Long Memory or Structural Breaks in Temperature and Proxy Time Series

[1]Rea, William, [1]Reale, Marco, and [1]Brown, Jennifer

[1]Department of Mathematics and Statistics
University of Canterbury
Ilam
Christchurch
New Zealand
E-Mail: marco.reale@canterbury.ac.nz

Keywords: *long range dependence, climate change*

EXTENDED ABSTRACT

Since the end of the last ice age the earth's climate has enjoyed a period of relative stability. As the earth is now in a period of rising global temperatures a number of authors have considered the stochastic properties of time series of both atmospheric and oceanic temperatures from instrumental and proxy records on time scales of a few decades to several millenia in an effort to estimate the natural variability of the earth's climate. These series almost universally exhibit the property of statistical long memory.

Long memory time series were brought to prominence by H.E. Hurst in 1951 in his study of river flows. Since then the physical cause or causes of the so-called Hurst phenomena have remained elusive. Two sets of competing models have been proposed. The fractional Gaussian noises (FGNs) and their discrete time counter-parts, the fractionally integrated processes of order d (FI(d)), possess genuine long memory in the sense that the present state of a system is temporally dependent on all past states. The alternative are models with a non-stationary mean. In these models the long memory is merely an artifact of the method of analysis. Some authors have proposed multifractals as a potential model. These are FGNs or FI(d) series in which the self-similarity parameter, H, or fractional integration order, d, is allowed to change with time.

A number of authors have attempted to develop statistical tests to distinguish between true long memory and other types of processes displaying statistical long memory. Most of these tests exploit, in some fashion, the fact that the self-similarity parameter, H, in the FGNs is required to be constant across the whole series.

It is known that structural break location methods tend to report breaks in simulated long memory series where no breaks exist. We have combined established methods for estimating H and/or d with a computationally fast structural break location method, Atheoretical Regression Trees (ART), to obtain empirical bivariate distributions of H or d and regime length for simulated FGNs.

These bivariate distributions are then compared with a 2649 year warm season temperature reconstruction using data from a stalagtite from Shihua Cave near Beijing, China and with an existing test for fit to an FGN due to Beran. We find the time series is not H-self-similar.

We further compared several other empirically determined bivariate distributions from the simulated data with the Shihua Cave data. In all but one case (mean vs regime length) the Shihua Cave data did not fit the empirical distributions for FGNs.

We can discount the FGNs and FI(d)s as appropriate models for the Shihua Cave data. However, we could not establish statistical primacy between multifractals and multiple regimes of short memory processes.

The implications for the climate change debate are minimal. There seems little doubt the current rising global temperatures are occurring because of increases in greenhouse gas concentrations as a result of human activity. Discounting of the H-self-similar and the FI(d) models for this data leaves the doubters with one less argument to support their case.

1 INTRODUCTION

The British hydrologist H.E. Hurst (1951) published a study of river flows which brought to prominence a phenomena now known variously as long memory, long range dependence, strong dependence, global dependence, or the Hurst phenomena. We shall use the term long memory.

This paper is concerned with the presence of long memory in temperature reconstructions and proxies. Long memory in the Moberg et al. (2005) Northern Hemisphere reconstruction was considered in detail by Mills (2007). Mills tentatively suggested the evidence favoured a shifting trends in temperature model over true long memory. Ballie and Chung (2002) considered long memory in several tree ring series as these series are often used in temperature reconstructions. Ballie and Chung found the series to be very well described by fractional differencing with the exception of the period 1800 to the present in two of their four data sets. Beran (1994) summarized some studies of long memory in instrumental temperature records.

A number of authors have attempted to develop statistical tests to distinguish between true long memory and other types of processes displaying statistical long memory. See for example Smith (2005) and Teverovsky and Taqqu (1999). These and other tests appear to be soundly based in theory but have not found wide application.

The use of structural break detection and location methods are regarded as problematic because they tend to find breaks in fractional Gaussian noises (FGNs) and fractionally integrated processes (FI(d)) (both defined in Section (2) below) even though the data generating process is uniform throughout. For example, Wright (1998) proved that when the standard cumulative summation (CUSUM) test (Brown et al. (1975)) for detecting structural breaks is applied to long memory series the probability of finding a break converges to one with increasing series length. Thus structural break location methods have generally been overlooked when attempting to distinguish between true long-memory and non-stationary means, of whatever type, in real data sets.

If a series is generated by a true long memory process, the use of a structural break location method to divide the series into a number of subseries of differing lengths should only yield subsamples of a single population. If, in fact, the series contains structural breaks which can be located by changes in the mean, using a structural break location method will, instead, divide the series into a number of subpopulations. In the former case our *a priori* expectation is that the subsamples will have the same statistical properties as the full series. In the latter the data generating process has one or more discontinuities and so statistical properties other than the mean way well have changed at the same time.

Despite this risk of model misspecification we could find no empirical study of the statistical properties of the “regimes”, (that is, the sections of the series between the reported breaks) in simulated FGNs or FI(d) series of finite sample size when they were incorrectly analyzed by applying structural break location methods to them. Cappelli et al. (2007) introduced ART , a computationally very fast structural break method, which has allowed large scale simulation studies, such as this one, to be conducted. These would have been computationally impractical with established techniques such as that described by Bai and Perron (1998, 2003).

2 MODELS

A number of models have been proposed to account for the extraordinary persistence of the correlations across time found in long memory series. There are two common sets of models applied across long-memory series from diverse fields. One set are true long memory models, in particular, the FGNs and FI(d) processes.

The other set are models with a non-stationary mean. For simplicity the types of non-stationary mean models studied are ones in which the time series can be broken in a series of “regimes” within which it is a reasonable assumption that the mean is stationary. Some examples are structural break and Markov switching models.

2.1 Fractional Gaussian Noises and Fractionally Integrated Series

Mandelbrot and Ness (1968) introduced FGNs to applied statistics as the stationary increments of a Gaussian H-self-similar stochastic process.

Definition 1 A real-valued stochastic process $\{Z(t)\}_{t\in\mathcal{R}}$ is self-similar with index $H > 0$ if, for any $a > 0$,

$$\{Z(at)\}_{t\in\mathcal{R}} =_d \{a^H Z(t)\}_{t\in\mathcal{R}}$$

where $=_d$ denotes equality of the finite dimensional distributions. H is also known as the Hurst parameter.

Definition 2 A real-valued process $Z = \{Z(t)\}_{t\in\mathcal{R}}$ has stationary increments if, for all $h \in \mathcal{R}$

$$\{Z(t+h) - Z(h)\}_{t\in\mathcal{R}} =_d \{Z(t) - Z(0)\}_{t\in\mathcal{R}}.$$

It is important to note in Definition 1 that H is constant for the whole series and hence for all subseries of an H-self-similar process. As a parameter only has meaning in the context of a model, if H varies over time then the process is, by definition, not H-self-similar.

FGNs are a continuous time process while Fractionally Integrated series (FI(d)) series introduced independantly by Granger and Joyeux (1980), and Hosking (1981) are their discrete time counter-parts.

FI(d)s are a generalization of the "integration" part of the Box-Jenkins ARIMA (p,d,q) (Autoregressive Integrated Moving Average) models to non-integer values of the integration parameter, d. Denoting by B the backshift operator, the operator $(1-B)^d$ can be expanded as a Maclaurin series into an infinite order AR representation

$$(1-B)^d X_t = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} X_{t-k} \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$. The operator in Equation (1) can also be inverted and written in an infinite order MA representation.

ARIMA models with non-integer d are known as Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. The AR(p) and MA(q) parameters in ARFIMA models may be used to model any additional short-range dependence present in the series. Both FGNs and FI(d)s have been extensively studied. See the volumes by Beran (1994), Doukhan et al. (2003), and Embrechts and Maejima (2002) and the references therein.

A variant of these models are the so-called multifractals in which the value of H or d is allowed to vary with time.

2.2 Non-Stationary Mean Model

Klemes (1974) argued that statistical long memory in hydrological time series was the result of non-stationarity in the mean. Klemes pointed out that the assumption of stationarity was often made to facilitate mathematical analysis of the data rather than being based on knowledge of the underlying physical mechanism(s) driving the data generating process. The types of non-stationary mean models which have been studied in any detail typically have stochastic shifts in the mean about some long term average.

A common non-stationary mean model is a series which has structural breaks in the mean. We define the structural break model as follows:

$$\mu_{y_t} = \sum_{i=1}^{p} I_{t_{i-1} \le t < t_i} \mu_i \quad (2)$$

where μ_{y_t} is the mean of the time series, $I_{t \in S}$ is an indicator variable which is 1 only if $t \in S$, t is the time, t_i, $i = 1, \ldots, p$, the breakpoint and μ_i is the mean of the regime i. In this case, a regime is defined as the period between breakpoints.

It is important to note that (2) is just a way to represent a sequence of different models (i.e. models subjected a structural breaks). However, this model only deals with breaks in mean. Given a true break each regime must be modeled separately. This will be important in what follows.

3 METHOD

In Section (4) we give details of the data sets for which we had obtained an estimate of H and d. We simulated by computer up to 26,000 FGN and FI(d) series for each length and value of H and d as estimated for our example data sets. We broke these simulated long memory series into "regimes" using Atheoretical Regression Trees (ART).

The standard deviation was standardized so the series standard deviation was one in all cases. For each "regime" we estimated the length, mean, standard deviation, skewness, kurtosis, normality by the Jarque-Bera test, H using the Whittle estimator. In addition, for the whole series we estimated H, the goodness-of-fit to a long memory process by the test of Beran (1992), the number of breaks detected by ART and the CUSUM range.

We obtained empirical (usually bivariate) distributions of the above quantities (e.g. regime length against standard deviation). We then compared the (usually bivariate) distributions obtained from the simulated series with the real data set to see if the real data set also resembled incorrectly analysed FGNs or FI(d) processes.

For the real data sets the whole series, the regimes and sometimes aggregations of the regimes discovered by ART were subjected to the Beran (1992) goodness-of-fit test for time series with long-range dependence. In this test the null hypothesis is that the series has a spectral density of the form

$$f(\lambda) = c_f |\lambda|^{-\gamma}; 0 < \gamma < 1 \quad (3)$$

where c_f is a constant and $-\gamma = 1 - 2H, H \in (1/2, 1)$. We applied the Beran test using functions implemented in the `R` package `longmemo`.

More formally, assume $Z^{(1,T)} = \{Z_t\}_{t=1}^T$ is a realization of a FGN and $B = \{t_1, t_2, \ldots, t_p\}$ the set of breakpoints identified by ART. The series is divided into $p+1$ sub-series or "regimes". Denote any sub-series i as $Z^{(t_{i-1},t_i)}$, $i = 1, 2, \ldots, p+1$ with $t_0 = 1$ and $t_{p+1} = T$. Then, define $L = \{l_1, \ldots, l_{p+1}\}$ as the sets of lengths of the "regimes".

For each $l_i \in L$ we estimated the various statistical parameters above. We illustrate our method for the H parameter. We obtain a set of estimates $h = \{H_1, H_2, \ldots, H_{n+1}\}$ for the "regimes". To evaluate the hypothesis that the real data sets are incorrectly analysed FGNs we test $P[(l_i, H_i) \in I_\alpha] < (1-\alpha)$ where P is a probability measure and I_α is the α-confidence set (equivalently, we check if $(l_i, H_i) \in I_\alpha$). This test is carried out by simulation as described below:

1. Simulate N true FGN(H) series each with T observations;
2. For each series calculate the sets L and h;
3. Estimate the empirical distribution and the confidence set I_α.
4. Verify if $(l_i, H_i) \in I_\alpha$ for the real data.

As the estimator we used for H exhibits bias in short sub-series which is dependent on the value of H, it is preferable to evaluate the hypothesis graphically (e.g. verify if the point (l_i, H_i) is inside the region defined by I_α). The structural break method we used, ART, breaks the series into regimes based on local changes in the levels. As indicated above there is no *a priori* reason to suspect that any other statistical property of the regimes should change with the level in an H-self-similar series.

4 THE DATA

The data we use is a warm season average temperature reconstruction by Tan et al. (2003) based on an analysis of a stalagtite from the Shihua Cave near Beijing, China. A temperature reconstruction is available for the period 665 BC to 1985 AD, giving a total of 2649 annual observations.

Figure (1) presents a plot of the reconstruction together with the break points determined by ART. Figure (2) presents the autocorrelation function for the Shihua data. The correlations decay at an exceptionally slow rate which is typical of long memory series. Figure (3) presents a smoothed periodogram for the Shihua data. The basic shape of the spectrum is known as red noise and is typical of long memory series.

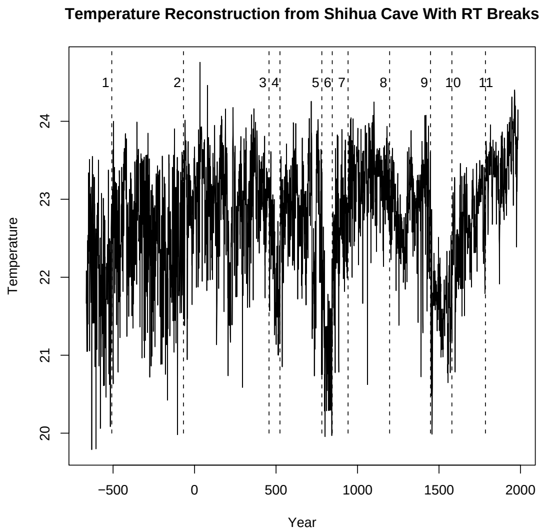

Figure 1. Shihua Cave temperature reconstruction and structural breaks located by the regression tree.

5 RESULTS

Table (1) presents the results of the Beran (1992) goodness-of-fit test for long memory time series. These results should be consider in conjunction with Figure (4). The first column gives the period of the reconstruction being considered. The second column gives the H estimate as returned by the Whittle estimator. The three and four columns gives the p-value of the Beran (1992) test using the value of H estimated from series and the regime respectively.

It is clear that on the basis of this test there is no period in which the null hypothesis of a FGN with H=0.838 is not accepted. Thus we would be lead to believe the long range dependence properties of the series is adequately modeled by a single value of H.

Figure (4) presents the results for the H estimates using the graphical method outlined above. The dots in the graph represent approximately five percent of the simulated data and are included to give a visual representation of the bivariate distribution. The empirical 95% and 99% confidence intervals, represented by the solid and dashed lines respectively, were determined by analysing the results from 26,000 simulated series. The "S" symbols are the Shihua temperature reconstruction regimes' estimated H values. As can be seen, five of the 12 H estimates for the regimes lie outside the empirical 95% confidence interval. Four of these five H values are outside the empirical 99% confidence interval.

In contradiction to Table (1) the evidence here is that this series is not H-self-similar in the sense of the above definition as the value of H does appear to vary

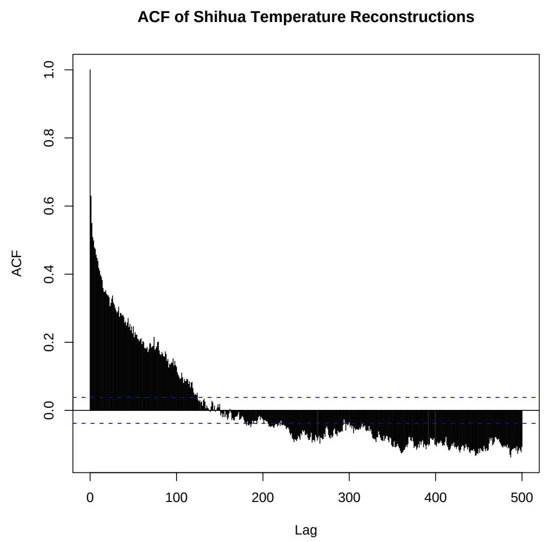

Figure 2. The autocorrelation function for the Shihua Cave temperature reconstruction.

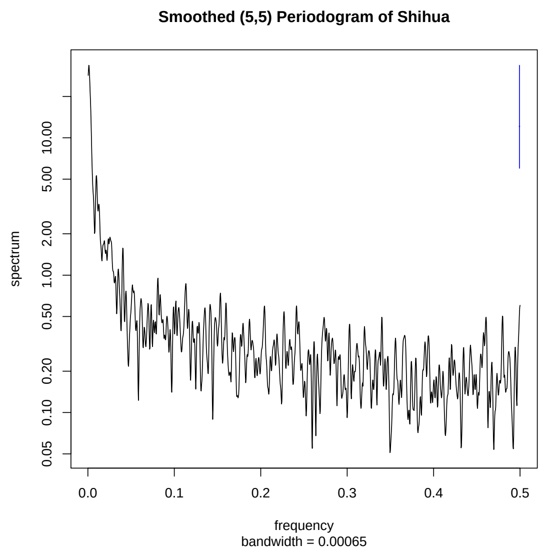

Figure 3. The smoothed periodogram for the Shihua Cave temperature reconstruction.

Period	H Est.	H=0.84	$H = H(t)$
1-2549	0.84	0.38	-
1-158	0.63	0.39	0.85
159-598	0.68	0.24	0.65
599-1123	0.81	0.56	0.49
1124-1190	0.80	0.57	0.49
1191-1447	0.82	0.80	0.80
1448-1511	0.69	0.81	0.90
1512-1608	0.67	0.46	0.65
1609-1863	0.66	0.64	0.76
1864-2114	0.78	0.88	0.88
2115-2245	0.87	0.78	0.80
2246-2451	0.84	0.91	0.91
2452-2549	0.94	0.10	0.26

Table 1. H Estimates and P-values for the Beran (1992) goodness of fit test for the Shihua Cave series and regimes.

with time. We suggest the graphical method is more sensitive to changes in H than the Beran (1992) test.

Figure (5) presents the results for the means of the regimes of both the simulated and real series. As can be seen the means of the regimes in the Shihua data had be adequately modeled by an FGN. Figure (6) plots the number of breaks reported by the regression tree against the CUSUM range. There are two measures used to detect structural breaks. As can be seen the Shihua data has an unusually high number of breaks. Only 16 data points of the 1000 simulations have an equal or higher number of reported breaks. Thus the Shihua Cave data is extreme at approximately the 0.02 level.

Figure (7) presents the results for the standard deviations. As can be seen fully half of the data points for the Shihua data lie below the empirically determined 95% confidence interval. Thus the data within the regimes are more homogeneous than we would expect if the data were generated by a uniform FGN throughout. This suggests we are dealing with distinct sub-populations rather than simply sub-samples of a single uniform population.

The results for skewness and kurtosis are not presented here for reasons of space but are available on request from the authors. For skewness two values are above the upper 95% confidence interval and one s below the lower 99% confidence interval. For the kurtosis one regime lies below the 95% empirical confidence interval and one well above the upper 99% interval.

6 DISCUSSION

There are a number of reasons why one would suspect that any long memory observed in temperature reconstruction time series would be spurious. There are a number of known cyclic influences on climate such as a several solar cycles and atmospheric oscillations such as the El Niño Southern Oscillation. Each of these have variable periods and may thus be difficult to discover in time series with traditional tools such as the periodogram. Theoretical work by Bhattacharya et al. (1983) and others showed that some estimators of H will report long memory when the data contains a small trend. In the long term the three Milankovich cycles are all of sufficient length that over short periods of time they would appear as a trend.

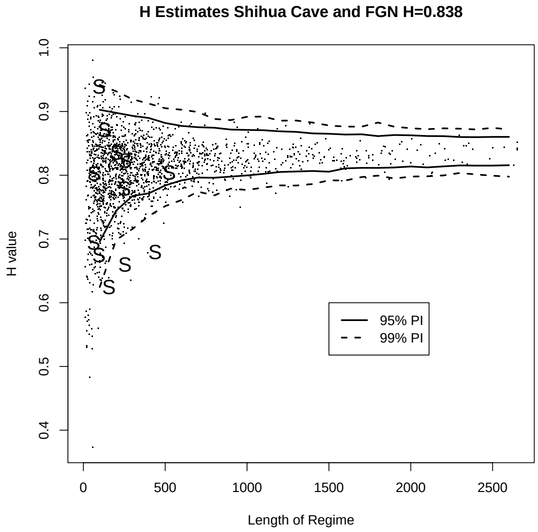

Figure 4. H estimates for Shihua Cave temperature reconstruction and simulated FGNs with H=0.838.

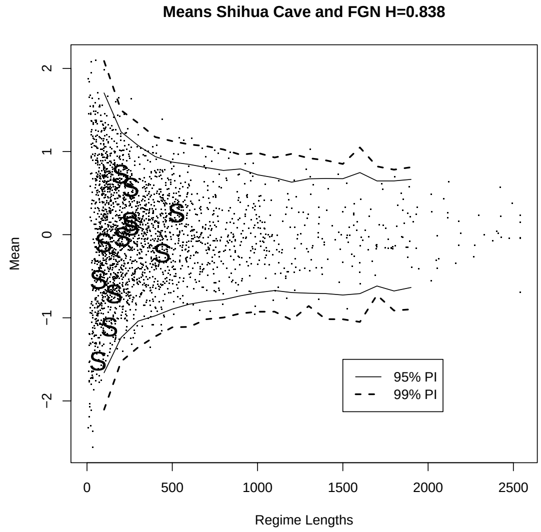

Figure 5. Estimates of regime means for Shihua Cave temperature reconstruction and simulated FGNs with H=0.838.

The particular hypothesis examined here is whether the time series exhibits H-self-similar behaviour as claimed in past literature. The evidence in Figure (4) is clearly that the time series is not H-self-similar.

As a parameter only has meaning in the context of a model, if H is allowed to vary with time we are led to consider the so-called multifractal models. These models appear perfectly adequate. The evidence from Table (1) is that each regime has a good fit to an FGN, the lowest p-value being 0.26. Indeed, seven of the 12 regimes show an improved fit by using the within regime H value rather than the series H value. To further discriminate between FGN and multiple regimes of short memory processes is difficult because there is inadequate data to make meaningful comparisons.

The alternative is to consider the types of models proposed by Klemes (1974) in which the mean is non-stationary. By considering a time-varying H we have already conceded that the series is non-stationary in H. To demonstrate adequate levels of statistical significance that the series is non-stationary in the mean cannot be done directly. FGNs are stationary models and Figure (5) shows that FGNs can indeed model the changes in the mean in a perfectly satisfactory manner.

However, by examining higher moments the FGN model fails to adequately account for the data. In particular the standard deviation evidence (Figure 7) is problematic. It indicates the data within the regimes is often more homogeneous that we would expect with an FGN. If there is a structural break at the points discovered by the regression tree we currently have no way to predict the properties of the new regime. We must estimate them from the data.

7 CONCLUSION

The evidence is clear that the Shihua Cave temperature reconstruction is not an H-self-similar time series. It is also clear that some type regime shifting model is appropriate for this data. But that is the extent of what we can say from the data. We cannot, at this stage, discriminate between shifting regimes of short memory process or a multifractal model.

The implications for the climate change debate are minimal. There seems little doubt the current rising global temperatures are occurring because of increases in greenhouse gas concentrations as a result of human activity. Discounting of the H-self-similar model for this data leaves the doubters with one less argument to support their case.

References

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18:1–22.

Ballie, R. T. and Chung, S.-K. (2002). Modeling and forecasting from trend-stationary long memory models with applications to climatology. *International Journal of Forecasting*, 18:215–226.

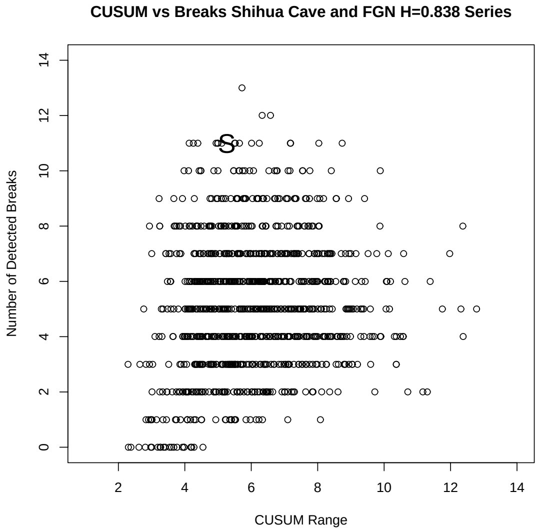

Figure 6. CUSUM range against reported breaks in the Shihua Cave temperature reconstruction and simulated FGNs with H=0.838.

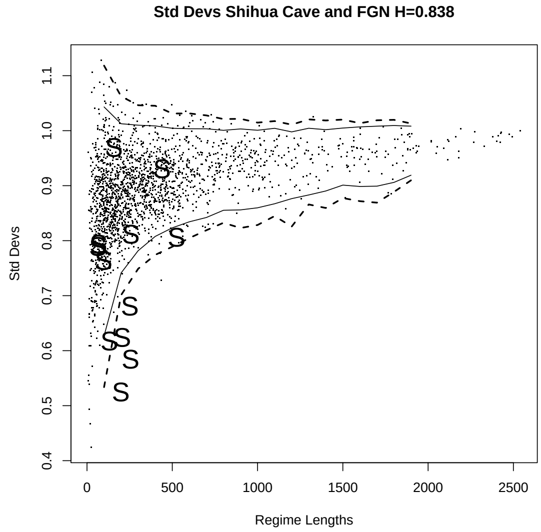

Figure 7. Estimates of regime standard deviations for Shihua Cave temperature reconstruction and simulated FGNs with H=0.838.

Beran, J. (1992). A Goodness-of-Fit Test for Time Series with Long Range Dependence. *Journal of the Royal Statistical Society B*, 54(3):749–760.

Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman & Hall/CRC Press.

Bhattacharya, R., Gupta, V., and Waymire, E. (1983). The Hurst effect under trends. *Journal of Applied Probability*, 20:649–662.

Brown, R. L., Durbin, J., and Evans, J. (1975). Techniques for Testing the Constancy of Regression Relationships over time. *Journal of the Royal Statistical Society Series B*, 37(2):149–192.

Cappelli, C., Penny, R. N., Rea, W. S., and Reale, M. (2007). Detecting multiple mean breaks at unknown points with Atheoretical Regression Trees. *Mathematics and Computers in Simulation*. forth coming.

Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). *Theory and Applications of Long-Range Dependence*. Birkhaüser.

Embrechts, P. and Maejima, M. (2002). *Selfsimilar Processes*. Princeton University Press.

Granger, C. W. J. and Joyeux, R. (1980). An Introduction to Long-range Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, 1:15–30.

Hosking, J. R. M. (1981). Fractional Differencing. *Biometrika*, 68(1):165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808.

Klemes, V. (1974). The Hurst phenomenon - a puzzle? *Water Resources Research*, 10:675–688.

Mandelbrot, B. B. and Ness, J. W. V. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437.

Mills, T. (2007). Time series modelling of two millenia of northern hemispere temperatures: long memory or shifting trends? *Journal of Royal Statistical Society Series A*, 170:83–94.

Moberg, A., Sonechkin, M. M., Holmgren, K., Datsenko, N. M., and Karlen, W. (2005). Highly variable Northern Hemispere temperatures reconstructed from low- and high-resolution proxy data. *Nature*, 433:613–617.

Smith, A. D. (2005). Level Shift and the Illusion of Long Memory in Economic Time Series. *Journal of Business and Economic Statistics*, 23(3):321–335.

Tan, M., Liu, T. S., Hou, J., Qin, X., Zhang, H., and Li, T. (2003). Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophysical Research Letters*, 30:1617.

Teverovsky, V. and Taqqu, M. (1999). Testing for Long-Range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-Type Estimator. *Journal of Time Series Analysis*, 18:279–304.

Wright, J. H. (1998). Testing for a Structural Break at Unknown Date with Long-Memory Disturbances. *Journal of Time Series Analysis*, 19(3):369–376.