Experimental Methods -Statistical Data Analysis - Assignment

In this assignment you will investigate daily weather observations from Alice Springs, Woomera and Charleville, from 1950 to 2006. These are towns in the Australian outback which all have a very dry climate. The final objective of the project is to determine the effect of the radiative forcing from the additional CO_2 in the atmosphere between 1950 and 2006 on the difference between the maximum and minimum temperatures on clear, dry nights.

The raw data you will analyse can be found on the baker machine in my directory.

/home/msevior/ExpMethods/ClimateData

The data is stored as plain ascii files in comma separated variable format (csv). This is "event mode data". Each line is a separate entry for every day of the year between 1950 and 2006.

Every day has room for entries for:

stn_num, Date, Maximum Temperature ,Minimum Temperature ,Terrestrial Minimum, Max Wind Gust (km/h),Max Gust Direction,Evaporation (mm),Bright Sunshine (hours),Precipitation (mm),Temperature at 9am ,Relative Humidty at 9am (%), Cloud Cover at 9am (oktas),Wind Speed at 9am (km/h),Wind Direction at 9am,Temperature at 3pm ,Relative Humidty at 3pm (%),Cloud Cover at 3pm (oktas),Wind Speed at 3pm (km/h),Wind Direction at 3pm

Some data values are missing for some days and you should not assume every single day from 1950 to 2006 has an entry.

You are free to perform the data analysis with any set of computer programs you like. You can use matlab, root or IDL. I can provide support for projects written in matlab or root.

The final report should contain all the plots you are asked to make in one self contained document. I'm perfectly happy to accept a report in a pdf format document.

This is a large project. I encourage you to investigate whatever scripting and automation tools are available to automate the production of the many quantities listed here. Learning these techniques will stand you in good stead in your future career in Physics.

Part 1 - Getting familiar with the data.

1. Decide what analysis tool you wish to use. The analysis tool - "root" was explicitly designed to analyse this kind of event mode data and if you are doing a research project that employs root I suggest you employ that package for this assignment.

I have written two C++ programs to import the csv data into a root TTree, readCSV.cpp and ReadAlice.cpp. These will help you get started.

There is very extensive documentation available for root. The Webpage:
http://root.cern.ch/root/doc/RootDoc.html

Lists the supported documentation.

For matlab simply use the supplied ezdata.m to read the into a matlab data structure. In addition I suggest you edit the original .csv data files to remove all the entries prior to the first entry that contains valid information for the humidity and cloudiness.

I have provided a collection of matlab .m files that read in matlab data and create a useful data structure for later analysis. There also examples of graphics and histogram related calls. These are in the subdirectory: ClimateData/matlab. Feel free to use this code as you like in your project.

Make sure you put as much as possible into matlab scripts so you can rerun your analysis later and also so that you have a set of useful templates for other data analysis jobs. For example you may find the code you write for this project is helpful and can be reused in your Masters project.

The first job will be to read the data into your analysis system, and record the date, min temp, max temp, % cloudiness at 9 AM and 3 PM and relative humidity at 9AM and 3 PM.

You will need to write a routine to calculate the number of hours of night on each date.

To start with, find the first year for which all the observations listed above are available at the 3 locations, (Alice Springs, Woomera and Charleville). Then for this year only do the following for the 3 locations.

For a collection of N data points:

the mean is: $\mu = \frac{\sum_i x_i}{N}$

The Standard Deviation is: $\sigma = \sqrt{\frac{\sum_i (x_i - \mu)^2}{N}}$

1. Find the mean and standard deviation for:
   (a) the minimum temperate,
   (b) the maximum temperature
   (c) the difference between the minimum and maximum temperature for the same day (dT1).
   (d) the difference between the days' maximum temperature and the minimum temperature for the next day (dT2).
   (e) the average of % cloudiness at 3PM and 9 AM of the next day, (C)
   (f) the average of relative humidity at 3PM and 9 AM of the next day, (H)

2. Make and plot histograms for each of the quantities listed in 1.

For a collection of N unbinned data points (x_i,y_i), with respective means (μ_x,μ_y) and standard deviations (σ_x,σ_y) the correlation coefficient r_{xy} is given by:

$$r_{\mathrm{xy}} = \frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{(N-1)\sigma_x \sigma_y}$$

3. Now calculate the correlation coefficient between:

(a) The minimum and maximum temperature of the same day.
(b) The maximum temperature and the minimum temperature the next day.
(c) The temperature difference, dT1, and the number of hours of darkness
(d) The temperature difference, dT2, and the number of hours of darkness
(e) The temperature difference, dT2, and the average of % cloudiness at 3PM and 9AM of the next day, (C)
(f) The temperature difference, dT2, and the average of relative humidity at 3PM and 9AM of the next day, (H)
(e) H vs C.

4. Make and plot scatter plots of each of the quantities listed in 3.

Part 2 - Employing MC methods

Construct Gaussian pdf's for the minimum and maximum temperature with the same expectation values and standard deviations you found in part 1.

5. Using the accept/reject method run a MC experiment for one years worth of data for the maximum and minimum temperatures. Calculate the expectation value for the years worth of data.

6. Now run 1000 MC experiments and calculate the standard deviation of the Expectation value for the maximum and minimum temperatures and the difference between the maximum and minimum temperatures on consecutive day/nights (dT2)over the ensemble of the 1000 experiments.

7. Do the same 1000 MC experiments for the 5 year Expectation value for the maximum and minimum temperatures and the difference between the maximum and minimum temperatures on consecutive day/nights (dT2)over the ensemble of the 1000 experiments.

What you have now is an estimate of the uncertainty in the 1 year and 5 year Expectation values for the maximum and minimum temperatures and dT2.

8. Now determine the expectation values of the maximum, minimum temperatures and dT2 for Alice Springs, Charleville and Woomera from 1950 to 2006. Do this for 1 year and 5 year averages. Is there a statistically significant trend in the data?

Part 3 - Investigating the rate of atmospheric cooling in the outback

The following is a plausible simple model that describes how locations in the Australian desert might cool at night. In principle it is sensitive to the effects that drive Global Warming. Your job is to determine the parameters of the model and to look for a statistically significant time variation in the parameters.

A Black body radiates heat with Power given by:

$$P = \sigma T^4$$

P = Power in watts per m^2
T = Temperature in Kelvin
σ = Stefan-Boltzman constant = $5.67 \times 10^{-8}\ Wm^{-2}T^{-4}$

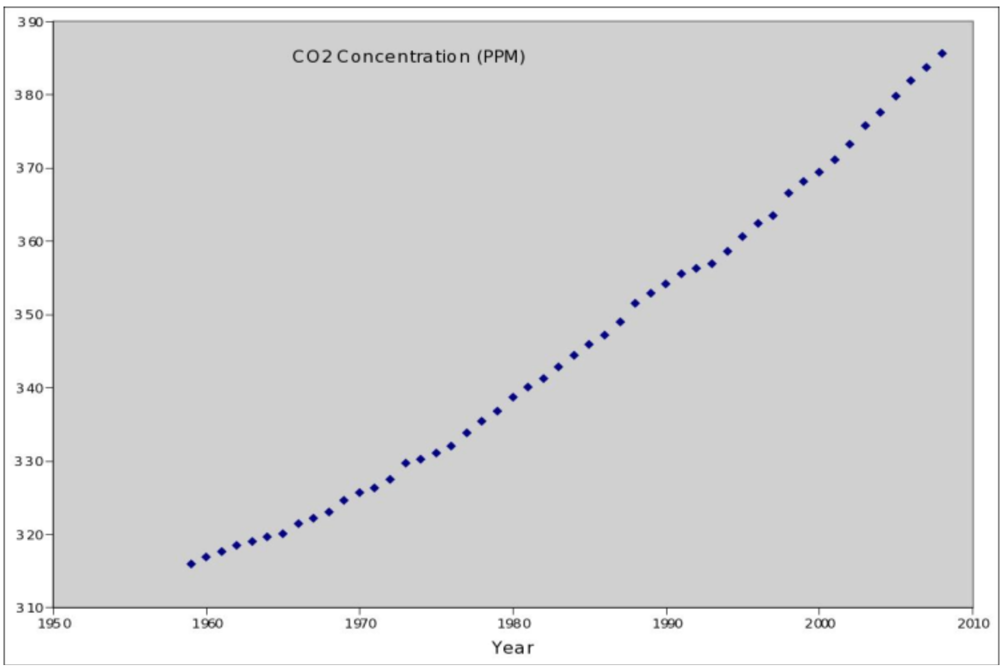
Plugging in T =300 K for air temperature gives P = 459 W/m^2.

As a rule of Thumb, the Radiant energy from the Sun is approximately 1000 W/m^2, but this is only present for 12 hours per day as opposed to the backbody radiation which is present 24 hours a day. Given that the Blackbody radiation is half that of the full intensity of the Sun, we can see that the Earth at 300 K is in approximate radiative balance with the heat input from the sun. My model employs radiative cooling as the first order effect that provides the night time decrease in temperature.

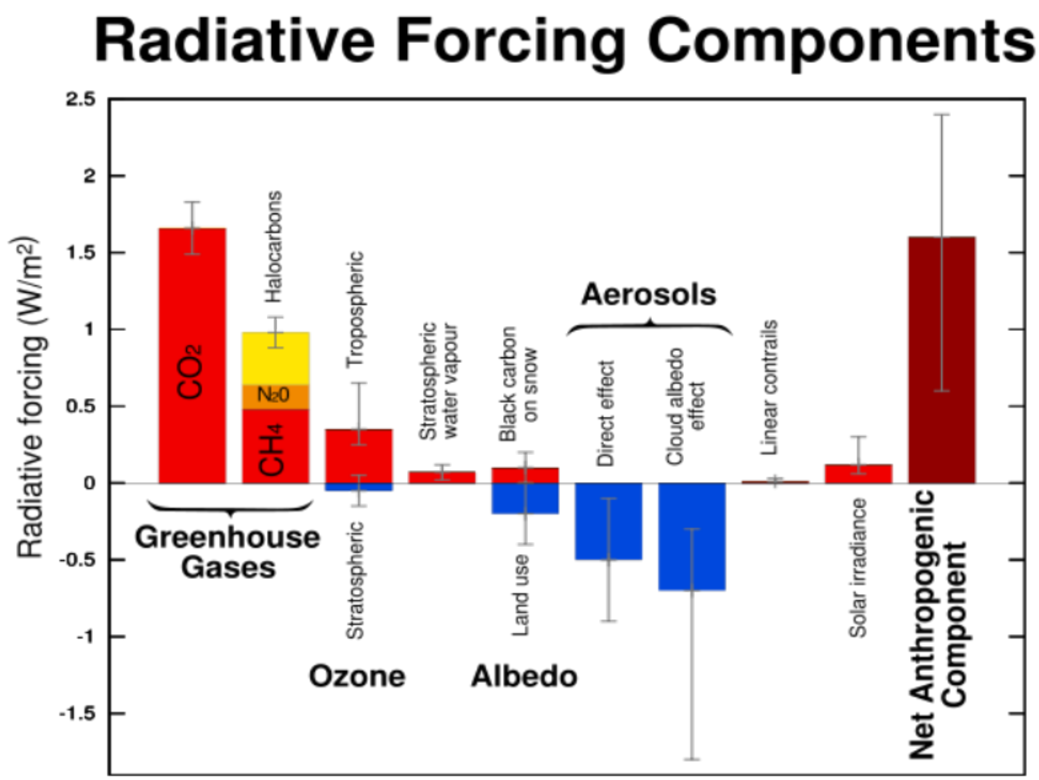
An interesting extension of the project is to somehow allow other modes of cooling in the model.

The Atmospheric CO2 concentration has been very well measured since 1959. The results are summerized at http://www.esrl.noaa.gov/gmd/ccgg/trends/.

An Excel spreadsheet with the yearly mean CO2 concentration is available at:
/home/msevior/ExpMethods/CO2_Conc.xls



Th effect of the increase in CO_2 has been predicted to be an increase radiative heating or “forcing" of energy back into the Atmosphere. The net effect of radiative forcing since the industrial revolution has been predicted to be 1.6 Wm^{-2}. See http://en.wikipedia.org/wiki/Radiative_forcing



At the time of the Industrial Revolution the CO_2 concentration was 280 PPM, it is now 383 PPM.

9. Assume that there is a linear relationship between CO_2 concentration and Radiative forcing ΔF and the CO_2 concentration ρ: ie

$\Delta F = F_0 + F_1 \times \rho$

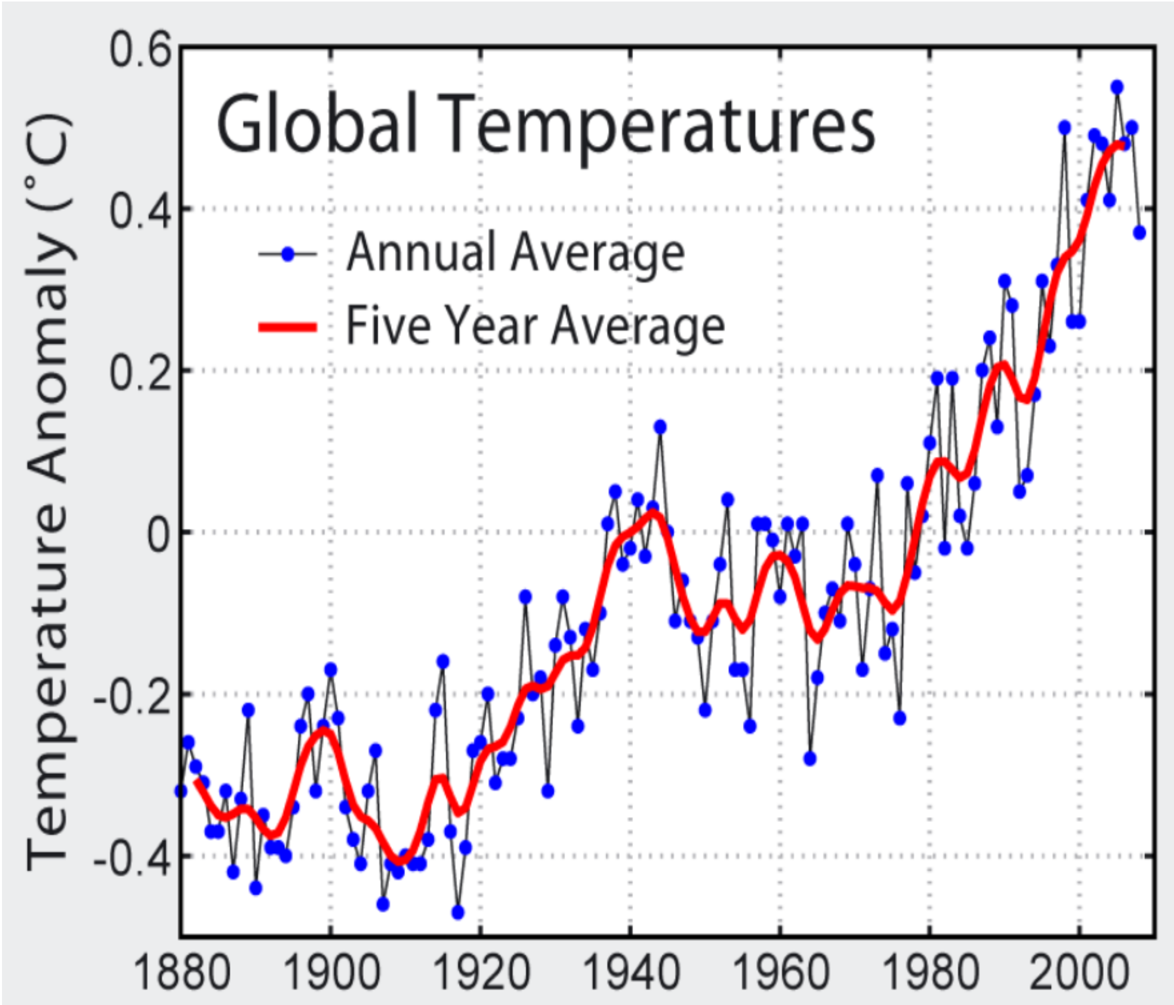
Use the data above to determine F_0 and F_1.

Using the BlackBody Radiation formula above, $P = \sigma T^4$, the temperature rise required to radiate this additional 1.6 Watts/m^2 can be estimated as follows:

$$T^4 \sigma = P$$
$$(T + \Delta T)^4 \sigma = P + \Delta P$$
$$\Rightarrow \frac{(T + \Delta T)^4 \sigma}{T^4 \sigma} = \frac{P + \Delta P}{P}$$
$$\Rightarrow (1 + \frac{\Delta T}{T})^4 = 1 + \frac{\Delta P}{P}$$
$$\Rightarrow 1 + 4\frac{\Delta T}{T} \approx 1 + \frac{\Delta P}{P}$$
$$\Rightarrow \frac{\Delta T}{T} \approx \frac{1}{4}\frac{\Delta P}{P}$$
$$\Rightarrow \Delta T \approx \frac{1}{4} T \frac{\Delta P}{P}$$

So plugging in T =300 and P = 477, ΔP = 1.6, gives $\Delta T = 0.25$ Degrees

Yet measurements of Global Temperature since pre-industrial times show ΔT = 0.8 Degrees.



http://en.wikipedia.org/wiki/Global_warming

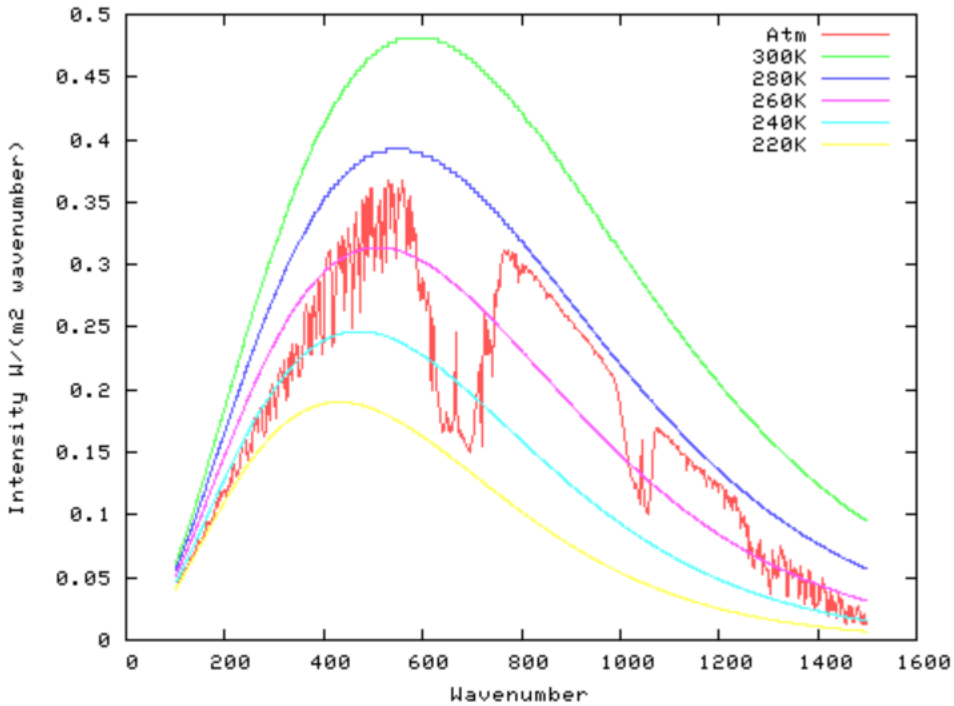
This additional temperature rise is thought to be caused by feedback mechanisms which increase the effective radiative forcing, ΔF.

You can calculate the effects of various types of atmospheric absorption at the website:

http://geodoc.uchicago.edu/Projects/modtran.html

The website employs a model to determine the effects of atmospheric absorption of infrared radiation due to water vapour, clouds, CO_2, methane, ozone and other trace gasses all as a function of their distribution through the atmosphere.

Some sample output with CO_2 at 384 ppm is shown below. Note how the many atmospheric effects significantly alter the total power emitted!



We will attempt to detect the presence of Radiative Forcing ΔF, to see if has changed with time since 1950, when the Global CO_2 concentration was ~310 PPM and measure its size.

To do this we will look at the Day-Night temperature difference in the outback and employ radiative cooling, while allowing for atmospheric absorption, to match the outward heat flow with the Heat capacity of the atmosphere.

We will assume that Stefan's Constant is modified by the relative humidity and the degree of cloudiness. We will also assume that the Heat Capacity of the changes as a function of the humidity and Cloudiness.

So the basic equation that governs how the atmosphere cools at night is:

$$Q = C\Delta T$$

Where Q is the energy that is radiated into space, C is the Heat Capacity of air and ΔT is difference between the minimum and maximum temperature.

Then Q, is given by:

$$Q = \int_0^N -\sigma' T(t)^4 \mathrm{dt}$$

Where σ' is effective radiation "constant". We will allow σ' to vary with humidity and cloudiness.
T(t) is the temperature in Kelvin as a function of time.
N is the number of seconds in the night.

We can approximate T(t) with:

$$T(t) = T_{\max} - \frac{(T_{\max} - T_{\min})t}{N}$$

Where:
$T_{\max}$ is the Maximum Temperature.
$T_{\min}$ is the minimum temperature of the next day.
N = number of seconds in the night.

Taking the constant out of the integrand, our Heat equation becomes:

$$Q = -\sigma' \int_0^N T(t)^4 \mathrm{dt}$$

Putting in the time dependence of the temperature

$$\Rightarrow Q = -\sigma' T_{\max}^4 \int_0^N \left(1 - \frac{(T_{\max} - T_{\min})\, t}{N T_{\max}}\right)^4 \mathrm{dt}$$

Since the temperature change is small compared to the total absolute temperature..

$$\Rightarrow Q \approx -\sigma' T_{\max}^4 (N - \int_0^N 4 \frac{(T_{\max} - T_{\min})\, t}{N T_{\max}} \mathrm{dt})$$
$$\Rightarrow Q \approx -\sigma' T_{\max}^4 (N - 2 \frac{(T_{\max} - T_{\min})\, N^2}{N T_{\max}})$$
$$\Rightarrow Q \approx -\sigma' \mathrm{N} \mathrm{T}_{\max}^4 \left(1 - 2 \frac{T_{\max} - T_{\min}}{T_{\max}}\right)$$

So this is the Heat radiated into space. We take account of atmospheric absorption with

$$\sigma' = \sigma_0 + \sigma_1 H + \sigma_2 S$$

Where:
H is the relative Humidity as measured at the station
S is the fractional cloudiness as measured in the station
σ_0, σ_1, σ_2 are parameters of the model which we will obtain from fits to the data
σ_0 Is sensitive to the Radiative Heating effect of Global warming. We’re particularly interested to see if there is a statistically significant time variation in this parameter between the earliest and latest measurements.

The Heat flux out, Q is balanced by the decrease in temperature times the Heat Capacity. We know the Heat Capacity, C, varies with Humidity, (H) and Cloudiness, (C) so we let:

$$C = C_0 + C_1 H + C_2 S$$

Where:
C_0, C_1, C_2 are parameters of the Model.

Finally we the first Law of Thermodynamics which requires:

$C(T_{\max} - T_{\min}) - Q = 0$

So our model is that for each measurement:

$$(C_0 + C_1 H + C_2 S)(T_{\max} - T_{\min}) - \sigma' \mathrm{N} \mathrm{T}_{\max}^4 \left(1 - 2\frac{T_{\max} - T_{\min}}{T_{\max}}\right) = 0$$

$$\Rightarrow (C_0 + C_1 H + C_2 S)(T_{\max} - T_{\min}) - N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 \left(1 - 2\frac{T_{\max} - T_{\min}}{T_{\max}}\right) = 0$$
$$\Rightarrow (C_0 + C_1 H + C_2 S) T_{\max} - (C_0 + C_1 H + C_2 S) T_{\min} - N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 + 2N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^3 (T_{\max} - T_{\min}) = 0$$
$$\Rightarrow (C_0 + C_1 H + C_2 S) T_{\max} - (C_0 + C_1 H + C_2 S) T_{\min} - N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 + 2N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 - 2N(\sigma_0 + \sigma_1 H + \sigma_2 S) T_{\max}^3 T_{\min} = 0$$

$$\Rightarrow (C_0 + C_1 H + C_2 S) T_{\max} - (C_0 + C_1 H + C_2 S) T_{\min} + N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 - 2N(\sigma_0 + \sigma_1 H + \sigma_2 S) T_{\max}^3 T_{\min} = 0$$

$$\begin{aligned} &\Rightarrow (C_0 + C_1 H + C_2 S) T_{\max} + N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4 \\ &= (C_0 + C_1 H + C_2 S) T_{\min} + 2N(\sigma_0 + \sigma_1 H + \sigma_2 S) T_{\max}^3 T_{\min} \end{aligned}$$

$$\Rightarrow T_{\min} = \frac{(C_0 + C_1 H + C_2 S) T_{\max} + N(\sigma_0 + \sigma_1 H + \sigma_2 S)\, T_{\max}^4}{(C_0 + C_1 H + C_2 S) + 2N(\sigma_0 + \sigma_1 H + \sigma_2 S) T_{\max}^3}$$

Your job is then to fit the parameters C_0, C_1, C_2, σ_0, σ_1, σ_2 to the data (C, H, $T_{\max}$, $T_{\min}$, N).

You do this by using C, H, $T_{\max}$, N to predict the value for $T_{\min}$ given the fit parameters C_0, C_1, C_2, σ_0, σ_1, σ_2.

If you make a histogram of predicted $T_{\min}$ compared measured $T_{\min}$ you can estimate the goodness of fit χ^2, and hence the uncertainties in your fitted parameters.
You will need to select data so that this model is close to reality. (ie in the range where the C and σ' are approximately linear in humidity and cloudiness.) You may need to investigate whether other types of weather measurements, like windiness, also affect the model.

Remember that the Temperature is in units of Kelvin in the model!

I suggest you do fits to blocks of 5 years worth of data to determine whether there are time variations in the parameters. Note that it is essential that you estimate the uncertainty in the parameters to determine whether any effect you see is statistically significant.

Overall we are looking to see if there is a time variation in σ_0 since this is proportional to the amount of radiative forcing due to the greenhouse effect. Since the CO_2 concentration in the atmosphere has significantly increased since 1950, a statistically significant decrease of σ_0 with time would be a strong indication of human-induced climate change.