Trend Curves for Climate Change

L.K. Jones

Recorded average temperatures exhibit year to year as well as longer term variations. A trend curve averages out short term changes and retains hypothesized behavior. Traditionally straight lines, which best fit the data, were used as trends. Over the last ¾ century temperatures have increased so dramatically that trend curves need more flexibility than linearity to adequately fit temperature data. See graph 1-lin.trend.

Our current research is aimed at verifying our hypothesis that the Arctic mega-dams built in the third quarter of the last century are responsible for at least 50% of the global warming of the northern hemisphere in the last 55 years. We use a trend curve, which we call a hinge, that consists of two straight lines joined at a year (called the hinge year) that best fit the temperature data for a given location. The hinge year is an appropriate mean or median time that the dams most affecting the given location went online. For instance a trend curve for the northern hemisphere might use 1968 as hinge year since 1968 is the median time of addition of mega-reservoir water volume by Canada and Russia. See graph 2-n.hemi.

CO2 data also provides a trend curve by starting at a given year with a temperature given by another trend and increasing the temperature by .01 times the change in CO2 ppm for each future year. See graph 3-s.hemi. Note that the CO2 trend is quite good for the southern hemisphere as is the hinge, even though 1968 might be harder to justify as hinge for the southern hemisphere. But,for the northern hemisphere, temperatures are increasing at twice the rate of the CO2 trend.

A recent controversial publication claims that only 10% of the atmospheric CO2 currently present is from burning fossil fuels.

Although there are some minor questions about some of the mathematical details in the latter paper, we (at IPA) are able to show that the % of atmospheric CO_2 from burning fossil fuels has remained under 20% since 1750. We further believe that the Arctic mega-dams are responsible for more than 50% of the northern hemisphere continental warming since 1968. (please read "Arctic Blue Deserts" by Stephen M. Kasprzak.)Hence, in our hemisphere, we can only blame fossil fuels for at most 20% of 50% = 10%. See graph 4-n.hemi.fossil.

In climate studies smoothed average temperature plots vs. time are also often called “hockey sticks” as they look like two connected linear trends. As we do,Leo Breiman called such functions “hinges” and he generalized the definition to multiple dimensions and characterized their use in regression. See “Hinging hyperplanes for regression, classification, and function approximation”, L.Breiman,IEEE Trans. Inf. Th. May,1993.

Classical Change Point

Suppose average temperatures Y_j are recorded at times t_j where $t_1 < t_2 \ldots\ldots < t_N$. Classical change point analysis finds t_p (or equivalently p, $1<p<N$) where one linear trend changes to the other with a possible “jump” in value. That case is easily solved by first fixing p and then finding the standard linear fit for the data at points less than t_p (equivalently for $j < p$) and finding the standard linear fit for the data at points greater than or equal t_p (equivalently for $j \geq p$). Now using these two fits compute the total squared error SS(p). Repeat the process by varying p and then choosing p that gives the smallest SS(p). Unfortunately this procedure often overfits, yielding an inaccurate p.

Hinge Change Point

In the case of the hinge or hockey stick, we must constrain the two linear trends to agree at the common point p. The hinge change time t_p is the time at which the 2 linear trends agree in value. We do not optimize over p. We prefer to fix p based on some climate hypothesis, as that described earlier concerning Arctic mega-dams.
Following are the mathematical details:

Estimation of hinge change point p (equivalently t_p) and hinge trend curve

Data: for a given location we have available average temperatures Y_j over N different time intervals (years, decades, etc.) $t_1, t_2, \ldots t_N$ arranged in increasing order; e.g. t_1= 1930, t_2= 1935, t_3 = 1937, … t_{93} = 2018. Choose p such that t_p is an estimate of an appropriate mean or median time that the dams most affecting the given location went online. We assume $N \geq 4$. The hinge trend curve takes values on the integers j =1,2,3,…..N (equivalently values on the time intervals $t_1, t_2, \ldots t_N$). When p is the fixed change point ($2 \leq p \leq N - 1$), the curve has the form

$$c + b(t_j - t_p) \qquad j = 1,2,\ldots p-1$$

$$c + a(t_j - t_p) \qquad j = p, p+1, \ldots N$$

To obtain the parameters c, a,and b, solve the maximum likelihood equations (equivalently minimize sum of squares, SS(p)).

Details: $SS(p) = \Sigma_1 (c + b(t_j - t_p) - Y_j)^2 + \Sigma_2 (c + a(t_j - t_p) - Y_J)^2$

where Σ_1 denotes summation over values of j = 1,…p-1 and Σ_2 denotes summation over values of j = p,…N.

We take partial derivatives of SS(p) wrt. c,a, and b . Setting them = 0 and dividing by 2, we get the 3 (normal) equations for c,a,b.

$$\Sigma_1 (c + b(t_j - t_p) - Y_j) \quad + \quad \Sigma_2 (c + a(t_j - t_p) - Y_J) \quad = \quad 0$$

$$\Sigma_2 (c + a(t_j - t_p) - Y_J)(t_j - t_p) \quad = \quad 0$$

$$\Sigma_1 (c + b(t_j - t_p) - Y_j)(t_j - t_p) \quad = \quad 0$$

These equations can be put in matrix form $A (c,a,b)^T = h$ where

A is the 3by 3 matrix with entries

$A_{11} = N$ $\quad A_{12} = \Sigma_2 (t_j-t_p)$ $\quad A_{13} = \Sigma_1 (t_j-t_p)$

$A_{21} = \Sigma_2 (t_j-t_p)$ $\quad A_{22} = \Sigma_2 (t_j-t_p)^2$ $\quad A_{23} = 0$

$A_{31} = \Sigma_1 (t_j-t_p)$ $\quad A_{32} = 0$ $\quad A_{33} = \Sigma_1 (t_j-t_p)^2$

And h is the column vector

$$(\Sigma_1 Y_j + \Sigma_2 Y_j \, , \, \Sigma_2 Y_j (t_j-t_p) \, , \, \Sigma_1 Y_j (t_j-t_p))^T$$

CODE { note I believe v^T is Python for v^T, transpose of vector(or matrix) v; also A^-1 in Python is A^{-1}, inverse of matrix A }

Finally to get the hinge parameters: Input p,Y_j ,N , t_j

Compute $(c,a,b)^T = A^{-1} h$. Calculate SS(p).
Output hinge curve using p, c, a, b and the residuals SS(p).