Statistical Test of the Global Warming Hypothesis

http://www.leapcad.com/Climate_Analysis/Statistical_Test_of_Global_Warming.xmcd

Statistical test of how well Solar Insolation and CO2 variation explain the variation of temperature over the last 50 years.

The change in temperature of a system is proportional to the net inflow of heat to that system. In the case of the Earth's surface the inflow is due to the influx of shortwave radiation from the Sun. The outflow is due to the longwave thermal radiation which is proportional to the fourth power of the absolute temperature of the surface. The level of greenhouse gases such as carbon dioxide (CO_2) affects the proportion of the thermal radiation retained.

Let S be the sunspot number. This is used as a proxy for the intensity of solar radiation at the top of Earth's atmosphere. If T is the absolute temperature of the Earth's surface then its outgoing radiation is proportional to σT^4, where σ is the Stefan-Boltzmann constant. The proportion of this thermal radiation which is not retained depends upon the concentration of greenhouses gases in the atmosphere, one of which is CO_2. For the statistical analysis below it is presumed that the proportion not retained is a linear function of the concentration of CO_2, p_{CO2}.The estimating equation is then

$$\Delta T = a_0 + a_1 \cdot S - \left(b_0 - b_1 \cdot p_{co2}\right) \cdot \sigma \cdot T^4$$

or, equivalently

$$\Delta T = a_0 + a_1 \cdot S - b_0 \cdot \sigma \cdot T^4 + b_1 \cdot p_{co2} \cdot \sigma \cdot T^4$$

where **ΔT is the January-to-January change in temperature**. It is important to use the change of temperature over an interval rather than the change in the annual average. When a variable is, as temperature is, the cumulative sum of disturbances the process of averaging introduces statistical artifacts that interfer with the statistical analysis. For more on this topic see stochastic structure. The temperature in **σT^4 is however the annual average**.

All three of the coefficients, a1, b0 and b1, should be positive if the hypothesis that an increase in the level of CO2 contributes to an increase in global temperature.

The Data

The data on CO_2 concentration are derived from air samples collected at Mauna Loa Observatory, Hawaii. The source is C.D. Keeling, T.P. Whorf, and the Carbon Dioxide Research Group at the Scripps Institution of Oceanography, University of California at La Jolla, May 2005. This data covers only from 1959 to 2004 so this is the interval of analysis.

The temperature data were constructed from the data set available from NASA. The temperature data goes back to 1880. It is worthwhile to look at some data scatter diagrams to get acquainted with the statistical characteristics of the data. First consider the times series for the January-to-January changes in temperatures.

What the diagram shows is that there were some extreme cases in the early years of the series that had more to do with the accuracy of the data than global climate. This is not a concern for the statistical work below because the analysis only covers the period for which there are data on CO_2 concentrations.

The theory suggests that there should be an inverse correlation between temperature change and the level of temperature, or more precisely the fourth power of the absolute temperature.

Stat Data for Global Temperature Change 1959-2004

(0)Year, (1)J-J Change Temp, (2) AnnualMeanTemp, (3) pCO2, (4) s*T4, (5) s*T4*pCO2, (6) #SS

TStats := READPRN("Climate Statistics.csv") n := rows(TStats)

$\Delta T = a_0 + a_1 \cdot S - b_0 \cdot \sigma \cdot T^4 + b_1 \cdot p_{co2} \cdot \sigma \cdot T^4$ $Yr := TStats^{\langle 0 \rangle}$

$Y_i = \beta_0 + \beta_1 \cdot X1_i + \beta_2 \cdot X2_i + \beta_3 \cdot X3_i + \varepsilon_i$ $y := TStats^{\langle 1 \rangle}$

$\sigma := 5.6704 \cdot 10^{-8}$ Temp SunSpots Radiation CO2

$T_{AnMean} := TStats^{\langle 2 \rangle}$ $x1 := TStats^{\langle 6 \rangle}$ $x2 := TStats^{\langle 4 \rangle}$ $x3 := TStats^{\langle 5 \rangle}$

The design matrix for our Temp Stats data can be constructed with the statements

$i := 0 .. n - 1$ $ONE_i := 1$ $c := 0 .. cols(TStats) - 1$ $p := 3$

$X^{\langle 0 \rangle} := ONE$ $X^{\langle 1 \rangle} := x1$ $X^{\langle 2 \rangle} := x2$ $X^{\langle 3 \rangle} := x3$

FLEX := augment(x1, x2, x3, y)

$b := (X^T \cdot X)^{-1} \cdot (X^T \cdot y)$ $b^T = (73.1787 \quad 8.10165 \times 10^{-4} \quad -0.201 \quad 33.96021)$

$\Delta T := b_0 + b_1 \cdot x1 + b_2 \cdot x2 + b_3 \cdot x3$

$corr(y, \Delta T) = 0.72245$ $RSquare := corr(y, \Delta T)^2 = 0.52193$ $SD_c := stdev(TStats^{\langle c \rangle})$

<u>The regression and estimation results are:</u>

$\Delta T = 73.18 + 0.000810\ S - (0.201 - 33.961\ pCO2)\ \sigma T^4$
(6.68) (1.60) (6.7) (5.87)

$R^2 = 0.522$ **Most of R^2 comes from the solar insolation**

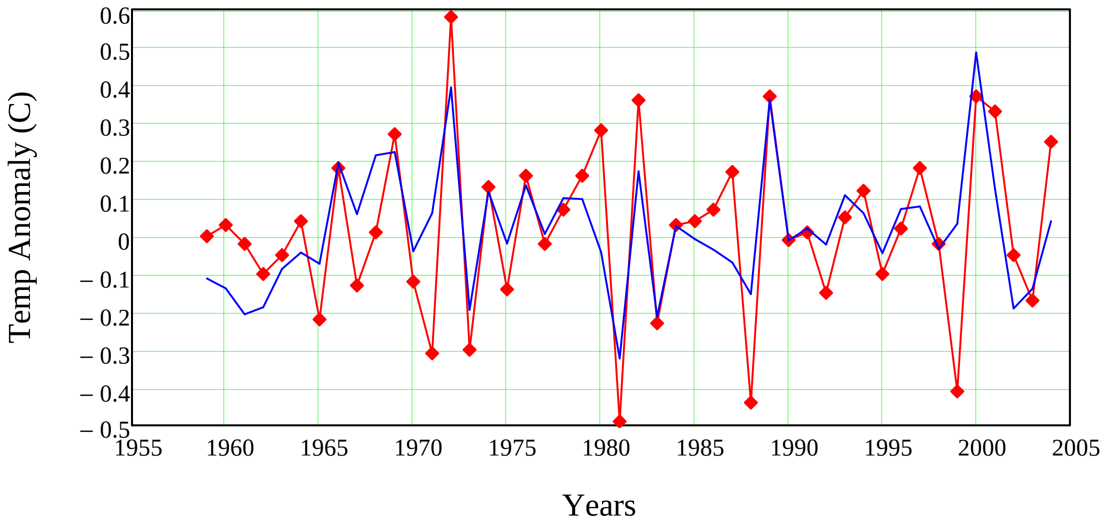
The numbers shown in parentheses below the regression coefficients are the magnitudes of their t-ratios; i.e. the coefficients divided by the standard deviation of the regression coefficient. All but the coefficient for Sunspot number are significantly different from zero at the 95 percent level of confidence and they are of the right sign.

Shown below is a comparison of the observed temperature change and the temperature change predicted by the regression equation. The observations are shown in red and the estimations from the regression equation are shown in blue.

Another way of viewing the comparison is in the scatter diagram below of the actual and regression predicted temperature changes.

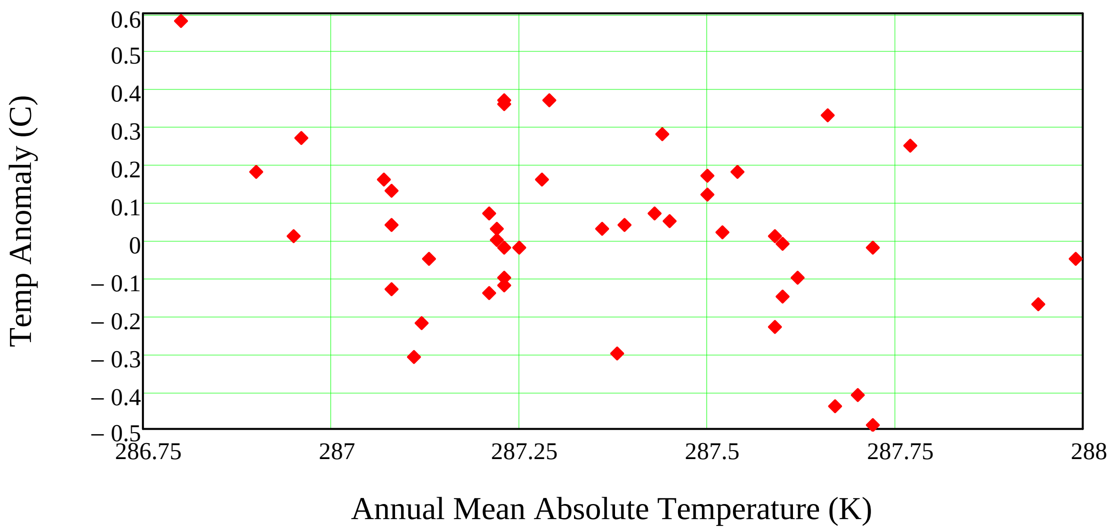
Although the t-ratios for the variables included in the regression equation are significant they **only explain <u>52.2 percent</u> of the variation in the year-to-year temperature change**. The insolation has most of the correlation, not CO2. Also the effect of the CO2 in the equation includes the effects of all variables influencing temperature change which are correlated with the general trend on CO2 concentration but are not in the equation. These would i**nclude the effects of anthropogenic water vapor and anthropogenic cloudiness.**

Comparison of Regression to Temp Data



Consider the scatter diagram for temperature change versus temperature. There is a satifying downward slope to the data plot. When thermal radiation σT^4 is used as the independent variable there is also a downward slope as seen below. But what is clear is that the outliers, the extreme cases, will dominate the statistical results. This raises a note of caution in interpreting the statistical results.

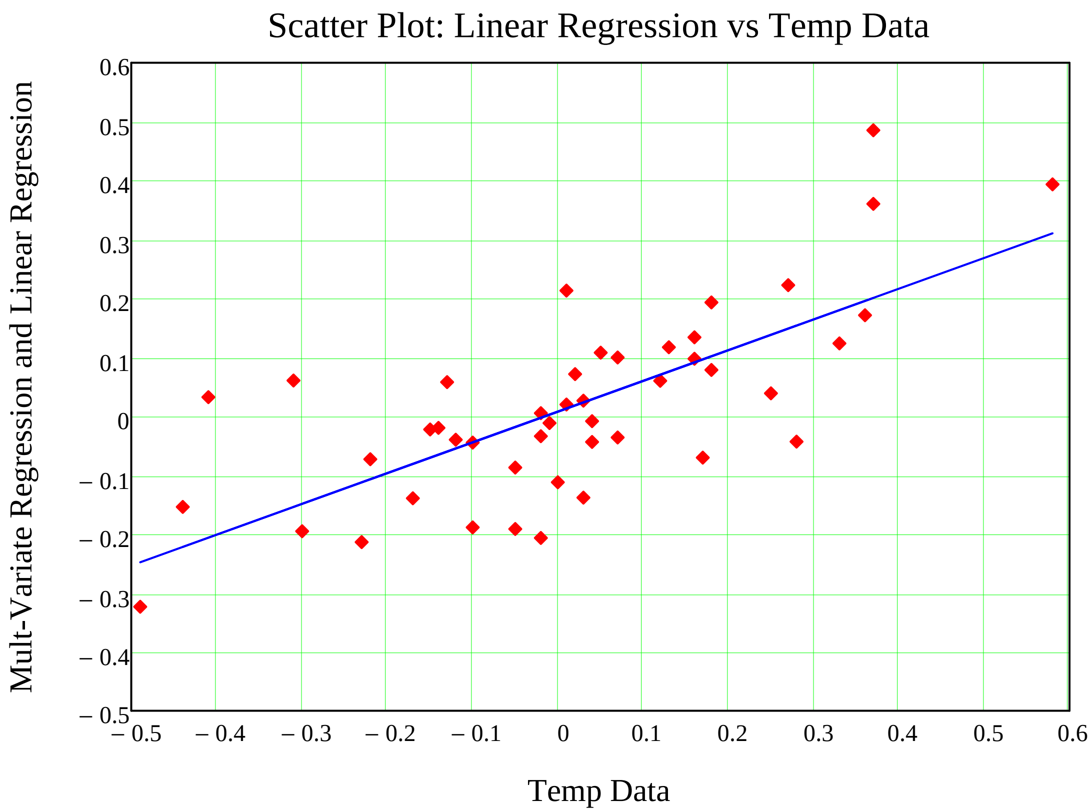
Comparison of Temp Anomaly Data vs Absolute Temperature



$int := intercept(y, \Delta T)$ $\quad$ $s := RSquare$

$int = 8.31423 \times 10^{-3}$ $\quad$ $s = 0.52193$ $\quad$ $\delta T := int + s \cdot y$

Result: Follows a general trend, but Insolation and CO_2 explain only half (52%) of the temperature variation. Must be other major factors



Test for Possible Regression

By extending this test to include p slope parameters

$\mathbf{H_0}$: $\beta_1 = \beta_2 = \beta_3 = ... = \beta_p = 0$

we have the equivalent test for the possibility of a multiple regression,

$\mathbf{H_0}$: no multiple regression relationship

As in simple linear regression, we can associate each data value with three types of deviations, specifically, the residual error, e

$yhat := X \cdot b$ $\qquad$ $e := y - yhat$

Sum of Squares

We can also use matrices to calculate the sum of squares for residual error,

$SSE := e^T \cdot e$ $\qquad$ $SSE = 1.08067$

as well as for regression, $SSR := yhat - mean(y)^T \cdot (yhat - mean(y))$

The total sum of squares equals

$SST := SSE + SSR$ $\qquad$ $DF\_REG := p$ $\qquad$ $DF\_RESID := n - (p + 1)$ $\qquad$ $DF\_RESID = 42$

Mean Squares

Again, as in simple linear regression, dividing each sum of squares by the corresponding degrees of freedom provides us with variance estimates. The mean square for residual error

$MSE := \frac{SSE}{DF\_RESID}$ $\qquad$ $MSR := \frac{SSR}{DF\_REG}$ $\qquad$ $DF\_TOTAL := n - 1$

F Test

The final entry in the table is the F statistic and corresponding p-value for the significance of an overall multiple regression. Under the null hypothesis of

$\mathbf{H_0}$: no regression relationship

the test statistic $F := \frac{MSR}{MSE}$ $\qquad$ $Rsq := \frac{SSR}{SST}$ $\qquad$ $Rsq = 0.52193$

has an F distribution with $n1 := DF\_REG$ $\qquad$ $n2 := DF\_RESID$

degrees of freedom. The p-value of the test, then, is given by

$p\_val := 1 - pF(F, n1, n2)$ $\qquad$ $p\_val = 7.21237 \times 10^{-7}$

Analysis of Variance Table

Summarizing the above for our example,

DF	SS	MS	F
DF_REG = 3	SSR = 1.17982	MSR = 0.39327	F = 15.28451
DF_RESID = 42	SSE = 1.08067	MSE = 0.02573	p-value
DF_TOTAL = 45	SST = 2.26049		$p\_val = 7.21237 \times 10^{-7}$

The amount of variability explained by the linear regression (MSR) is greater than the amount due to residual error (MSE). The difference is large enough (the p-value is, in fact, close to 0) to strongly reject the null hypothesis,

Correlations between each pair of variables in the model can be displayed in matrix form as

$$j := 0..p \qquad k := 0..p$$

$$CORR_{j,k} := corr\left(FLEX^{\langle j\rangle}, FLEX^{\langle k\rangle}\right)$$

Result: Variables are independent -
Little Correlation between "independent" variables

x1 x2 x3 y

$$CORR = \begin{pmatrix} 1 & 0.10515 & 0.05357 & 0.09464 \\ 0.10515 & 1 & 0.84069 & -0.3362 \\ 0.05357 & 0.84069 & 1 & 0.05104 \\ 0.09464 & -0.3362 & 0.05104 & 1 \end{pmatrix} \begin{pmatrix} x1 \\ x2 \\ x3 \\ y \end{pmatrix}$$

$$Var\_Covar\_b := \left(X^T \cdot X\right)^{-1} \cdot MSE \qquad k := 0..p$$

$$se\_b_k := \overrightarrow{\sqrt{Var\_Covar\_b_{k,k}}} \qquad se\_b^T = \left(10.96004 \quad 5.05663 \times 10^{-4} \quad 0.03 \quad 5.7882\right.$$

t tests

$$t := \overrightarrow{\frac{b}{se\_b}} \qquad t^T = (6.67687 \quad 1.60218 \quad -6.69929 \quad 5.86705)$$