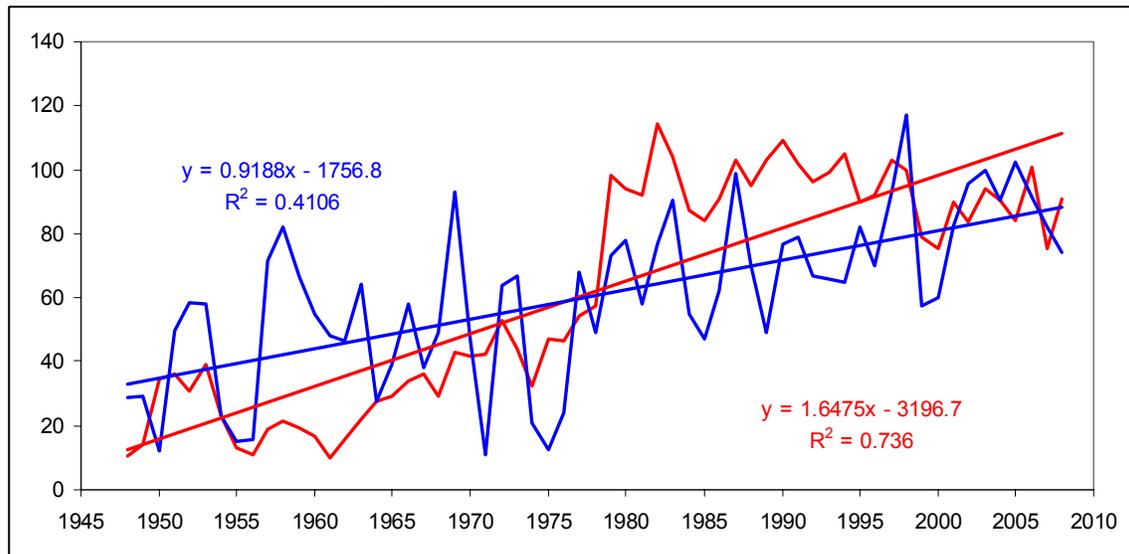
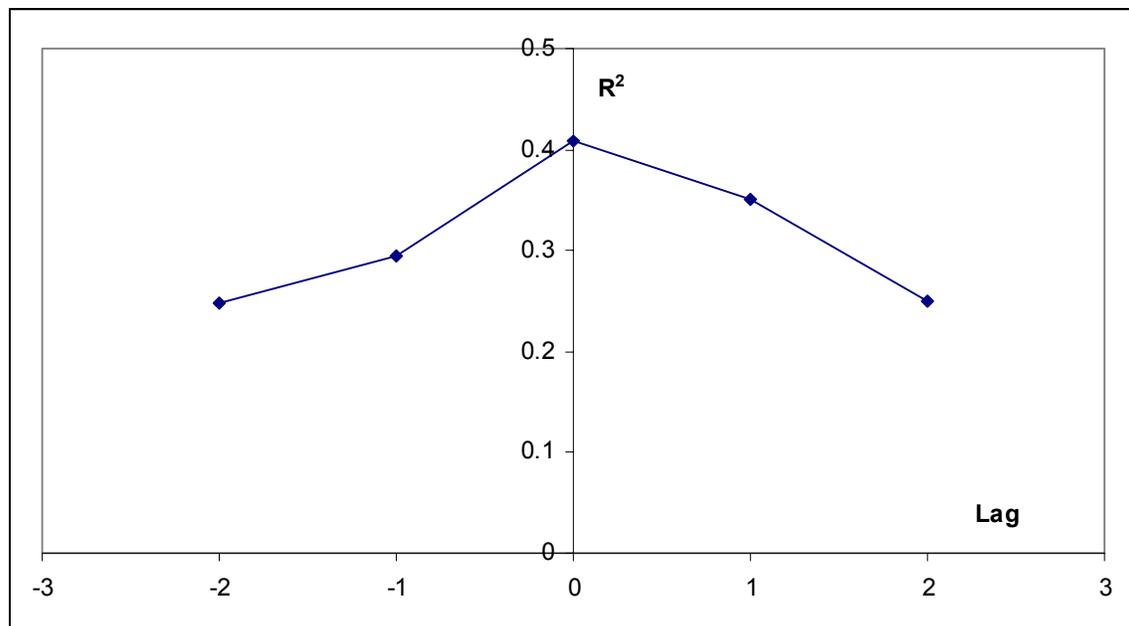


200 hPa versus SST in tropics:



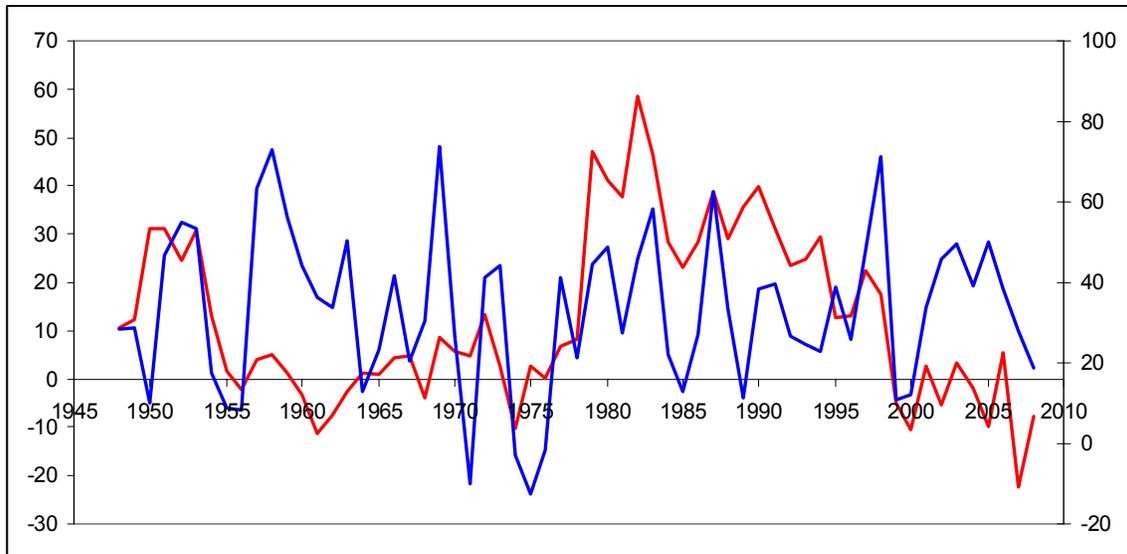
Red = 200 hPa 30S-40S, 240-260E, blue = SST tropics 20N-20S

We now calculate the ‘coefficient of determination’, also known as R^2 , for the cross correlation between the two curves. This number says how large a fraction of the variation in one curve matches or ‘causes’ – if you believe they are causally related. E.g. if $R^2 = 0.35$, 35% of the variation matches that in the other curve. One can calculate R^2 for different ‘lags’, i.e. move one series up a year [or down for a negative lag], then correlate again and note R^2 . Here is a plot of R^2 for lags between -2 and +2 years:

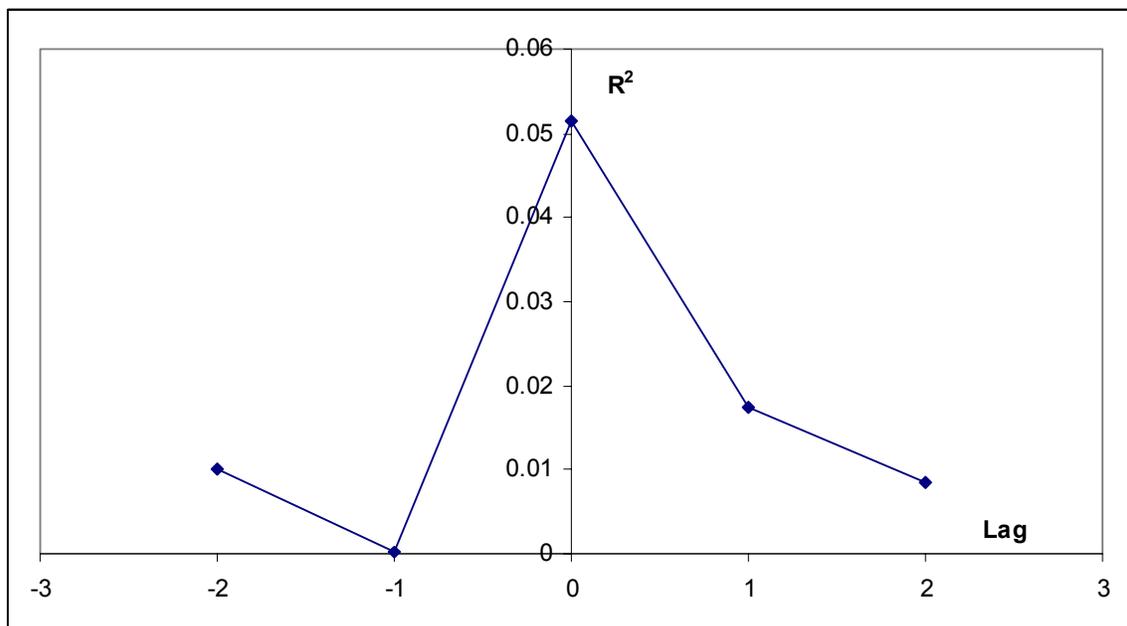


The largest R^2 occurs for lag = 0, but the coefficient is large for several lags. There is not much difference. The reason for this is obvious: There is a large 'auto-correlation' present: and both curves show an upwards trend, so the analysis so far just says that there is a similar upwards trend in both curves. Moving one curve over a year does not change the trend much. It is clear that the trend must first be removed if we want to compare the fine details.

Removing the trend, the curves look like this:

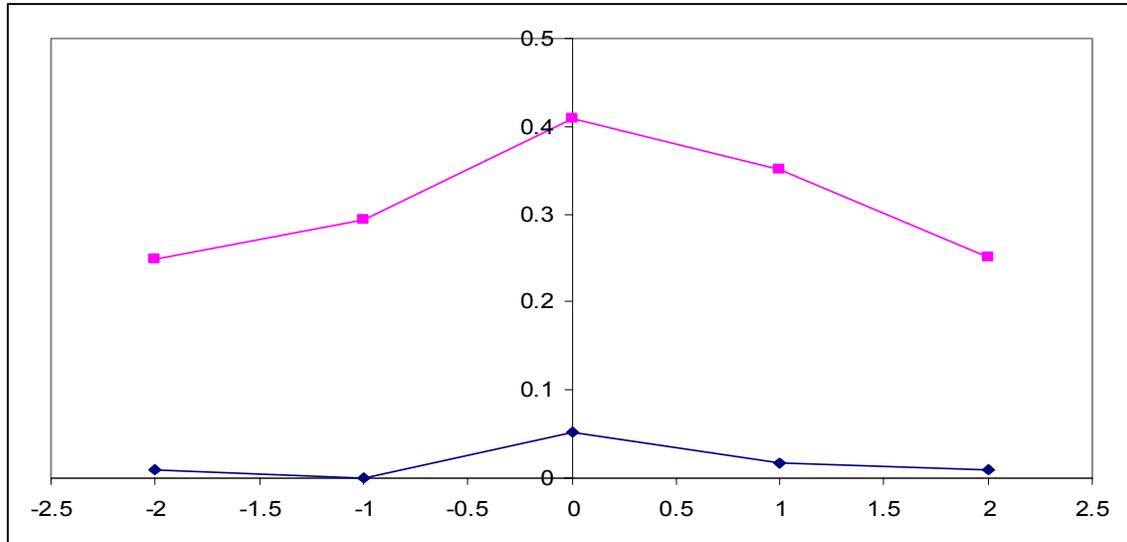


And the plot of R^2 as a function of lag looks like this:



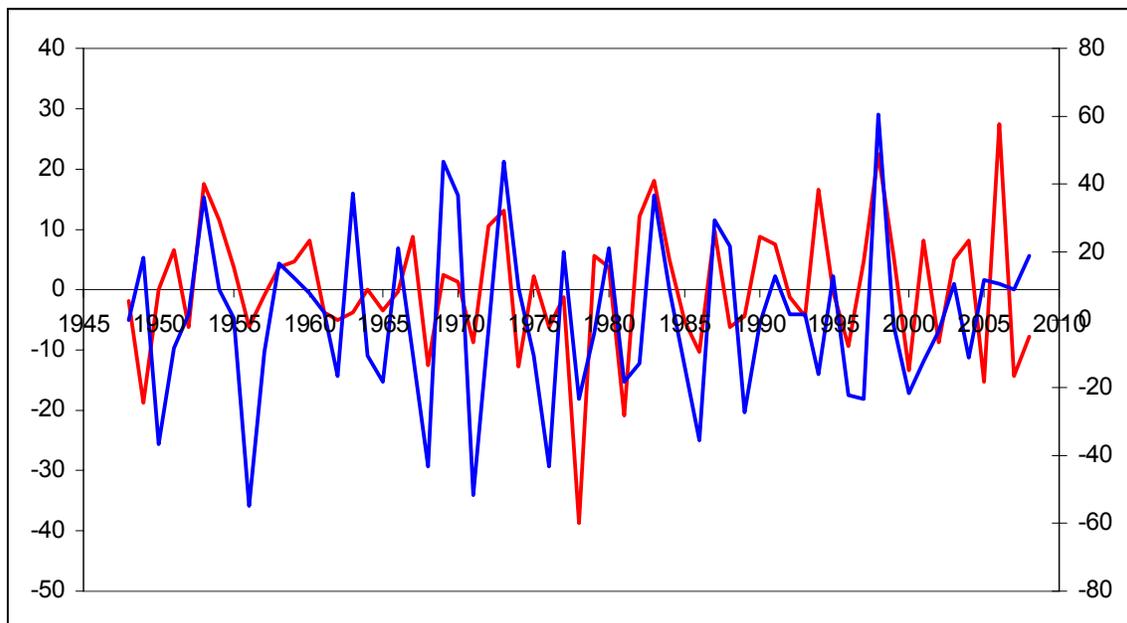
Thus about 5% of the variation of one curve matches that of the other curve at lag zero years.

If we plot both set of coefficients as function of lag, we have:

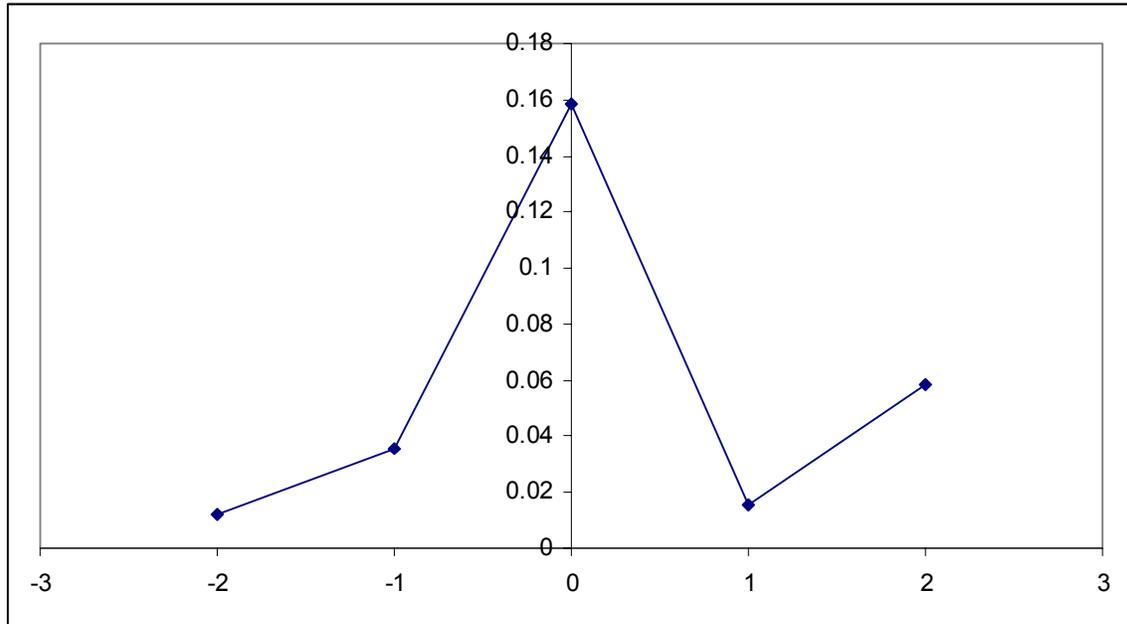


Just showing at a glance the effect of the high autocorrelation. The conclusion is that 95% of the variations are unrelated, and the 5% that match do that at zero lag.

Now, people are often very creative in ‘massaging’ the data. One standard trick is to plot the difference between successive values. This accentuates the variation:

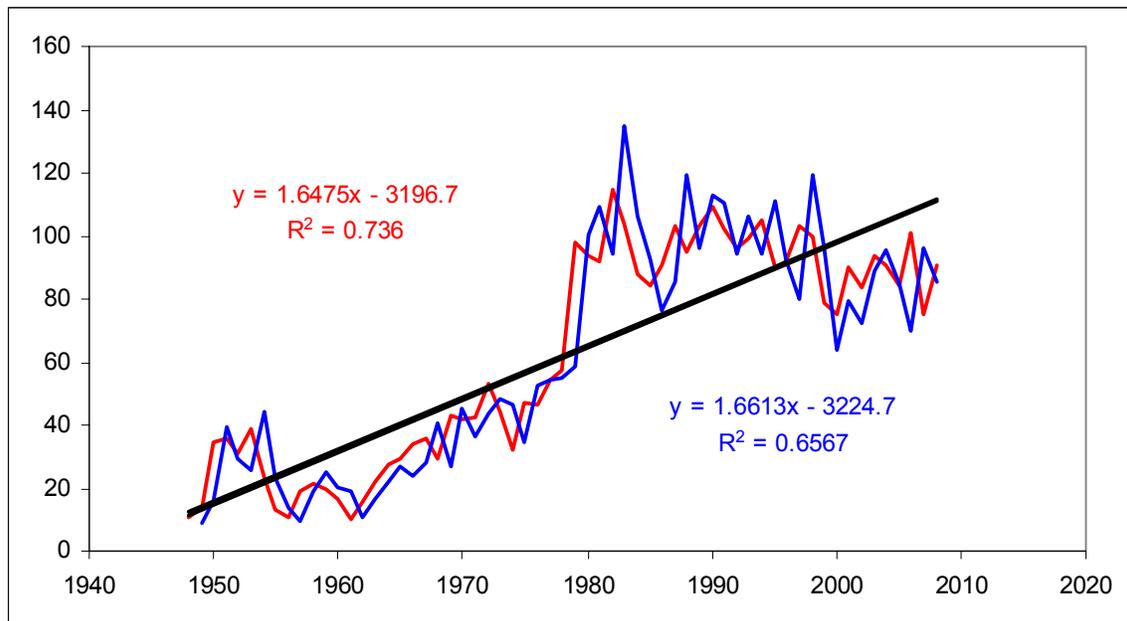


And indeed, the coefficient of determination more than triples:



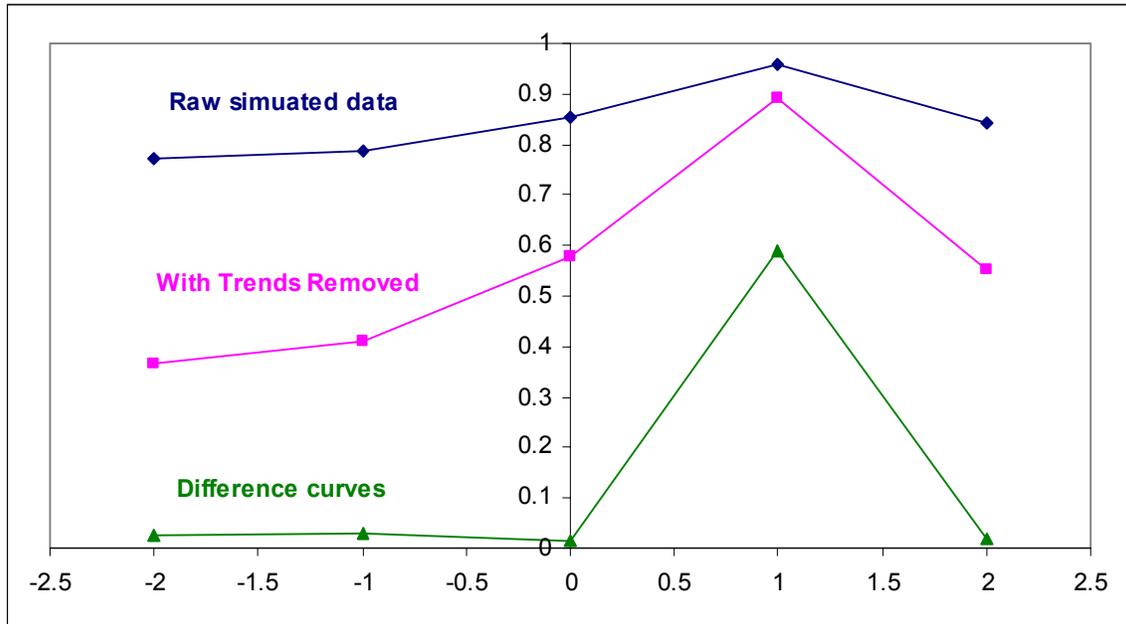
But still with the best match at zero lag. At this point it becomes clear to an old hand as I in data ‘massaging’ that further and more desperate measures will not alter the conclusion.

How good is this method? To answer that we take one of the time series and construct the other one by copying the data shifted one year plus adding some random noise:



Here it is clear there is a relationship with a lag of 1 year.

We now go through the same procedure as above, and plot the resulting R^2 s on one plot:



You can clearly see the peak at lag one year.

Extensive trials with more and more random noise added confirm that if there is a signal, this method will find it.