Long memory in temperature reconstructions

William Rea, Marco Reale & Jennifer Brown

Climatic Change
An Interdisciplinary, International Journal Devoted to the Description, Causes and Implications of Climatic Change

ISSN 0165-0009
Volume 107
Combined 3-4

Climatic Change (2011)
107:247-265
DOI 10.1007/s10584-011-0068-y


Climatic Change

An Interdisciplinary, International Journal Devoted to the Description, Causes and Implications of Climatic Change

Editor: STEPHEN H. SCHNEIDER

Volume 69 – No. 1 – March 2005

SPECIAL ISSUE:
CLIMATE CHANGE IMPACTS FOR THE CONTERMINOUS USA
An Integrated Assessment

Guest Editors:
Norman J. Rosenberg
James A. Edmonds

CODEN CLCHDX

ISSN 0165-0009

Springer

Climatic Change (2011) 107:247–265
DOI 10.1007/s10584-011-0068-y

Long memory in temperature reconstructions

William Rea · Marco Reale · Jennifer Brown

Received: 24 June 2008 / Accepted: 16 July 2010 / Published online: 17 May 2011

Abstract Ever since H. E. Hurst brought the concept of long memory time series to prominence in his study of river flows the origins of the so-called Hurst phenomena have remained elusive. Two sets of competing models have been proposed. The fractional Gaussian noises and their discrete time counter-part, the fractionally integrated processes, possess genuine long memory in the sense that the present state of a system has a temporal dependence on all past states. The alternative to these genuine long memory models are models which are non-stationary in the mean but for physical reasons are constrained to lie in a bounded range, hence on visual inspection appear to be stationary. In these models the long memory is merely an artifact of the method of analysis. There are now a growing number of millenial scale temperature reconstructions available. In this paper we present a new way of looking at long memory in these reconstructions and proxies, which gives support to them being described by the non-stationary models. The implications for climatic change are that the temperature time series are not mean reverting. There is no evidence to support the idea that the observed rise in global temperatures are a natural fluctuation which will reverse in the near future.

1 Introduction

Since the end of the last ice age the earth's climate has enjoyed a period of relative stability. The earth is now in a period of rising global temperatures. A number of authors have considered the stochastic properties of univariate 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 often exhibit the property of statistical long memory.

W. Rea · M. Reale (✉) · J. Brown
University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
e-mail: marco.reale@canterbury.ac.nz

Long memory time series were brought to prominence by the hydrologist Hurst (1951). Long memory is also known as long range dependence, strong dependence, global dependence, or the Hurst phenomena. We shall use the term long memory.

There is no definitive test for the presence of long memory in a time series. Long memory tends to be identified by a subjective assessment of the time series plot, the autocorrelation and partial autocorrelation functions (ACF and PACF respectively) and a spectral estimate. Long memory time series typically have the following three qualitative features visible in the series plot (Beran 1994);

1. There are relatively long periods where the observations tend to be either above or below the global mean.
2. When short subseries are examined there appear to be local cycles or trends, but no discernible persisting trend or cycle over the whole series.
3. Overall, the series looks stationary.

When examining the common diagnostic tools of the ACF, PACF, and spectral estimate, in long memory series we typically find the following;

1. The ACF decays at an exceptionally slow rate. Often there are statistically significant correlations at lags of hundreds and sometimes thousands of data points apart.
2. The PACF decays more quickly than the ACF and a high order autoregressive (AR) model is required to adequately model the data.
3. The spectral estimate shows no statistically significant periodicities, but rather an exponential increase in power with decreasing frequency. This type of spectrum is often called red noise.

Fractional Gaussian Noises and Fractionally Integrated series (FGNs and FI(d) respectively, defined in Section 2 below) have been widely used as models for long memory series. FGNs and FI(d) series have three attractive properties; they are (1) stationary, (2) linear and (3) only a single parameter is required to model the long memory. In the literature they have been applied to a number of different temperature time series, climatic proxies and closely related atmospheric and oceanic processes. Some examples are the study by Mandelbrot and Wallis (1969) who reported long memory to be present in a wide variety of geophysical time series including instrumental temperature records. Bloomfield (1992) and Bloomfield and Nychka (1992) considered several time series models including FI(d) series to determine whether the observed global warming in instrumental records could be accounted for by natural fluctuations. Bloomfield and Nychka (1992) concluded the observed rate of temperature rise could not be accounted for by a stationary FI(d) series. Gil-Alana (2005) considered using FI(d) series as models for Northern Hemisphere temperatures. Mills (2007) considered in detail long memory in the Moberg et al. (2005) Northern Hemisphere temperature reconstruction. Mills tentatively suggested the evidence favoured a shifting trends in temperature model over true long memory.

Stephenson et al. (2000) considered FI(d) series as a possible model for the North Atlantic Oscillation. Baillie and Chung (2002) considered long memory in several tree ring series which are often used in temperature reconstructions. They reported the series they examined to be very well described by FI(d) series with the exception of the period 1800 to the present in two of their four data sets. Overland et al.

(2006) considered three models of the North Pacific Ocean sea surface temperatures; autoregressive order one (AR(1)), FI(d) and a square wave oscillator. Overland and his co-authors could not establish the statistical primacy of any of the three models.

Some of these authors were aware that a series with a mean which is non-stationary but which is constrained for physical reasons to lie in a bounded range may be indistinguishable from a long memory time series. Those authors who tried to distinguish between a non-stationary mean and true long memory found the task difficult or impossible with the statistical tools they had available. Sibbertsen (2004) stated that this difficulty arose because the finite sample properties of both long memory series and series with structural breaks (e.g. series with shifts in the mean) are similar and so standard methodologies often fail.

There is a significant body of literature which has applied other models to temperature time series. It is not the purpose of this paper to examine them in detail, see for example Allen and Smith (1994), Mann and Lees (1996), Gipp (2001), and Smith et al. (2003).

A number of authors have developed statistical tests or procedures to identify true long memory series. Among these, Beran and Terrin (1996, 1999) developed a test for the constancy of the self-similarity parameter H (defined in Section 2.1 below). Teverovsky and Taqqu (1999) introduced a procedure which involved the use of the the aggregated variance and differenced variance estimators to distinguish between long memory on the one hand and shifting means or deterministic trends on the other. Smith (2005) modified the widely used GPH (Geweke and Porter-Hudak 1983) estimator to discriminate between infrequent level shifts and long memory. Ohanissian et al. (2008) developed a test of self-similarity based on different levels of aggregation of the series. All of these tests appear to be soundly based in theory but have not yet found wide application.

An alternative approach to these tests are the use of structural break detection and location methods. These methodologies have largely been overlooked when attempting to distinguish between true long-memory and non-stationary series because they tend to report breaks in true long memory series where no breaks exist.

There are a growing number of millennial scale temperature reconstructions available and in this paper we examine in some detail whether long memory models are in fact appropriate for this type of data from an empirical viewpoint. In essence – is the data described well by the model? Along with several established methodologies we present a new computational method based on obtaining through simulation a bivariate distribution of the long memory parameters H or d (defined in Section 2 below) with regime length and comparing this distribution with the parameter values obtained from the temperature reconstructions.

The remainder of the paper is organized as follows. Section 2 sets out the competing models. Section 3 discusses the methods used in this paper. Section 4 describes the data. Section 5 presents the results of the investigation. Section 6 contains the discussion, and Section 7 gives final concluding remarks.

2 Models

A small number of models have been proposed to account both for the exceptionally slow decay of the correlations with time and the red noise spectrum found in long memory series.

2.1 Fractional Gaussian noises and fractionally integrated series

Mandelbrot and van Ness (1968) introduced Fractional Gaussian Noises (FGNs) into applied statistics, in part, in an attempt to model the famous Nile River annual low flow data. They are 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 $a > 0$,

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

where $=_d$ denotes equality of the finite dimensional distributions and $\mathcal{R}$ denotes the real numbers. 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 $h \in \mathcal{R}$

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

H-self-similar series with $0.5 < H < 1$ exhibit long memory. When $H \geq 1.0$ the series are non-stationary but mean reverting, while when $H < 0.5$ the series are anti-persistent.

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.

Independently, Granger and Joyeux (1980) and Hosking (1981) obtained the discrete time counter-parts of FGNs by generalizing 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 the backshift operator by B, the operator $(1-B)^{-d}$ can be expanded as a Maclaurin series into an infinite order MA representation

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

where ϵ_t are the disturbances at time t and $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ is the well-known Gamma function. The operator in Eq. 1 can also be inverted and written in an infinite order AR representation. In these models stationary long memory is exhibited in the range $0 < d < 0.5$.

ARIMA(p,d,q) models with non-integer d are known as Autoregressive Fractionally Integrated Moving Average (ARFIMA) models. Sometimes they are referred as Fractional ARIMA models (FARIMA). The ARMA(p,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), Embrechts and Maejima (2002), and Palma (2007) and the collections of Doukhan et al. (2003) and Robinson (2003) and the references therein.

2.2 Constrained non-stationary models

Klemes (1974) argued that statistical long memory in hydrological time series was the result of analyzing non-stationarity series with statistical tools which assume stationarity. Klemes pointed out that the assumption of stationarity was often made to facilitate mathematical analysis of the data rather than being based on a knowledge of the underlying physical mechanism(s) driving the data generating process. Thus long memory was a simply a meaningless statistical artifact arising from incorrect analysis. His arguments are easily seen to be applicable to reconstructed temperature time series such as we examine here.

Many times series which display long memory are constrained for physical reasons to lie in a bounded range and thus on visual inspection appear to be stationary.

The types of non-stationary models which have been studied in detail typically have stochastic shifts in the mean about some long term average. These mechanisms produce models with a number of "regimes" within which stationarity is assumed.

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

$$\mu_{y_t} = \sum_{i=1}^{p+1} I_{t\in(t_{(i-1)},t_i]}\mu_i \tag{2}$$

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

It is important to note that Eq. 2 represents 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

To motivate our new methodology consider a series generated by true long memory model such as an FGN. If we apply a structural break location method to the series, and it reports one or more breaks, we only divide that series into a number of subseries of differing lengths. This should only yield subsamples of a single population. If, on the other hand, the series contains true structural breaks or some other more general type of constrained non-stationarity which may be located by finding changes in the mean, then the use of a structural break location method will, instead, divide the series into a number of subpopulations. In the former case our expectation is that the subsamples will have the same statistical properties as the full series. In the latter case the data generating process may have at least one discontinuity and so the statistical properties of the individual regimes should differ.

Thus we propose that it should be possible to distinguish between true long memory and constrained non-stationary series on the basis of examining the statistical properties of any "regime" reported by a structural break location method.

In Section 4 we give details of the six data sets analyzed. We assumed these data sets had true long memory and estimated the long memory parameters H and d for each of them. We simulated up to 26,000 FGN series for each length and value of H or d as estimated for our example data sets. We then subjected these simulated long memory series to a computationally fast structural break detection and location method, Atheoretical Regression Trees (ART) (Cappelli et al. 2008). The subseries between identified breaks are referred to as "regimes".

For each "regime" we estimated the length, mean, standard deviation, skewness, kurtosis, normality by the Jarque-Bera test using functions in the base packages in `R` (R Development Core Team 2005) and `fBasics` of Wuertz (2005a), and H using the Whittle estimator as implemented in the `R` package `fSeries` (Wuertz 2005b) or d using the estimator of Haslett and Raftery (1989) as implemented in the `fracdiff` package (Fraley et al. 2006) in `R`. In addition, for the whole series we estimated H, the goodness-of-fit to a long memory process by the test of Beran (1992). FGNs and FI(d) series with long-range dependence have a distinctive spectral density at the origin of the form

$$f(\omega) \sim b\,|\omega|^{1-2H} \tag{3}$$

where b is a constant and $H \in (1/2, 1)$. By building on the work of Milhoj (1981), Beran (1992) introduced a goodness-of-fit test in the frequency domain for Gaussian models with spectra of the form in Eq. 3. It should be noted that Beran test was shown by Deo and Chen (2000) to use an asymptotically incorrect distribution which leads, in practice, to a small amount of over rejection of the null hypothesis. We applied the Beran test using functions implemented in the `R` package `longmemo` of Beran et al. (2006).

ART was implemented using the package `tree` (Ripley 2005). The CUSUM test (Brown et al. 1975) was applied as implemented in `strucchange` (Zeileis et al. 2002). Spectral estimates were obtained with `spectrum` in `stats` within `R`.

We obtained empirical (usually bivariate) distributions of the above quantities (e.g. regime length against standard deviation) for simulated FGNs or FI(d) series. These are the distributions obtained when the null hypothesis of long memory was known to be true. We then subjected the real datasets to the same structural break detection and location method and compared the distributions obtained from the simulated series with the real data set. The idea behind this approach is that if the statistical properties of the real data set differ from the properties of the spurious regimes from the simulated long memory series then the real data could be considered not to be a true long memory series.

More formally, let $Z^{(1,T)} = \{Z_t\}_{t=1}^{T}$ be a realization of a FGN and $B = \{t_1, t_2, \ldots, t_p\}$ the set of (spurious) 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 $l_i \in L$ estimate 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 not self-similar time series 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 interval (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 long memory FGN(H) series each with T observations and H (or d) as estimate for the data series under investigation;
2. For each series calculate the sets L and h;
3. Estimate the empirical distribution and the confidence interval I_α.
4. Verify if $(l_i, H_i) \in I_\alpha$ for the data.

As both the Whittle estimator for H and the Haslett and Raftery (1989) estimator for d exhibit bias in short series which is dependent on the value of H or d, 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 by visual inspection of the bivariate plot). The structural break method we used, ART, breaks the series into regimes based on local changes in the mean. 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.

We used the ten H estimators implemented in the `R` package `fSeries` of Wuertz (2005b). These were the aggregated variance, absolute value method, boxed periodogram, differenced variance, Higuchi (1988), Peng et al. (1994), periodogram, rescaled range (R/S), Whittle and wavelet. Details of the first nine of these estimators and further references to the literature are given in Taqqu et al. (1995), details of the wavelet estimator can be found in Abry and Veitch (1998), Abry et al. (1998), Jensen (1999) and Veitch and Abry (1999). We also used the estimator of Haslett and Raftery (1989) as implemented in `fracdiff` of Fraley et al. (2006). FGN series were simulated with the function `fgnSim` in `fSeries` of Wuertz (2005b). Theoretical ACFs for FGNs were obtained with `fgnTrueacf` as implemented in `fSeries`.

4 Data

In this section we briefly list the six temperature reconstructions considered in this study.

A number of data sets were discarded from the study. Details of the reasons for discarding are available from the authors on request. These included; (i) a 1241 year summer temperature reconstruction based on laminae thickness in sediments from Donard Lake, Baffin Island, Canada by Moore et al. (2001), (ii) a 633 year temperature reconstruction based on grape harvest records in the Burgundy region of France by Chuine et al. (2004) (iii) a 1980 year Northern Hemisphere temperature reconstruction by Moberg et al. (2005), (iv) a 1230 year reconstruction of temperatures in West Greenland from ice core data by Fisher et al. (1996) (v) a 4,000-year summer temperature reconstruction in the the Yamal Peninsula, Siberia from larch tree data by Hantemirov and Shiyatov (2002).

The series considered in detail are as follows.

Colorado A 2247 year temperature reconstruction between 250BC and 1992AD for the Colorado Plateau regions by Salzer and Kipfmueller (2005) based on tree ring data.

Northern Hemisphere Two 1283 year Northern Hemisphere temperature reconstructions between 713 and 1995AD by D'Arrigo et al. (2006).

Shihua A 2650 year warm season temperature reconstruction by Tan et al. (2001) between 665BC and 1985AD based on stalagmite data from the Shihua Cave near Beijing, China.

Tasmania Two 3592 year reconstructions of warm season temperatures in Tasmania, Australia between 1600BC and 1991AD by Cook et al. (2000) based on Tasmanian Huon pine tree rings.

Torneträsk A 1993 year reconstruction of temperatures at Torneträsk, Sweden from tree ring data by Briffa et al. (1992) between 1 and 1993AD. This data set appears to have been updated since the publication of Briffa et al. (1992).

Western USA A 1780 year reconstruction of temperatures in the Western USA by Mann et al. (1998) between 200 and 1980AD based on tree ring data. (Note: The reference to Mann et al. (1998) for this data set was given in Jones and Mann (2004) but this data set seems too long for the reconstruction discussed by Mann et al. (1998).)

5 Results

We present the results in detail for the Shihua Cave reconstruction of Tan et al. (2001), and Torneträsk reconstruction of Briffa et al. (1992). Where practical we state the results of the other series, full details are available on request from the authors.

Table 1 reports the p-values from the Beran (1992) test of goodness-of-fit of a series to an FGN.

In Table 2 are the estimates of the H or d parameter as reported by 11 different estimators for the six temperature time series.

5.1 Shihua cave

When the Shihua Cave temperature time series was analyzed using ART to obtain candidate break points 12 regimes were identified ranging from less than 100 years

Table 1 H estimates using the Whittle estimator and p-values for the Beran (1992) goodness-of-fit test for the six series

Series	H Est.	p-value
Colorado	0.955	0.44
Nth. Hemisphere	0.949	0.50
Shihua	0.838	0.38
Tasmania	0.997	0.009
Torneträsk	0.842	0.17
West USA	0.740	0.61

Table 2 *H* and *d* estimates (*d* estimates have been converted to *H* equivalents) for six temperature reconstructions reported by 11 different estimators

Estimator	Colo	Nth Hem.	Shihua	Tasmania	Torne	W USA
Abs. val.	0.508	0.779	0.895	0.604	0.773	0.672
Agg. var.	0.488	0.724	0.858	0.592	0.740	0.639
Box Per	0.911	0.934	0.884	0.768	0.889	0.701
Dif. var	0.787	1.122	1.084	0.844	1.173	0.954
Higuchi	0.966	0.982	0.967	0.967	0.810	0.631
Peng	0.875	0.973	0.929	0.822	0.926	0.758
Periodogram	0.990	0.977	1.039	0.733	1.015	0.810
R/S	0.895	0.639	0.690	0.809	0.732	0.657
Whittle	0.955	0.949	0.838	0.997	0.842	0.740
Wavelet	0.729	0.998	0.887	0.875	1.014	0.643
Haslett-Raftery	0.998	0.994	0.893	0.999	0.912	0.794

The data column abbreviations are; *Colo* Colorado Plateau, *Nth Hem.* Northern Hemisphere, *Shihua* Shihua Cave, *Torne* Torneträsk, *W USA* Western USA reconstruction of Mann et al. (1998)

to over 450 years in length. The time series plot with the locations of the breaks reported by ART are presented in Fig. 1.

Figure 2 presents an ACF plotted out to 2000 lags overplotted with three theoretical ACFs for the FGNs with *H* estimates as reported by the Higuchi, R/S and Whittle estimators.

Figure 3 presents spectral estimates using 3,3 modified Daniell windows for the full series and three sub-periods of 800 years starting at 148BC, 521AD, and 1186AD respectively.

Figure 4 presents the results for the conditional bivariate *H* estimates given regime length for simulated FGNs with $H = 0.838$. The empirical median, 75%, 95% and

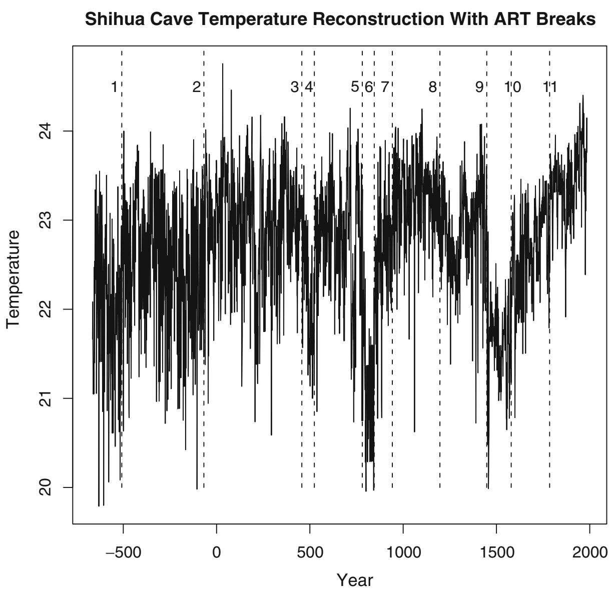
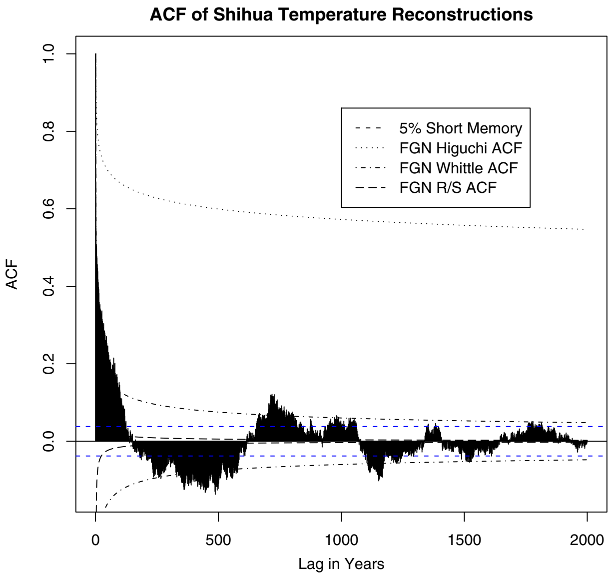


Fig. 1 Time series plot of Shihua reconstruction with breaks reported by ART marked by *dashed vertical lines*

Fig. 2 ACF for Shihua Cave temperature reconstruction plotted to 2000 lags



99% confidence intervals were determined by analyzing the results from 26,000 simulated series. The 12 "S" symbols are the Shihua temperature reconstruction regimes' estimated H values overplotted on the null distribution.

Figure 5 presents the results for the conditional bivariate standard deviation given regime length obtained using the same data as in Fig. 4. The 12 "S" symbols are the 12 Shihua Cave data points overplotted on the null distribution. Results for the skewness and kurtosis are available from the authors on request.

5.2 Torneträsk

When the Torneträsk data was analyzed with ART to obtain candidate break points, 11 regimes were identified ranging from less than 20 years to almost 500 years. The time series plot with the ART breaks marked by dashed vertical lines is presented in Fig. 6.

Figure 7 presents an ACF plotted out to 1500 lags overplotted with three theoretical ACFs for the FGNs with H estimates as reported by the Peng, R/S and Whittle estimators.

Figure 8 presents spectral estimates using 3,3 modified Daniell windows for the full series, and three 800 year subseries starting at 388AD, 778AD, and 1192AD respectively.

Figure 9 presents the results of the conditional bivariate H estimate given regime length for the simulated FGNs. The 11 Torneträsk data points over plotted as "T"s on the null distribution for FGNs with H=0.842.

The results corresponding to Fig. 5 are not presented here as they are not statistically significant but are available on request. Results for the skewness and kurtosis are also available on request.

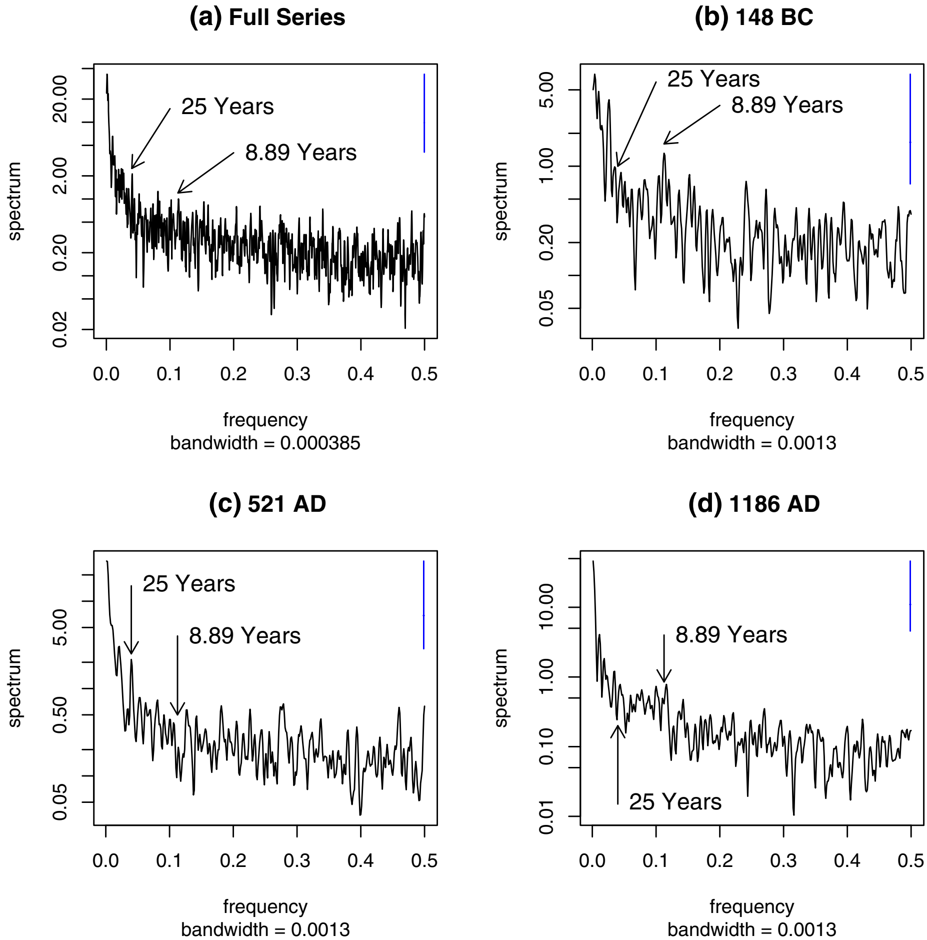


Fig. 3 Spectral estimates for the Shihua Cave temperature reconstruction using 3,3 modified Daniell windows. **a** Full series. **b–d** 800 year periods beginning at 148BC, 521AD and 1186AD respectively

6 Discussion

A parameter only has meaning in the context of a model. For FGNs and FI(d) series, the parameters H and d have a clear mathematical meaning. It is our purpose to examine whether the apparent long memory properties of temperature reconstructions are well described by a single value of H or d.

The time series plots in Figs. 1 and 6 have the qualitative features outlined in the introduction which are typical of long memory processes. Local trends and local cycles are easily spotted but none persist across the whole series. Further, the ACFs and spectral estimates in Figs. 2, 3, 7 and 8 are clearly of the long memory type. When a formal test of goodness-of-fit is made as presented in Table 1 for five of the six data sets the null of long memory spectrum is not rejected.

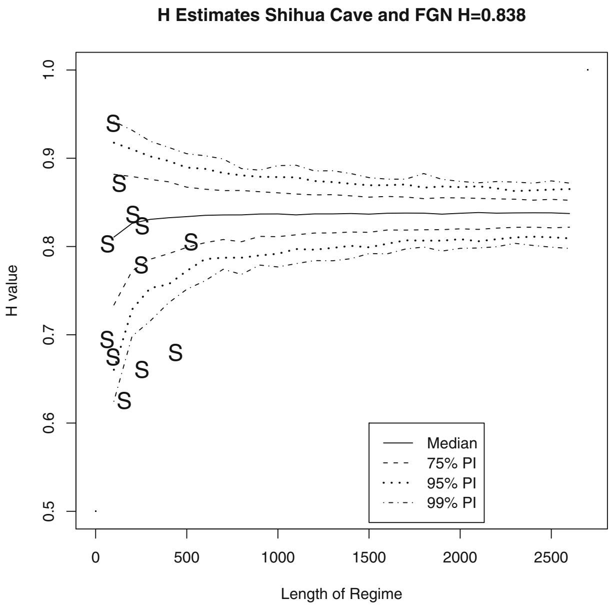


Fig. 4 Conditional bivariate distribution of *H* estimates given regime length for Shihua Cave temperature reconstruction and simulated FGNs with H = 0.838

Thus the initial evidence is strongly in favor of identifying five of these series as FGNs. Particularly in light of the evidence presented in Table 1 it may seem churlish to question their use in modeling these series.

The results in Table 2 give us our first clue that the long memory model FGN does not describe these data sets well. We have applied 11 different estimators of

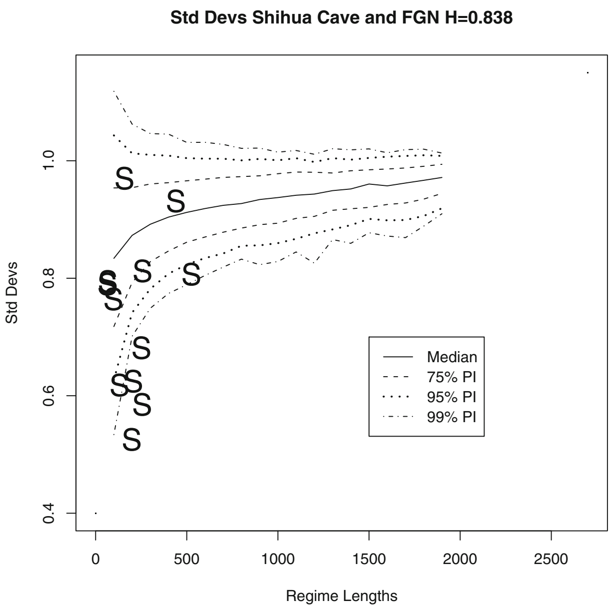
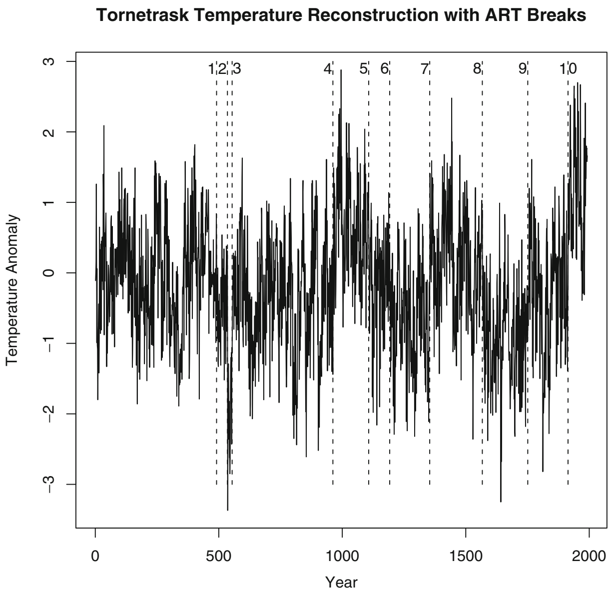


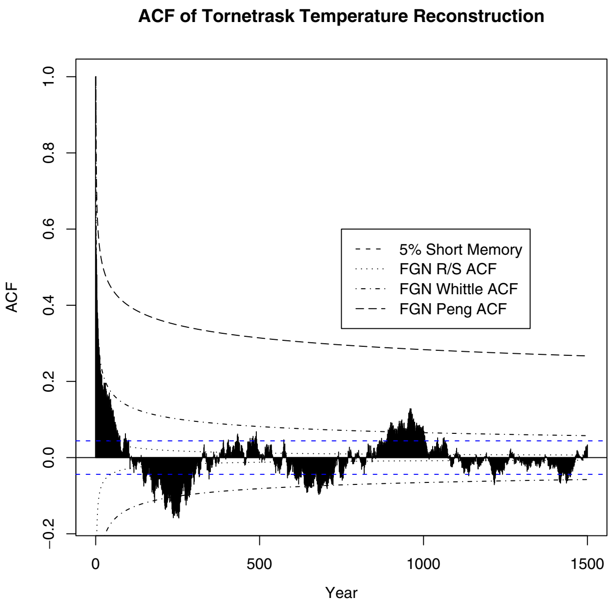
Fig. 5 Estimates of regime standard deviations for Shihua Cave temperature reconstruction and simulated FGNs with H = 0.838

Fig. 6 Temperature plot of the Torneträsk data with breakpoints reported by ART marked by *dashed vertical lines*



H or *d* to the data. All are implemented as estimators of *H* except the Haslett and Raftery (1989) estimator which was implemented as an estimator of *d*. These estimators are validated by either an appeal to some aspect of self-similarity or by an asymptotic analysis of their distributional properties. Some them have been discussed in reasonable detail by Taqqu et al. (1995) and Beran (1994, Ch.4–6). Each estimator

Fig. 7 ACF of the Torneträsk data plotted to a lag of 1500 years overplotted with theoretical ACFs for FGNs with *H* estimates as reported by the Peng, R/S and Whittle estimators. The *horizontal lines* are the 5% significance level under short memory assumptions



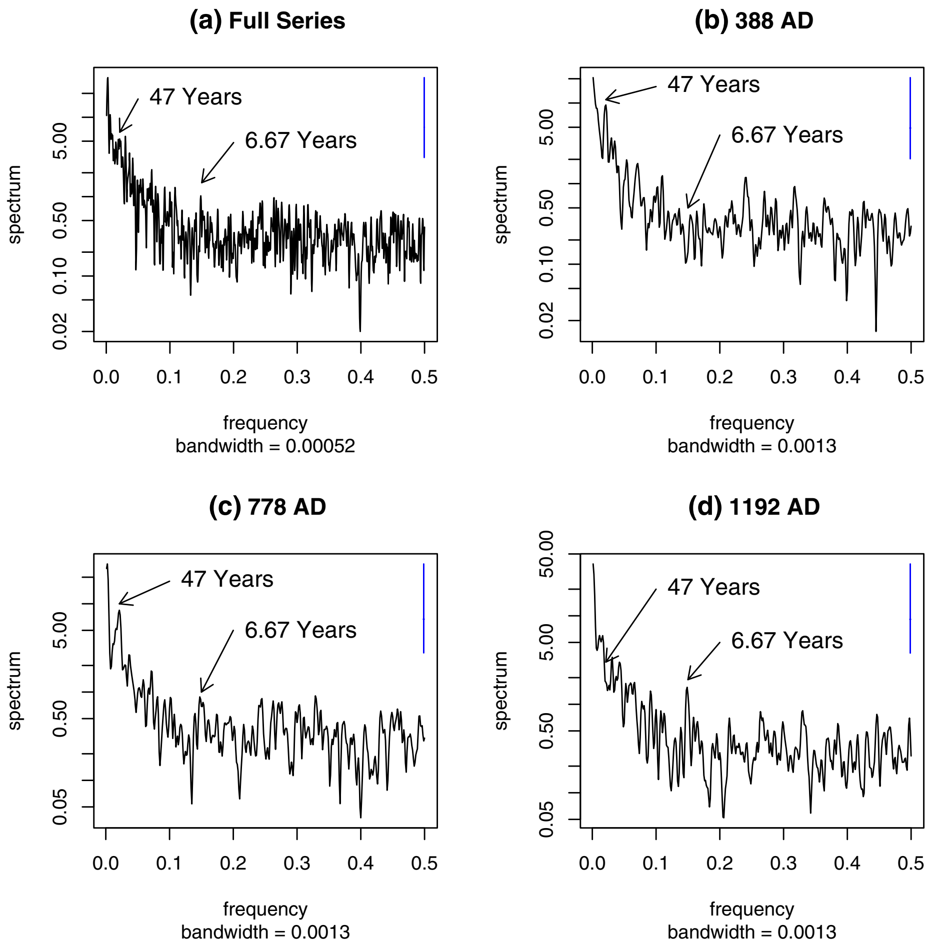


Fig. 8 Spectral estimates for the Torneträsk temperature reconstruction using 3,3 modified Daniell windows. **a** Full series. **b**–**d** 800 year periods beginning at 388AD, 778AD, and 1192AD respectively

is, ultimately, attempting to estimate the same quantity. The wide range of estimates reported is of concern. For example, for the Colorado Plateau data the estimates ran from a low of 0.488 to a high of 0.990. We note that long memory occurs in FGNs with $0.5 < H < 1.0$. Three of the series (Northern Hemisphere, Shihua, and Torneträsk) have one or more H estimates in the non-stationary range ($H \geq 1.0$).

The ACFs of the Shihua and Torneträsk series are plotted out to 2,000 and 1,500 lags in Figs. 2 and 7 respectively. Over plotted on these estimated ACFs are theoretical ACFs for FGNs with three different H estimators as reported by the Higuchi (1988), R/S, and Whittle estimators for the Shihua data and the Peng et al. (1994), R/S, and Whittle estimators for the Torneträsk data. These were selected for reporting "low", "middle" or "high" H estimates for each of the series. It is clear that the theoretical ACFs for some of these estimates do not match the ACFs for the data.

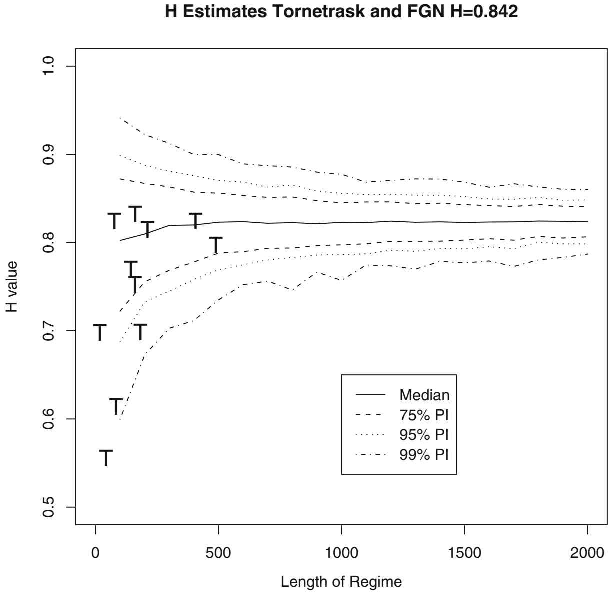


Fig. 9 Conditional bivariate distribution of H estimates given regime length for Torneträsk temperature reconstruction and simulated FGNs with H = 0.842

If the series are not FGNs a simple statistical explanation for these diverging H estimates and the mismatch between theoretical and estimated ACFs would be that the estimators are not estimating anything meaningful. This is essentially the argument put forward by Klemes (1974) but which is still unresolved.

Figures 4 and 9 present the bivariate distributions of H estimate and regime length, using the Whittle estimator to estimate H, over-plotted on the distributions of the null hypothesis obtained by simulation. The null distribution is that which we would obtain if the series were an FGN and H was constant across the series. For the Shihua Cave data it is clear that the 12 data points do not fit this null distribution. Attempting to improve the fit by raising or lowering the H value tested would only result in small improvements in fit. Lowering the tested H to bring the three data points below the lower 99% confidence interval within with lower 95% confidence interval would result in at least one more data point lying above the upper 95% confidence interval. Conversely, lowering the test H value to bring the single data point above the upper 95% confidence interval would result in one more data point lying below the lower 95% confidence interval. Also, we should note that we cannot arbitrarily change H as the model still needs to fit the evidence presented in Fig. 2.

If the parameter H is meaningful in these series we must conclude that it varies with time. Thus the evidence here is that the Shihua Cave series is not H-self-similar in the sense of the above definition because evidence is that the value of H is not constant.

When we examined the bivariate distribution of standard deviation with regime length the FGN model also failed to adequately account for the data. The evidence of Fig. 5 was problematic. Five of the 12 regimes in the Shihua data have standard deviations below the empirically estimated 95% confidence interval. Because the standard deviation is just a measure of the variability in the data this evidence indicates the data within the regimes was often more homogeneous that we would expect with an FGN. This suggests we are dealing with distinct sub-populations rather than simply sub-samples of a single uniform population.

For the Torneträsk data in Fig. 9 it would be possible to improve the fit by lowering the tested H value. The best fit which could be obtained this way would either leave one data point above the upper 95% confidence interval and one data point below the lower 95% confidence interval or two data points below the lower 95% confidence interval. However, lowering the H value would then be problematic for the data in Fig. 7 in much the same way as it was for the Shihua Cave data.

A question we should consider is how to account for the excellent fits to an FGN which appeared in Table 1 provided the value of H was well chosen.

In this type of data we could expect to see some periodic components. At very long time scales there are systematic changes to the earth's orbital characteristics known as the Milankovitch cycles. At shorter time scales there are a number of known quasi-cyclic influences on climate such as a number of solar cycles (e.g. the Hale, Schwabe, Gleisberg, Suess, and Hallstatt cycles), and atmospheric oscillations such as the El Niño, Arctic and North Atlantic Oscillations, and systematic sea-surface temperature changes such as the Pacific Decadal Oscillation. With the exception of the Milankovitch cycles, each of these have variable periods and amplitudes and the cycles are often asymmetric in shape. Thus these cycles are sometimes difficult to discover in time series with traditional tools such as the periodogram. But even sophisticated techniques such as multi-taper methods often report few or no significant periodicities, see for example Thomson (1990).

In Figs. 3 and 8 we have plotted spectral estimates for three sub-periods of the data. In each of these sub-periods there was evidence of periodic components data but we see that these periodic phenomena did not persist across all of the sub-periods. For example, in the Shihua Cave data in the segment beginning 148BC there was evidence of an 8.89 year period oscillation in the data. However, in the period beginning 521AD the spectral estimate at 8.89 years displayed exceptionally low power. We saw a similar occurrence in the Torneträsk data where there was evidence of 6.67 year periodicity in the segment beginning at 1192AD which was absent in the segment being 388AD. There are also longer period cycles indicated in both figures which were present in some segments and not in others. In the other four data sets we either observed transient periodic components as in the Shihua and Torneträsk data or large changes in the structure of the estimated spectrum.

This type of phenomenon has been reported before in the literature. For example, when Thomson (1990) split the Campito Mountain bristlecone pine data into smaller blocks he reported "Numerous lines with high apparent statistical significance were found in individual blocks but did not persist between blocks." Thus the evidence is that some periodicities are excited in some time segments but not in others. It is beyond the scope of this paper to consider physical candidates for these transient periodicities. This is an area worthy of future research.

These observations are typical of non-linear systems in which the spectra are not invariant with time (Gipp 2001). Gipp (2001) has shown that the earth's climate behaves non-linearly at much longer time scales than considered here. Even with the relative stability of the climate in recent millenia the climate system is not approximated well by a linear FGN or FI(d) model.

The excellent approximation of the estimated spectra for these data to a spectrum with the form of Eq. 3 appears to be the consequence of the data having a number of transient periodic phenomena each with high leakage when present. It is perhaps an unfortunate coincidence that the final spectral estimate resembles that of an FGN when the two processes may have little else in common.

7 Conclusion

The evidence is clear that none of the temperature reconstructions considered here are true long memory time series. Thus FGNs or FI(d) series are inappropriate models for these data.

The apparent good fits to FGNs or FI(d) series mask a lot of important detail in the data. In particular, there are transient periodic phenomena clearly present in different segments of some of the reconstructions.

Discounting the appropriateness of FGN and FI(d) models for these data indicates that the apparent long memory is merely an artifact of the method of analysis. Each of these series, perhaps with the exception of Tasmania, is non-stationary. While over the course of the period covered by these data the temperatures have been relatively stable, from a time series analysis point of view there is nothing to assure us that this stability will continue into the future. In particular, the mean reverting nature of FGN or ARFIMA models cannot be appealed to to give us comfort that recent observed temperature increases in instrumental records will naturally reverse themselves in the near future.

References

Abry P, Veitch D (1998) Wavelet analysis of long-range-dependent traffic. IEEE Trans Inf Theory 44(1):2–15

Abry P, Veitch D, Flandrin P (1998) Long-range dependence: revisiting aggregation with wavelets. J Time Ser Anal 19(3):253–266

Allen MR, Smith LA (1994) Investigating the origin and significance of low-frequency modes of climate variability. Geophys Res Lett 21(10):883–886

Baillie RT, Chung S-K (2002) Modeling and forecasting from trend-stationary long memory models with applications to climatology. Int J Forecast 18:215–226

Beran J (1992) A Goodness-of-fit test for time series with long range dependence. J R Stat Soc B 54(3):749–760

Beran J (1994) Statistics for long memory processes. Chapman & Hall/CRC Press

Beran J, Terrin N (1996) Testing for a change in the long-memory parameter. Biometrika 83(3):627–638

Beran J, Terrin N (1999) Testing for a change in the long-memory parameter. Biometrika 86(1):233

Beran J, Whitcher B, Maechler M (2006) Longmemo: statistics for long-memory processes (Jan Beran)—data and functions. R package version 0.9-3

Bloomfield P (1992) Trends in global temperature. Clim Change 21:1–16

Bloomfield P, Nychka D (1992) Climate spectra and detecting climate change. Clim Change 21:275–287

Briffa KR, Jones PD, Wigley TML, Picher JR, Ballie MGL (1992) Fennoscandian summers from 500AD: temperature changes on short and long timescales. Clim Dyn 7:111–119

Brown RL, Durbin J, Evans J (1975) Techniques for testing the constancy of regression relationships over time. J R Stat Soc Ser B 37(2):149–192

Cappelli C, Penny RN, Rea WS, Reale M (2008) Detecting multiple mean breaks at unknown points with Atheoretical Regression Trees. Math Comput Simul 78(2–3):351–356

Chuine I, Yiou P, Viovy N, Seguin B, Daux V, Ladurie ELR (2004) Grape ripening as a past climate indicator. Nature 432(7015):289–290

Cook ER, Buckley BM, D'Arrigo RD, Peterson MJ (2000) Warm-season temperatures since 1600BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface temperature anomalies. Clim Dyn 16:79–91

D'Arrigo R, Wilson R, Jacoby G (2006) On the long-term context for late twentieth century warming. J Geophys Res 111. doi:10.1029/2005JD006352

Deo RS, Chen WW (2000) On the integral of the squared periodogram. Stoch Process Their Appl 85:159–176

Doukhan P, Oppenheim G, Taqqu M (2003) Theory and applications of long-range dependence. Birkhaüser

Embrechts P, Maejima M (2002) Selfsimilar processes. Princeton University Press

Fisher DA, Koerner RM, Kuiviner K, Clausen HB, Johnsen SJ, Steffensen J-P, Gundestrup N, Hammer CU (1996) Intercomparison of ice core and precipitation records from sites in Canada and Greenland over the last few centuries using EOF techniques. In: Jones PD, Bradley RS, Jouzel J (eds) Climatic variations and forcing mechanisms of the last 2000 years, NATO ASI series, vol 41. Springer

Fraley C, Leisch F, Maechler M, Reisen V, Lemonte A (2006) fracdiff: fractionally differenced ARIMA aka ARFIMA(p,d,q) models. R package version 1.3-0

Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. J Time Ser Anal 4:221–237

Gil-Alana LA (2005) Statistical modeling of the temperatures in the northern hemisphere using fractional integration techniques. J Climate 18(24):5357–5370

Gipp MR (2001) Interpretation of climate dynamics from phase space portraits: is the climate strange or just different? Paleoceanography 16(4):335–351

Granger CWJ, Joyeux R (1980). An introduction to long-range time series models and fractional differencing. J Time Ser Anal 1:15–30

Hantemirov RM, Shiyatov SG (2002) A continuous multimillenial ring-width chronology in Yamal, northwestern Siberia. The Holocene 12(6):717–726

Haslett J, Raftery AE (1989) Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). Appl Stat 38(1):1–50

Higuchi T (1988) Approach to an irregular time series on the basis of fractal theory. Physica D 31:277–283

Hosking JRM (1981) Fractional differencing. Biometrika 68(1):165–176

Hurst HE (1951) Long-term storage capacity of reservoirs. Trans Am Soc Civ Eng 116:770–808

Jensen MJ (1999) Using wavelets to obtain a consistent ordinary least square estimator of the long-memory parameter. J Forecast 18:17–32

Jones PD, Mann ME (2004) Climate over the past Millennia. Rev Geophys 42:1–42 RG2002

Klemes V (1974) The Hurst phenomenon—a puzzle? Water Resour Res 10(4):675–688

Mandelbrot BB, van Ness JW (1968) Fractional brownian motions, fractional noises and applications. SIAM Rev 10(4):422–437

Mandelbrot BB, Wallis JR (1969) Global dependence in geophysical records. Water Resour Res 5:321–340

Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392:779–787

Mann ME, Lees JM (1996) Robust estimation of background noise and signal detection in climatic time series. Clim Change 33:409–445

Milhoj A (1981) The test of fit in time series models. Biometrika 68(1):177–188

Mills T (2007) Time series modelling of two millenia of northern hemispere temperatures: long memory or shifting trends? J R Stat Soc Ser A 170:83–94

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

Moore JJ, Hughen KA, Miller GH, Overpeck JT (2001) Little ice age recorded in summer temperatures from varved sediments of Donard Lake, Baffin Island, Canada. J Paleolimnol 25:503–517

Ohanissian A, Russell JR, Tsay RS (2008) True or spurious long memory? A new test. J Bus Econ Stat 26(2):161–175

Overland JE, Percival DB, Mofjeld HO (2006) Regime shifts and red noise in the North Pacific. Deep-sea Res Part 1 53:582–588

Palma W (2007) Long-memory time series theory and methods. Wiley-Interscience

Peng CK, Buldyrev SV, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685–1689

R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
Ripley B (2005) Tree: classification and regression trees. R package version 1.0-19
Robinson PM (2003) Time series with long memory. Oxford University Press
Salzer MW, Kipfmueller KF (2005) Reconstructed temperature and precipitation on a millennial timescale from tree-rings in the Southern Colorado Plateau, USA. Clim Change 70:465–487
Sibbertsen P (2004) Long memory versus structural breaks: an overview. Stat Pap 45(4):465–515
Smith AD (2005) Level shift and the illusion of long memory in economic time series. J Bus Econ Stat 23(3):321–335
Smith RL, Wigley TML, Santer BD (2003) A bivariate time series approach to anthropogenic trend detection in hemispheric mean temperatures. J Climate 16(8):1228–1240
Stephenson DB, Pavan V, Bojariu R (2000) Is the North Atlantic oscillation a random walk? Int J Climatol 20:1–18
Tan M, Liu TS, Hou J, Qin X, Zhang H, Li T (2001) 2650-year Beijing Stalagmite Layer Thickness and Temperature Reconstruction IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series No. 2003-050
Taqqu M, Teverovsky V, Willinger W (1995) Estimators for long-range dependence: an empirical study. Fractals 3(4):785–798
Teverovsky V, Taqqu M (1999) Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. J Time Ser Anal 18(3):279–304
Thomson DJ (1990) Time series analysis of Holocene climate data. Philos Trans R Soc Lond A 330:601–616
Veitch D, Abry P (1999) A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Trans Inf Theory 45(3):878–897
Wuertz D (2005a) fBasics: financial software collection—fBasics. R package version 220.10063
Wuertz D (2005b) fSeries: financial software collection. R package version 220.10063
Zeileis A, Leisch F, Hornik K, Kleiber C (2002) strucchange: an R package for testing for structural change in linear regression models. J Stat Softw 7(2):1–38