# CLIMATE CHANGE

There have been many studies documenting that the average global temperature has been increasing over the last century. The consequences of a continued rise in global temperature will be dire. Rising sea levels and an increased frequency of extreme weather events will affect billions of people.

In this problem, we will attempt to study the relationship between average global temperature and several other factors.

The file climate_change.csv contains climate data from May 1983 to December 2008. The available variables include:

- *Year*: the observation year.
- *Month*: the observation month.
- *Temp*: the difference in degrees Celsius between the average global temperature in that period and a reference value. This data comes from the Climatic Research Unit at the University of East Anglia.
- *CO2*, *N2O*, *CH4*, *CFC.11*, *CFC.12*: atmospheric concentrations of carbon dioxide ($CO_2$), nitrous oxide ($N_2O$), methane ($CH_4$), trichlorofluoromethane ($CCl_3F$; commonly referred to as CFC-11) and dichlorodifluoromethane ($CCl_2F_2$; commonly referred to as CFC-12), respectively. This data comes from the ESRL/NOAA Global Monitoring Division.
  - CO2, N2O and CH4 are expressed in ppmv (parts per million by volume -- i.e., 397 ppmv of CO2 means that CO2 constitutes 397 millionths of the total volume of the atmosphere)
  - CFC.11 and CFC.12 are expressed in ppbv (parts per billion by volume).
- *Aerosols*: the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space. This data is from the Godard

Institute for Space Studies at NASA.

- *TSI*: the total solar irradiance (TSI) in W/m$^2$ (the rate at which the sun's energy is deposited per unit area). Due to sunspots and other solar phenomena, the amount of energy that is given off by the sun varies substantially with time. This data is from the SOLARIS-HEPPA project website.
- *MEI*: multivariate El Nino Southern Oscillation index (MEI), a measure of the strength of the El Nino/La Nina-Southern Oscillation (a weather effect in the Pacific Ocean that affects global temperatures). This data comes from the ESRL/NOAA Physical Sciences Division.

## Problem 1.1 - Creating Our First Model

(2/2 points)

We are interested in how changes in these variables affect future temperatures, as well as how well these variables explain temperature changes so far. To do this, first read the dataset climate_change.csv into R.

Then, split the data into a *training set*, consisting of all the observations up to and including 2006, and a *testing set* consisting of the remaining years (hint: use subset). A training set refers to the data that will be used to build the model (this is the data we give to the lm() function), and a testing set refers to the data we will use to test our predictive ability.

Next, build a linear regression model to predict the dependent variable Temp, using MEI, CO2, CH4, N2O, CFC.11, CFC.12, TSI, and Aerosols as independent variables (*Year* and *Month* should NOT be used in the model). Use the training set to build the model.

Enter the model R$^2$ (the "Multiple R-squared" value):

0.7509 ✔

*You have used 1 of 5 submissions*

## Problem 1.2 - Creating Our First Model

(1/1 point)

Which variables are significant in the model? We will consider a variable signficant only if the p-value is below 0.05. (Select all that apply.)

- [x] MEI
- [x] CO2
- [ ] CH4
- [ ] N2O
- [x] CFC.11
- [x] CFC.12
- [x] TSI
- [x] Aerosols

✔

*You have used 1 of 2 submissions*

## Problem 3 - Simplifying the Model

(2/2 points)

Given that the correlations are so high, let us focus on the N2O variable and build a model with only MEI, TSI, Aerosols and N2O as independent variables. Remember to use the training set to build the model.

Enter the coefficient of N2O in this reduced model:

0.02532 ✔

(How does this compare to the coefficient in the previous model with all of the variables?)

Enter the model $R^2$:

0.7261 ✔

*You have used 1 of 5 submissions*

Please remember not to ask for or post complete answers to homework questions in this discussion forum.