Applications of Hurst Coefficient Analysis to Chaotic Response, Temperature Data, and GCM Results: A First draft

Dan Hughes
November 2008

Abstract
I have coded the process for calculating Hurst coefficients described on Page 19 of this preprint (http://www.itia.ntua.gr/en/docinfo/673/) by D. Koutsoyiannis:

Koutsoyiannis, D., Nonstationarity versus Scaling in Hydrology, *Journal of Hydrology*, **324**, 239–254, 2006.

The objective was to apply the analysis method to (1) results from classic systems of non-linear ODEs that exhibit Chaotic Response, (2) measured temperature data, and (3) numbers calculated by a GCM.

The results indicate that the Global Mean Surface Temperature, as reported from both measured data and GCM calculations, does not contain sufficient information to warrant investigations using the Hurst Coefficient approach. Applications to finer-scale data might prove useful.

Introduction
I was struck by these statements at the bottom of Page 20 of the preprint:

"The Hurst or scaling behaviour has been found to be omnipresent in several long time series from hydrological, geophysical, technological and socio-economic processes. Thus, it seems that in real world processes this **behaviour is the rule rather than the exception**. The omnipresence can be explained based either on **dynamical systems with changing parameters** (Koutsoyiannis, 2005b) or on the principle of maximum entropy applied to stochastic processes at all time scales simultaneously (Koutsoyiannis, 2005a)."

I wanted to explore the presence, or not, of this "omnipresence" behavior in the systems listed in the title. The characterization by Koutsoyiannis of "dynamical systems with changing parameters" also happened to fit with my previous discussions of the Chaotic Response of some systems of non-linear ODEs. The *phenomenology* of these systems has been adopted as the expected response from GCM codes.

Additionally, the discussions by Koutsoyiannis regarding fitting **assumed** trends to data series also struck a cord. The systems of non-linear ODEs cannot produce response having a trend. I wanted to see how this fact fit into the accepted wisdom of Chaotic Response and GCM results. The difficulties associated with assuming trends for measured data were introduced in the Abstract of the paper as follows:

"The most common modelling approach is to assume that long-term trends, which have been found to be omnipresent in long hydrological time series, are "deterministic" components of the time series and the processes represented by the time series are nonstationary. In this paper, it is maintained that this approach is contradictory in its rationale and even in the terminology it uses. As a result, it may imply misleading perception of phenomena and estimate of uncertainty."

Almost every discussion about the Global Mean Surface Temperature (GMST) includes assigning a linear trend to the recent measurements of the surface temperature. I wanted to see what those time-series look like relative to Hurst Coefficients.

Verification of the Code and Calculations

The equations needed to do the analysis have been given in the reference and I won't repeat them here. Instead, I'll report the results of Verification exercises that I used to determine that the equations were correctly coded and were being correctly solved.

I made a routine that returns the results of the Hurst-coefficient analysis when it is fed an array of values for the quantity of interest. This way I can use the same routine in any arbitrary application and not have to change the coding for this step in the analysis procedure.

Because application of the method to numbers calculated by use of Lorenz-like systems of non-linear ODEs that exhibit Chaotic Response was an early objective, I used these equations to get started. I used the original Lorenz system of 1963. The equation system has been given in this post, discussed several numerical solution methods in this post, and discussed some results in this post. The classic values of the parameters; Ra = 28.0, Pr = 10.0, and b=8.0/3.0 were used. The explicit Euler solution method was used and I'm certain that use of any other solution method would not change the results reported here.

The first Verification calculations were done by forcing the numerical integration routine to return constant values; those values being the initial conditions for the integration. This test is frequently termed a null transient and is very useful for finding incorrect coding. The explicit Euler method is especially simple in this regard; simply set y(i) = yold(i) after initially setting yold(i) to the IC for each dependent variable. So as to make the Verification as simple as possible, I used numerical values of 1.0, -10,0. and 10.0 for the ICs. With these I could do some of the calculations in my head to compare with those calculated by the routine. I generated "solutions" for 100,000 steps of the "integrations" and sent these to the Hurst coefficient routine. This test was successful, returning 0.00 for the standard deviations for all three equations for all lengths of the series.

I next generated 1,000,000 random numbers, broke the results down into 18 series, and plotted the results as indicated by Koutsoyiannis. I used more than the recommended number of series just for fun.

The results are shown in Figure 1 nearby. The standard deviation for each series is shown as a function of the number of items in the series. The left side of the plot represents the case of many short series (small number of items in the series) and the right side represents the case of few longer series. The last series case I ran (n = 100,000) has a total of 10 series on the right side. The second case on the left side (n = 2) has a total of 500,000 series.

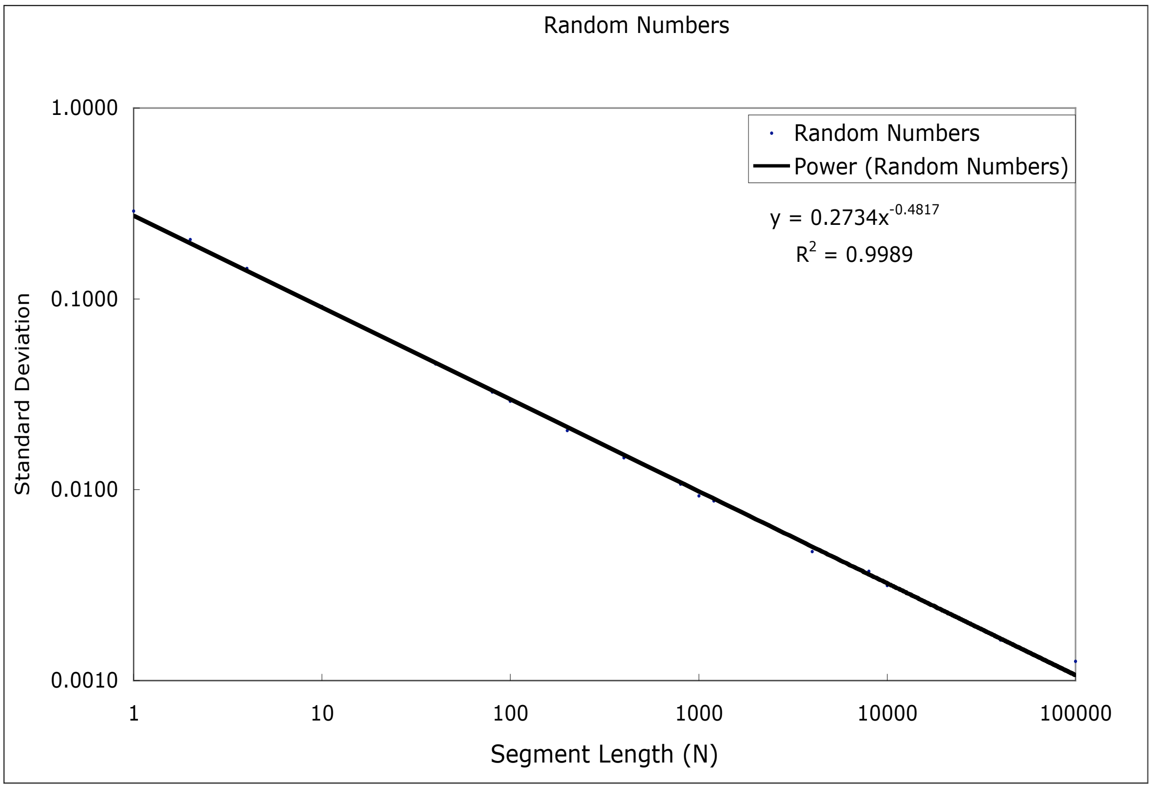


Figure 1. Standard deviations for one million random numbers.

The results in the Figure are in accord with the description of the process given by Koutsoyiannis. A power-law regression line is shown on the Figure and the built-in routine in Excel gives the slope to be –0.4817, compared to the theoretical value of –0.50. The fit would very likely be better if I removed the last long series or two. I didn't try that yet.

Application to the Original Lorenz System of 1963

So, now we want to see what Chaotic Response looks like in the Koutsoyiannis method. As noted previously, chaotic response cannot return results that have a trend associated with them; that has been demonstrated an almost uncountable number of times in all investigations of chaotic system response. The parameter values and ICs used in the null transient calculation above were also used in this analysis. Again 100,000 values were generated by running the solution out to 100.0 Lorenz Time Units (LTU) using a step size of 0.001.

The calculated results for one of the dependent variables, usually denoted to be X(t) is shown in Figure 2. Nothing new there.

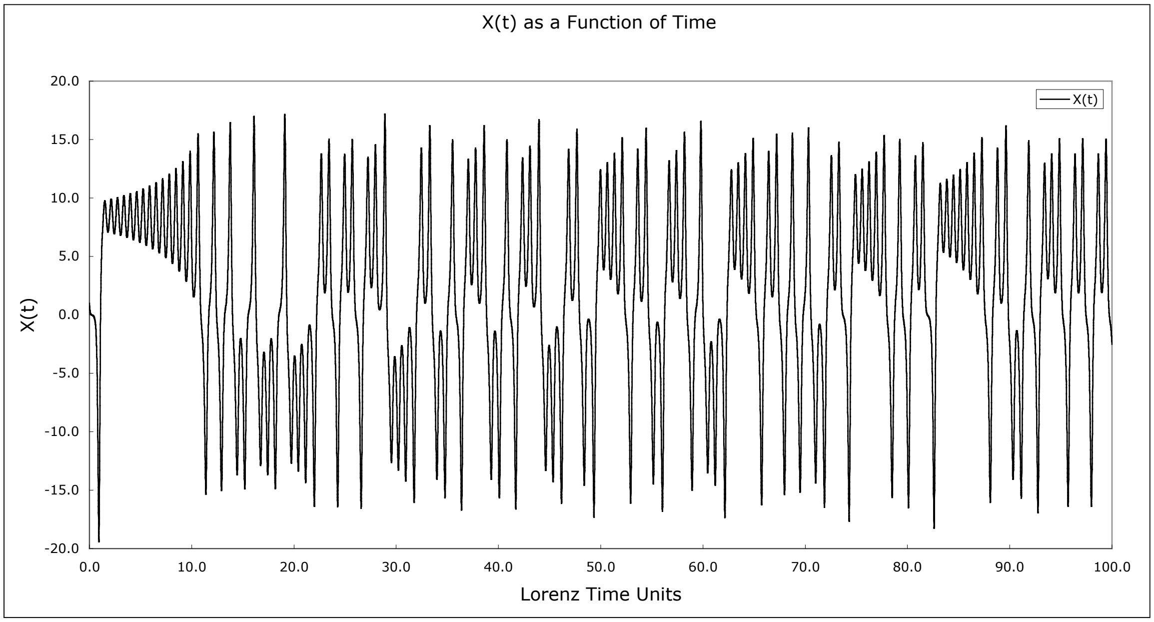


Figure 2. X(t) as a function of time.

The results of the Hurst coefficient analysis is given in Figure 3 nearby. The results shown in Figure 3 are for the complete run time of 100,000 steps and for segment lengths from 1 to 25,000. Note that the latter value corresponds to four segments. Now we see something interesting.

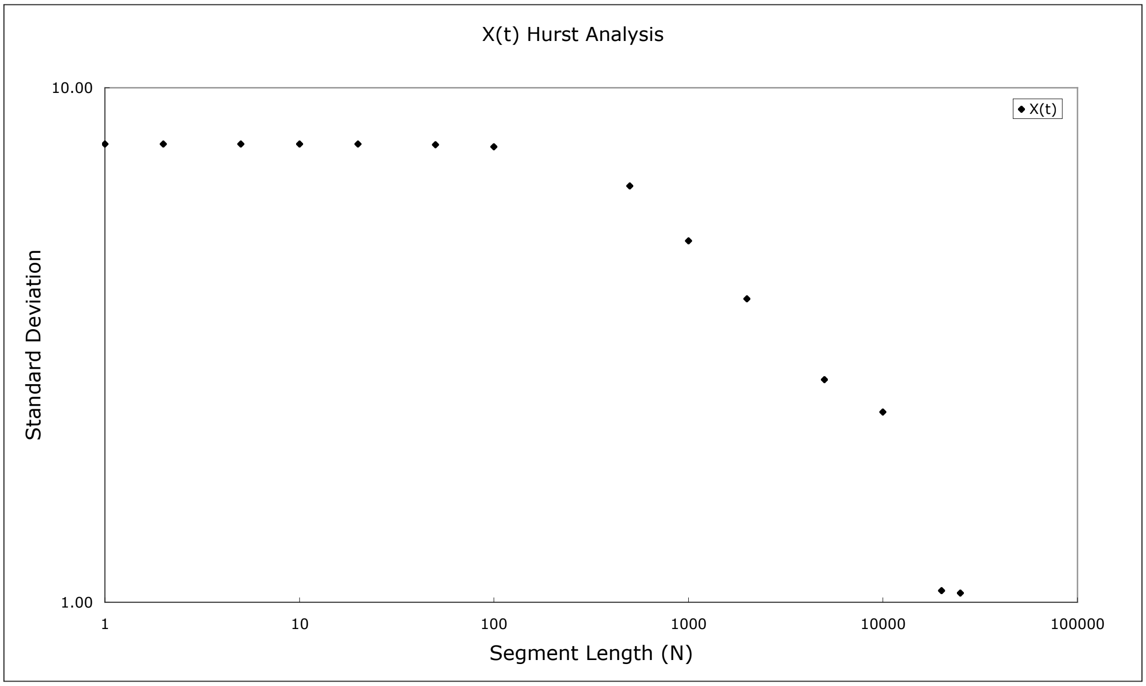


Figure 3. Hurst analysis plot for the complete Lorenz calculation.

For segment lengths from 1 to about 500 the standard deviation is very nearly constant, at about 7.77 for this particular case. Many analyses of the Lorenz systems eliminate a “start-up” part of the results. Based on experience from previous analyses, I suggest that about 20 LTUs is a good place to start. Re-doing the analysis, now having 80,000 points, from 20 LTUs to 100 LTUs gives the results in Figure 4 nearby.

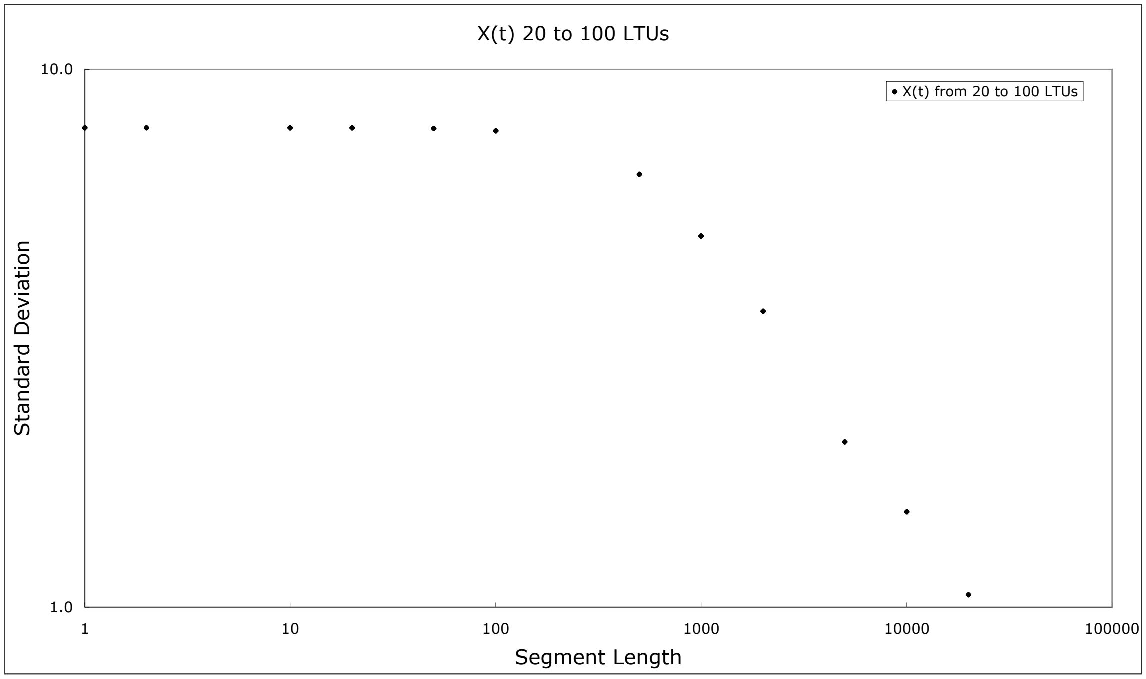


Figure 4. Hurst analysis plot for 20 to 100 LTUs; total length 80,000.

Well, we see that the results look the same as those for the complete run shown in Figure 3 above. Not surprising, in my opinion. The constant value is now about 7.79.

The definite break in the results at about segment length of 500 needs to be investigated. So, I plotted these for the complete run of Figure 3 and got Figure 5.

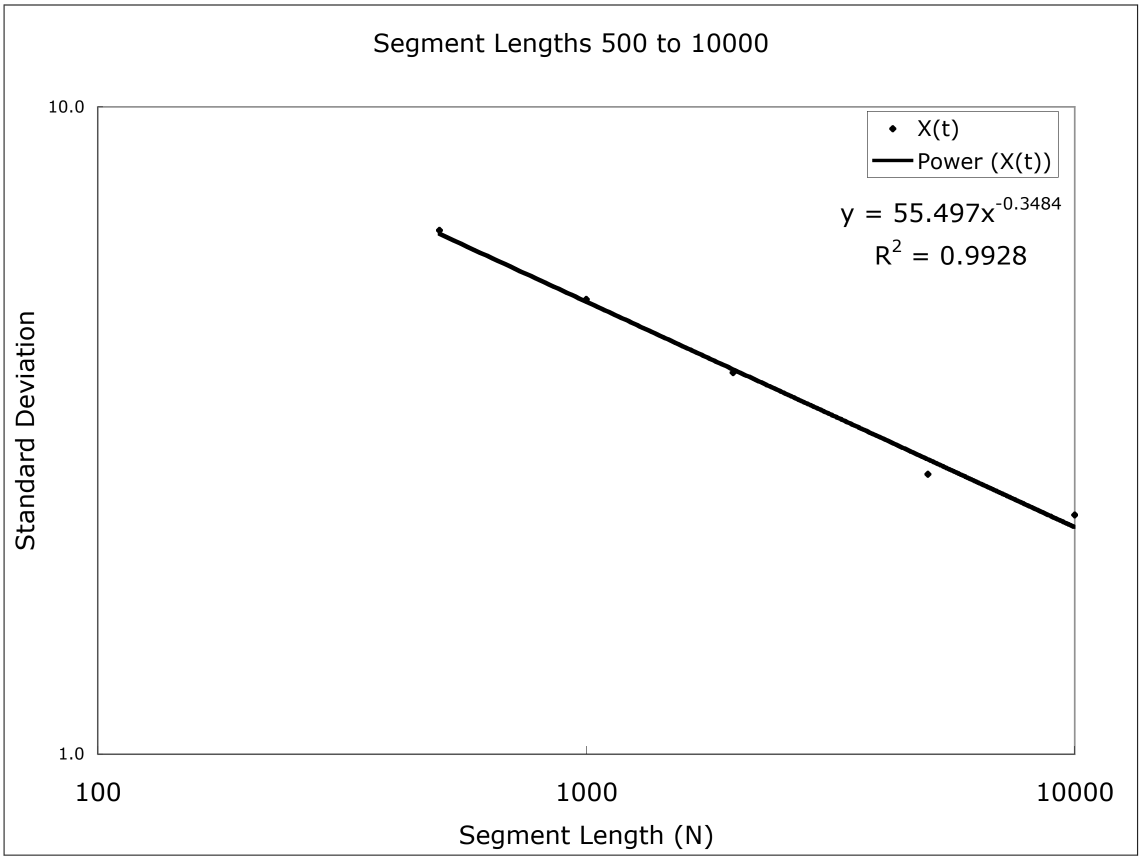


Figure 5. Hurst analysis for segment lengths 500 to 10,000 for the complete run.

A power-law fit to the data points gave a slope of –0.3484; Hurst coefficient 0.652. The corresponding plot for the run from 20.0 to 100.0 LTUs is shown in Figure 6.

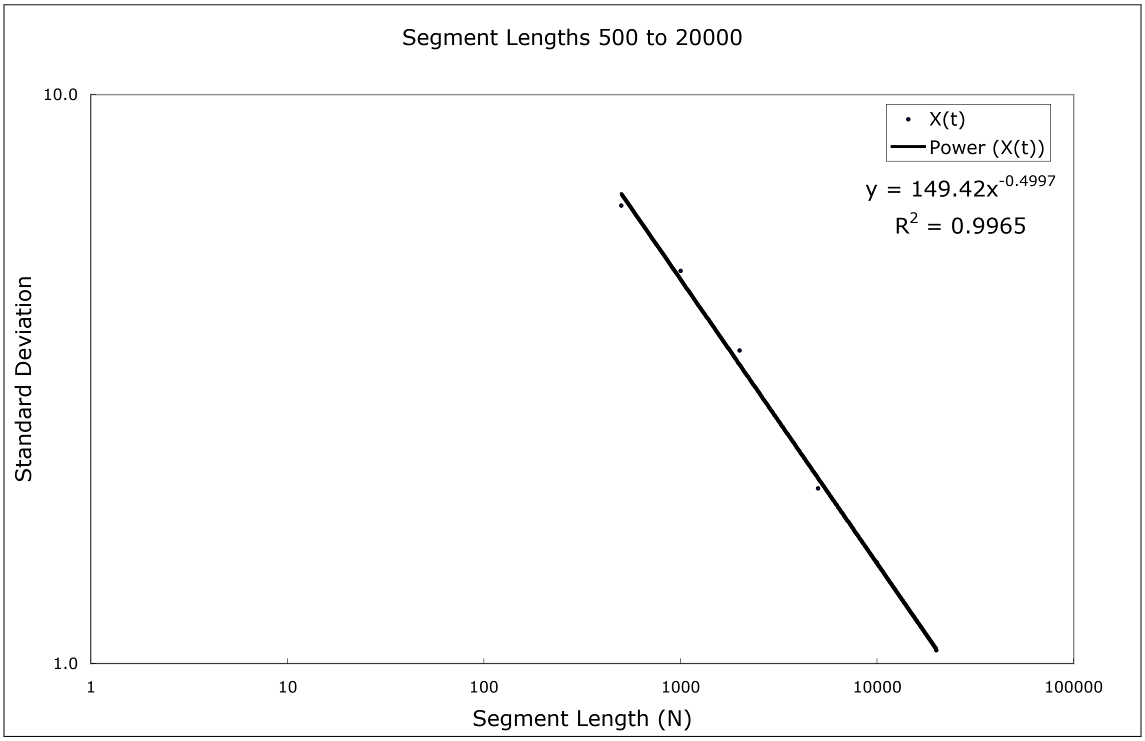


Figure 6. Hurst analysis for segment lengths 500 to 10,000 for 20 to 100 LTUs.

The Hurst coefficient is now about 0.50, as in Figure 1 above in the case of random numbers. When using fewer and longer segment lengths, the stats approach classical stats. I'm not yet sure what this means. It is clear that the over-all Chaotic Response is not random.

Repeating the above analysis using anomalies gives the same results.

I have not completed an analysis for Y(t) and Z(t), but a preliminary first-look indicates those results will be the same as for X(t) given here.

Temperature Data

Next let's look at what measured data for the temperature looks like. These results should be comparable with the physical processes discussed by Koutsoyiannis.

In a comment on Climate Audit someone posted a link to some temperature data from somewhere in Sweden. The data for January from about 1802 to 2002 are shown in Figure 7. Note that this data do not appear to be chaotic.

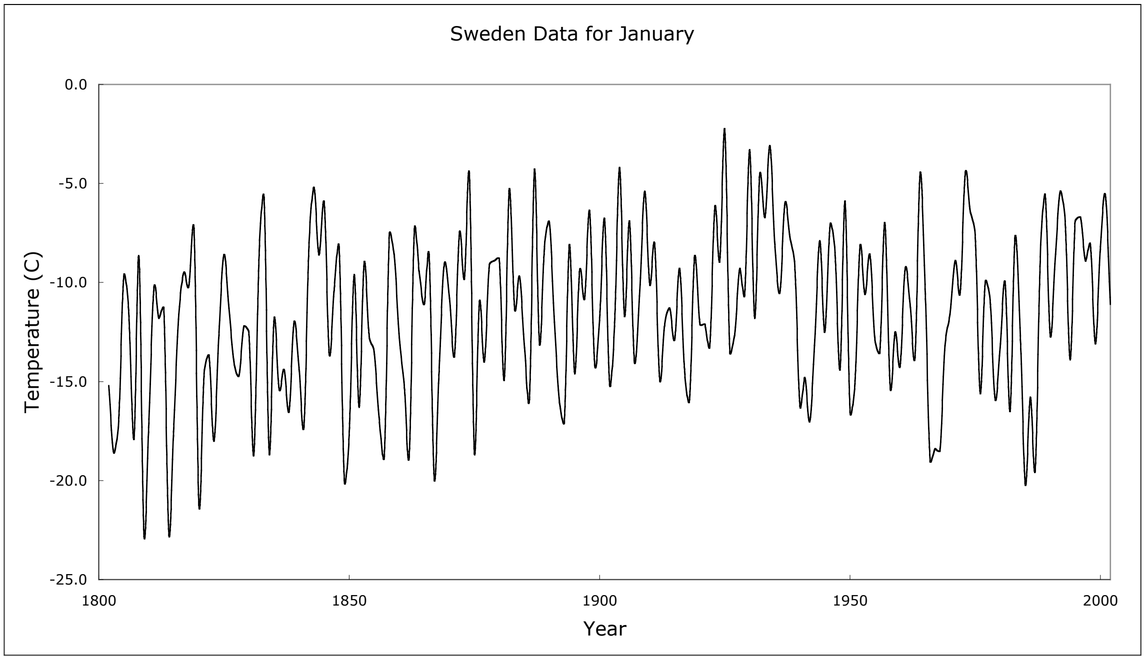


Figure 7. Temperature for January in Sweden from about 1802 to 2002.

I used a 200 year chunk of these data for the Hurst analysis, using 11 segments having segment lengths from 1 to 100, and got the results shown in Figure 8.

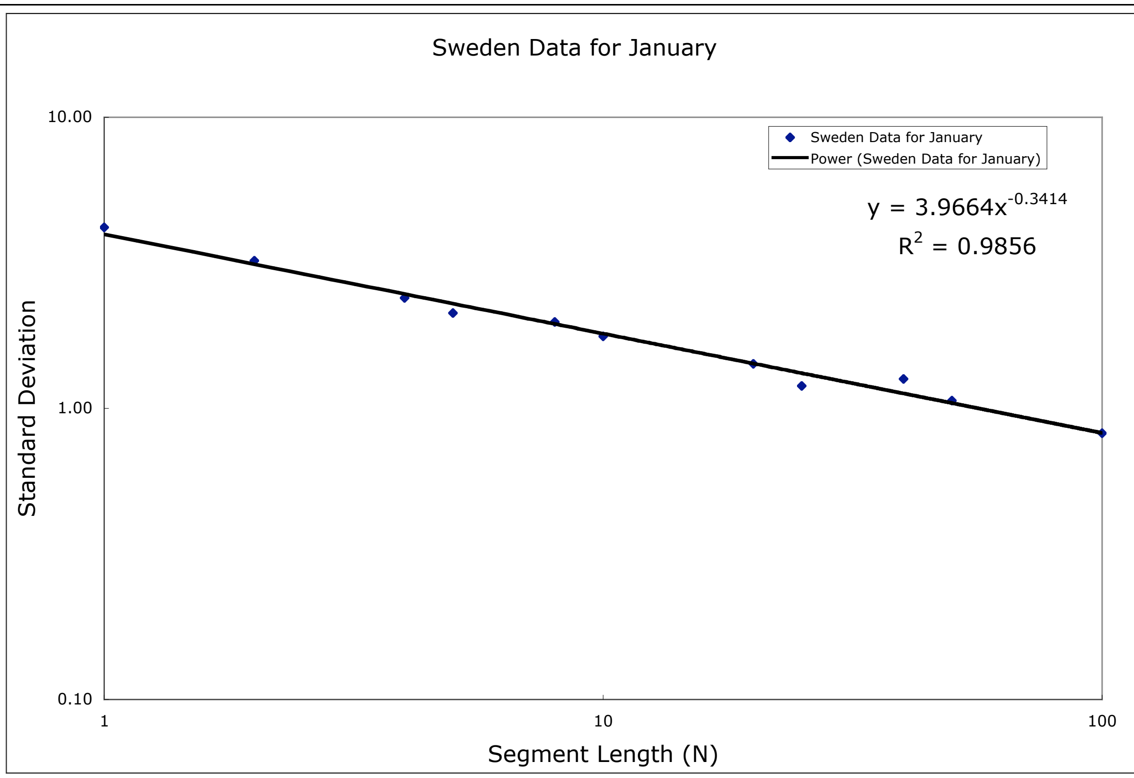


Figure 8. Hurst coefficient analysis for the data in Figure 7.

The slope of the power-law fit to the data is about –0.3414, indicating that the data are not random. Note that the measured data do not have the characteristics of Chaotic Response as seen in the analysis of the numerical solutions of the Lorenz equation system. And the data are not random.

Looking for a longer record to investigate, I chose the monthly-average Central England (European) Temperature, a very long record. A 350 year chunk with segment lengths from 1 to 70 was used. The Hurst coefficient plot for the month of March is shown in Figure 9.

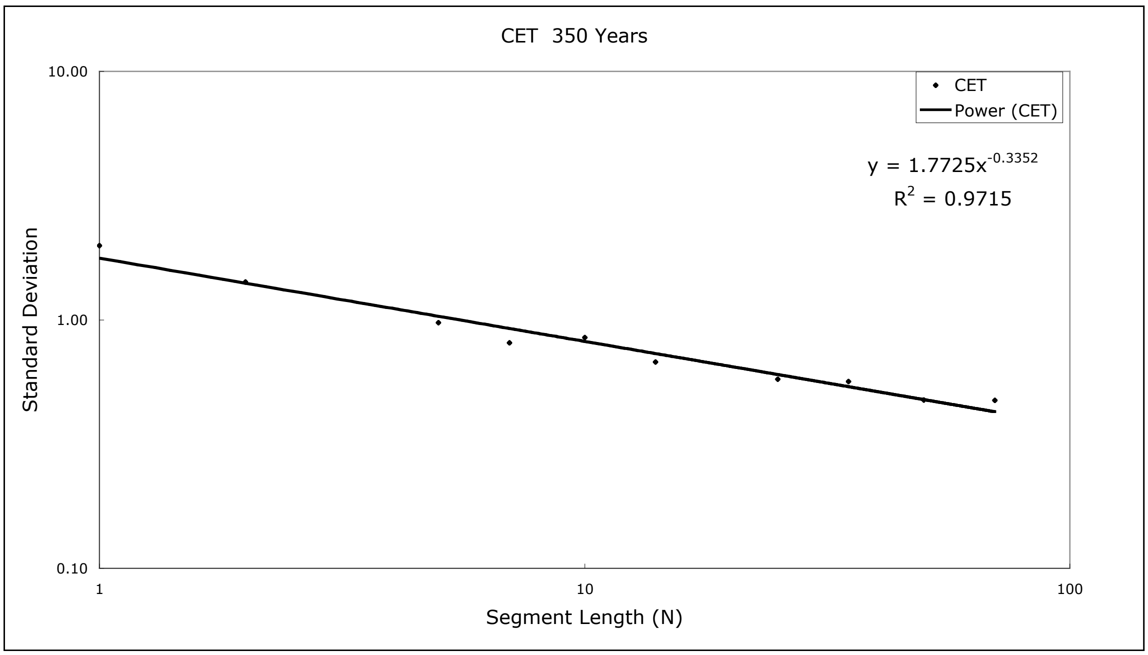


Figure 9. Hurst coefficient analysis for 350 years of CET data.

The slope of the power-law fit to the data is about –0.3352, indicating that the data are not random. Note that again the measured data do not have the characteristics of Chaotic Response as seen in the analysis of the numerical solutions of the Lorenz equation system. And the data are not random.

I will expand this analysis to the daily-average data Real soon now.

Global Mean Surface Temperature

Next I used the HadCRUT yearly-mean GMST data and the results are shown in Figure 10. The HadCRUT data are reported as anomalies. The analysis used 160 data points broken into 11 segments.

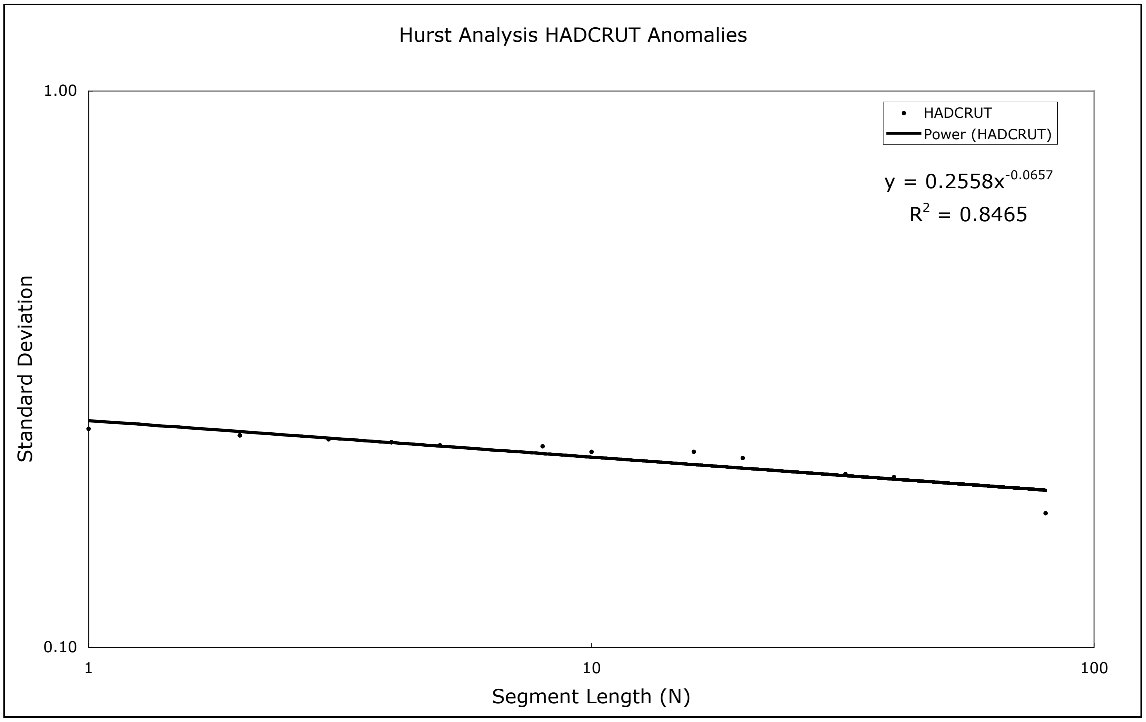


Figure 10. Hurst analysis for the HadCRUT yearly GMST.

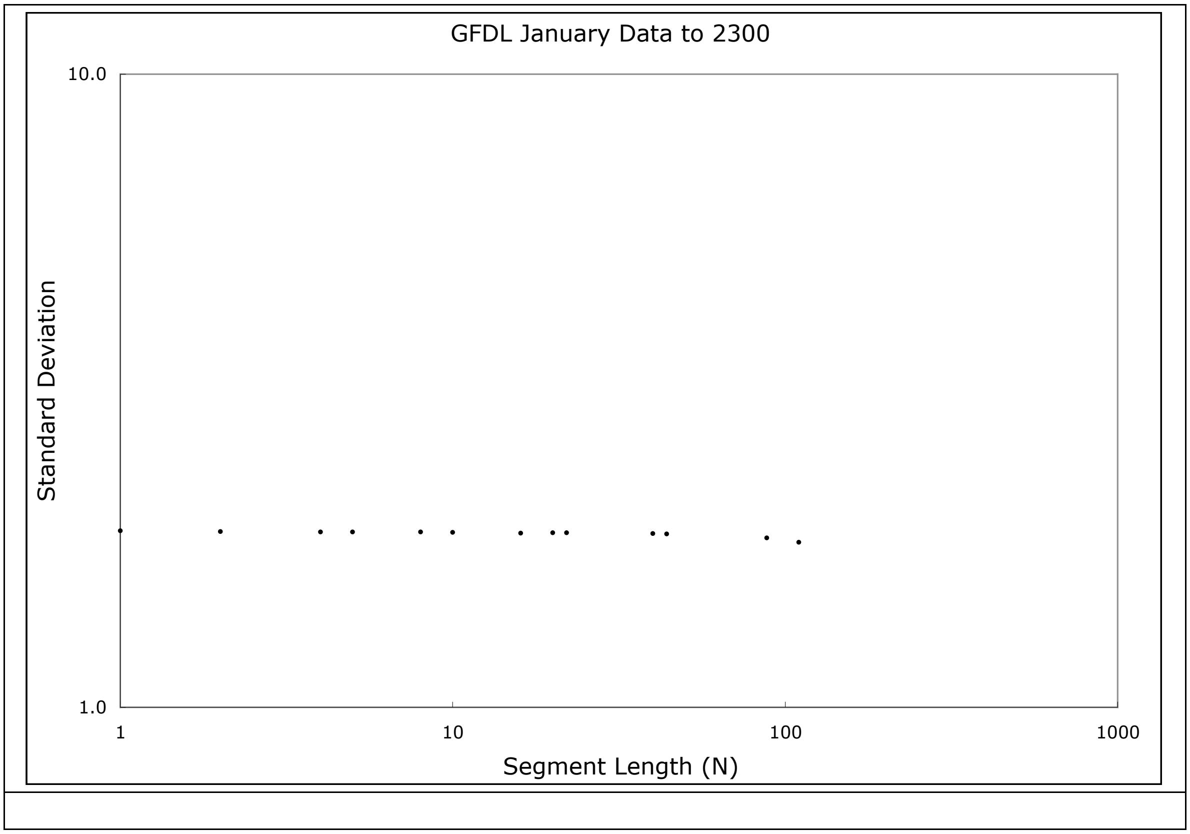
The results have characteristics between those of Chaotic Response from systems of ODEs and the temperature data for Sweden and CET. Averaging over the entire planet annihilates an enormous amount of information content. I'm not yet sure about any other ramifications of the results.

GCM Data

A representative example of application of the Hurst methodology to data produced by a GCM calculation is of interest. Using the facilities at Climate Explorer, I got the monthly-average results of a calculation:

```
# using minimal fraction of valid points 30.00
# tas [Celsius] from GFDL CM2.0, 20C3M (run 1) climate of the 20th Century
experiment (20C3M) output for IPCC AR4 and US CCSP
```

The GMST for January was used in the analysis of 440 years of data with 13 segment lengths ranging up to 110. The results are shown in Figure 11.



As shown in the Figure, there is no useful information left in this grand-average approach. And we're back to what the what the results of the Lorenz systems look like.

Conclusions

The results indicate that the Global Mean Surface Temperature, as reported from both measured data and GCM calculations, does not contain sufficient information to warrant investigations using the Hurst Coefficient approach. Applications to finer-scale data might prove useful.