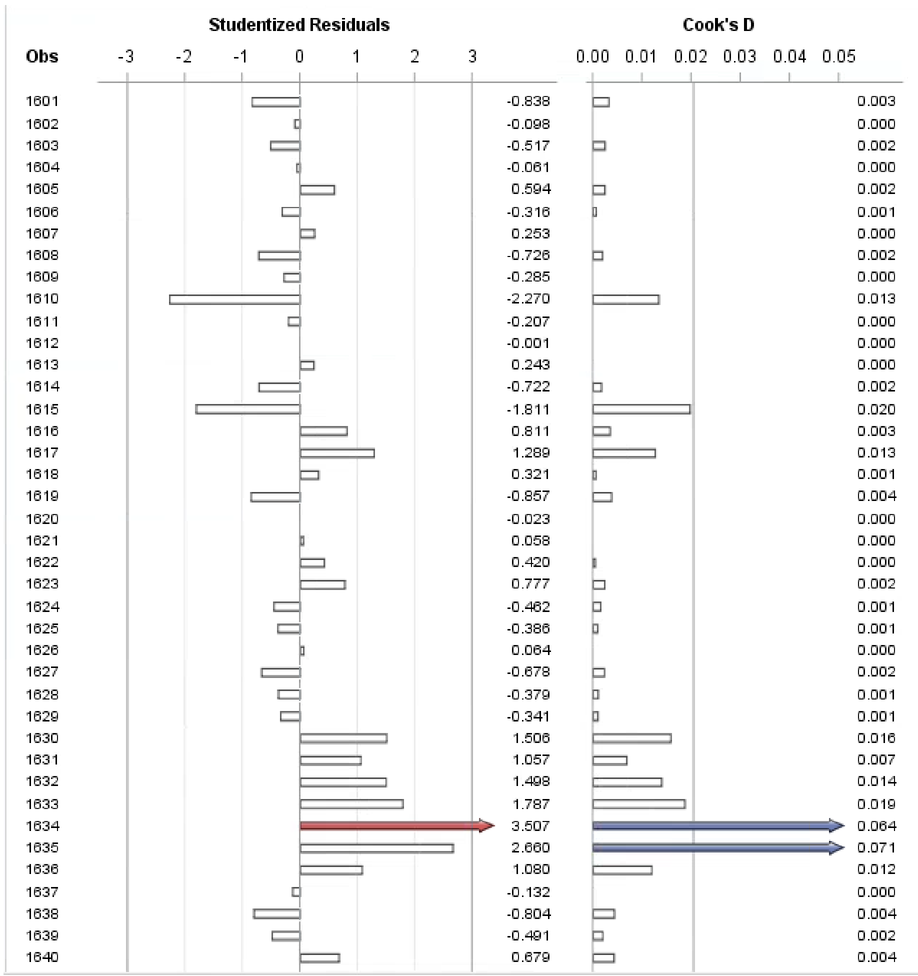
After running the regression, we now have determined the estimate and standard errors, which are shown below, with standard errors in parentheses underneath the estimate

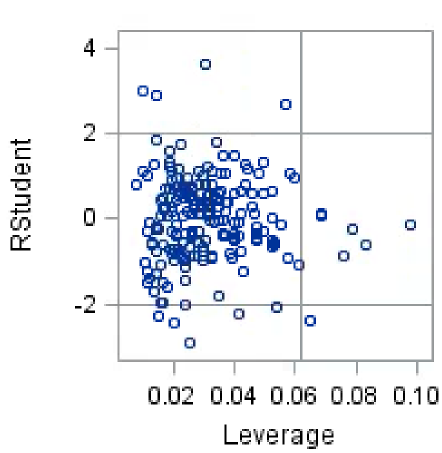
$$\widehat{Temp} = \underset{(1409.587)}{-5241.699} + \underset{(0.382)}{-0.254}\mathrm{CO}_2 + \underset{(0.072)}{0.178}\mathrm{methane} + \underset{(4.652)}{16.680}\mathrm{N}_2\mathrm{O} + \underset{(0.950)}{-5.350}\mathrm{HCFC} + \underset{(13.407)}{-29.870}\mathrm{SF}_6$$

It appears that there are a few compounds with regression coefficient signs that are different from the expected signs from the correlation table. Specifically, $CO_2$, HCFC, and $SF_6$ have negative signs. Additionally, we see a lack of significance of expected effects (especially with $CO_2$ and methane). Possible reasons for this occurence include: misspecified model, omitted variables, influential observations, and multicollinearity. A graph of the collinearity present in the data is shown below:




It is clear to see that there are distinct linear relationships between many of the gases, which produces problems in our estimates if we want to determine the effect of a particular gas on the average global temperature. We also have a handful of influential observations and outliers. The first plot below is a sample of the observation (month) with the corresponding studentized residuals and Cook's Distance (a measure of leverage). The second plot shows a scatterplot of the leverage and studentized residuals for all data points.




Out of all of the gases, it seems that $N_2O$ and sulfur are the two most highly correlated gases, with a VIF of 517 and 466, respectively. Because of this incredibly high level of correlation, it is best to leave out sulfur to improve the estimates of the other gases. Unfortunately, doing so will mean that our final model will not be able to include sulfur, so we will not be able to determine a more precise estimate for that gas. The new model, then, is:

$$\widehat{Temp} = \underset{(317.521)}{-2188.681} + \underset{(0.386)}{-0.285}CO_2 + \underset{(0.0703)}{0.112}\text{ methane} + \underset{(1.487)}{6.978}N_2O + \underset{(0.945)}{-4.908}HCFC$$

# 5 Conclusion

While our first analysis showed a statistically significant effect of both $CO_2$ and methane, correcting for multicollienarity and adding in additional data led us to understand differently. Also, while there was a positive correlation between HCFC and temperature, the regression coefficient for this gas was both negative and statistically significant. Finally, $N_2O$ appeared to have a positive effect on temperature, both in terms of correlation and the regression coefficient.

The data features that match the strengths of the analysis are that the model incorporates multiple explanatory variables, and that it allows us to test each one independently of the other. However, the model failed to incorporate nonlinear relationships or interaction effects. The prediction was also not very good because of this and the endogeneity problems.

A research task with the same characteristics as this data is determining housing prices based off of the square footage, region, number of bathrooms and bedrooms, etc. The data for this can be found at https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data.