JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES, VOL. 118, 7046–7062, doi:10.1002/jgrd.50399, 2013

# Long-range memory in Earth's surface temperature on time scales from months to centuries

K. Rypdal,[1] L. Østvand,[2] and M. Rypdal[1]

Received 7 October 2012; revised 4 April 2013; accepted 8 April 2013; published 10 July 2013.

[1] The paper explores the hypothesis that the temporal global temperature response can be modeled as a long-range memory (LRM) stochastic process characterized by a Hurst exponent $0.5 < H \lesssim 1.0$ on time scales from months to decades. The LRM is a mathematical representation of the multitude of response times associated with the various subsystems. By analysis of instrumental and reconstructed temperature records, we verify LRM on time scales from months to centuries. We employ well-known detrending methods to demonstrate that LRM increases when one goes from local and regional ($H \approx 0.65$) to global ($H \approx 0.75$) land temperature records, and LRM is highest in records strongly influenced by the ocean ($H \approx 1.0$). The increasing trend through the last century cannot be explained as an unforced LRM fluctuation, but the amplitude of the observed 60 year oscillation can be reconciled with the LRM process. We investigate statistical bias and error of the analysis methods employed, and conclude that, for these short record lengths, the error in estimated $H$ is $\pm 0.07$ for the instrumental records. Analysis of a northern-hemisphere reconstruction confirms that the LRM-scaling prevails up to at least 250 years with $H = 0.9 \pm 0.1$. We show that, if this reconstruction is correct, the temperature difference between the Medieval Warm Period and the Little Ice Age cannot be explained as an LRM fluctuation.

**Citation:** Rypdal, K., L. Østvand, and M. Rypdal (2013), Long-range memory in Earth's surface temperature on time scales from months to centuries, *J. Geophys. Res. Atmos.*, *118*, 7046–7062, doi:10.1002/jgrd.50399.

## 1. Introduction

[2] The standard paradigm of natural climate variability up to millennial time scales is that global fields of climatic variables can be decomposed into a diverse set of quasi-coherent modes imbedded in a red-noise stochastic field. This field has spatial correlation length of a few thousand kilometers and autocorrelation time of the order of a year [*Mann and Park*, 1994; *Mann and Lees*, 1996; *Mann and Park*, 1999]. The red-noise hypothesis has replaced an older white-noise assumption and is motivated by a number of empirical studies which suggest that the climate noise can be adequately described as a first-order autoregressive (AR(1)) process $x_k = \phi x_{k-1} + w_k$, characterized by the lag-one autocorrelation $\phi$. The red-noise stochastic process exhibits short-range memory (SRM), i.e., the temporal autocorrelation function $C(t)$ is typically exponentially decaying. Another class of processes is characterized by long-range memory (LRM) and exhibit autocorrelation functions of power-law form $C(t) \sim t^{\beta-1}$ for which the integral $\int_0^\infty C(t)\,\mathrm{d}t$ diverges. Such processes may be Gaussian or non-Gaussian and monofractal or multifractal [*Franzke et al.*, 2012]. In fact, they do not even have to belong to this wide class. It is sufficient that the process is stationary with finite second-order structure function, which is a power-law in the time lag [*Rypdal and Rypdal*, 2012]. The Gaussian approximation is valid for deseasonalized surface temperature records, which are averaged over synoptic spatiotemporal scales (e.g., monthly means averaged over spatial scales $\gtrsim 10^3$ km). Such records are also devoid of signatures of multifractality *Rypdal and Rypdal* [2010]. For all data records analyzed in the present paper, Gaussianity of the deseasonalized and detrended records has been tested by the standard Q-Q-plot technique [*Wilk and Gnanadesikan*, 1968], suggesting that fractional Gaussian noise (fGn) is a proper model for the LRM in these data [*Beran*, 1994]. The power spectral density (PSD) of an fGn has the form $S(f) \sim f^{-\beta}$, and the range $0 < \beta < 1$ describes persistent LRM noise. Here $\beta = 0$ corresponds to uncorrelated (white) noise and $\beta = 1$ to strongly persistent (pink) noise.

[3] The majority of papers dealing with LRM properties in climatic records are confined to analysis of local time records. Thus, *Koscielny-Bunde et al.* [1996, 1998]; *Weber and Talkner* [2001]; *Govindan et al.* [2003]; and *Eichner et al.* [2003] apply the detrended fluctuation analysis (DFA) method to atmospheric instrumental temperature records from localized sites. *Király et al.* [2006] apply it to

[1]Department of Mathematics and Statistics, University of Tromsø, Norway.
[2]Department of Physics and Technology, University of Tromsø, Norway.

Corresponding author: K. Rypdal, Department of Mathematics and Statistics, University of Tromsø, Norway. (kristoffer.rypdal@uit.no)


2169-897X/13/10.1002/jgrd.50399

localized records over land, *Monetti et al.* [2003] over the oceans, and *Bunde and Havlin* [2002] supplement these with records from atmospheric measurements on islands, coastal, and continental stations, and compare with corresponding records from climate models. *Bunde et al.* [2001]; *Govindan et al.* [2002]; and *Vjushin et al.* [2002] also focus on comparing LRM in localized records with climate model results, pointing out a lack of correspondence between observations and models. There are also a few studies of global land and ocean records which suggest LRM properties on time scales from months to decades [*Pelletier and Turcotte*, 1999; *Lennartz and Bunde*, 2009a; *Rybski et al.*, 2006; *Rypdal and Rypdal*, 2010; *Efstathiou et al.*, 2011], and of zonally averaged temperature data which indicate stronger LRM at higher latitudes than in the tropics [*Varotsos and Kirk–Davidoff*, 2006]. It is possible to infer from these papers that temperatures over oceans are more persistent than over land and that global records are more persistent than local. A common feature of virtually all these studies is that biases and uncertainties, arising from the limited record lengths, are not estimated. Thus, one of the main objectives of the present paper is to provide proper error bars on the estimated LRM exponents.

[4] The trends in the instrumental global temperature record (1850–2012) are dominated by a monotonic rise superposed on an oscillation with period of approximately 60 years. It is debated how much of the rising trend, which is of anthropogenic origin, and the nature of the oscillation is poorly understood. *Schlesinger and Ramankutty* [1994] found some evidence that the oscillation is of internal origin, but it has also been suggested that it is related to the motion of the giant planets in the solar system [*Scafetta*, 2010, 2011a, 2011b]. Hence, it is a challenge to determine to what extent this oscillation and the rising trend are driven by some natural or anthropogenic forcing, or are natural fluctuations internal to the climate system. For internal fluctuations, it is also important for predictability on multidecadal time scales to determine if they are constituents of coherent climate modes detectable with high confidence under an LRM-noise null hypothesis, or if they are plausibly explained as fluctuations consistent with a long-memory process. The length of the global instrumental records does not allow us to estimate LRM properties of the records on time scales longer than about 20 years. To establish LRM on longer time scales, we need records of reconstructed temperatures. *Rybski et al.* [2006] employ DFA to establish Hurst exponent of six different reconstruction records, among these the Moberg record analyzed here, and for the latter, they establish a spectral index $\beta = 0.86 \pm 0.03$. The method by which they obtain the error estimates is not explained, but it seems to be based on a standard regression analysis, which assumes a linear model for the log-log fluctuation function with a Gaussian noise superposed. This method, which is based on only one realization of the record, is completely inadequate for testing an fGn-model of the signal, and gives too low error bars and no information about statistical bias. The proper method is to employ Monte Carlo simulations which explores the variability of different realizations of the LRM process. In the present paper, we obtain error bars on this estimate, which allows us to address the important question of whether the millennium oscillation in the reconstruction record, separating the medieval warm period (MWP) from the little ice age (LIA), can be completely described as a realization of an fGn process with the estimated memory exponent. This possibility was suggested by *Rybski et al.* [2006], but without quantitative assessment.

[5] In a recent study, *Vyushin et al.* [2012] compared the performance of the AR(1) statistical model and an LRM model for temperature time series from local observations distributed in a global grid. They also used corresponding data from multimodel ensemble simulations associated with the Coupled Model Intercomparison Project 3, and concluded that both statistical models describe these local data equally well. The persistence in both statistical models are higher over oceans than over continents, and in the climate models, the persistence is independent of the forcing, hence, the LRM properties are associated with the climate response rather than with correlation structures in the forcing. *Vyushin et al.* [2012] do not extend their study to time series of regional and global averages and therefore miss the opportunity to observe that the SRM properties fade away in favor of LRM as one goes from local to global behavior. One of the purposes of the present study is to demonstrate that strong LRM is a fundamental characteristic of global climate response. The methods employed here does not allow us to make tests which discriminate more clearly between SRM models like AR(1) and LRM models like fGn for local climate records. In a forthcoming paper, we will employ methods which utilize the information in available records of global radiative forcing and allow us to test the validity of the two models in describing the recorded climate responses to the known forcing. The result is that also local temperature series are consistent with an LRM process and inconsistent with an AR(1) process.

[6] In section 2 of this paper, we present a stochastic-dynamic model (SDM) of a global climate variable exhibiting LRM response to external deterministic and internal stochastic forcing. This model allows us to estimate $\beta$ in those cases where time series of the deterministic component of the forcing are available. What we estimate by this method are the LRM-properties of the climate response function, independent of correlation structures present in the forcing. Due to space limitations, we will have to show the results of this method in a forthcoming paper. The reason for sketching the method here is to point out that the problem of separating stochastic signal and trend disappears when forcing data is available and is taken into account, and that systematic methods of analysis exist. Section 3 gives a brief summary of more conventional detrending methods which do not require knowledge of deterministic forcing, but have to devise ways to eliminate deterministic trends in the signals. We also present here some new results on how to evaluate the consistency of a given record with the LRM hypothesis using Monte Carlo simulations, and estimates of biases and uncertainties of $\beta$ for the different estimation methods. In section 4 we present detailed analyses of global, regional, and local instrumental records, utilizing methods and results presented in section 3. Section 5 extends these results to centennial time scales by analyzing a northern hemisphere temperature reconstruction covering the last two millennia. The LRM estimates of instrumental as well as reconstructed temperature records are presented with an evaluation of statistical biases and uncertainties resulting from the finite record length. The results allow us to

draw general conclusions about the spatiotemporal origin of LRM (local versus global) and about the roles of land and ocean in its genesis. Section 6 provides further discussion of the results, and we conclude that the oscillations on multi-decadal and multicentennial time scales can be considered as inherent parts of a realization of a long-memory fGn model with $\beta \approx 1$, while the rising trend over the last century cannot be reconciled with such a null hypothesis.

## 2. Estimation of LRM-Response to Known Forcing

[7] For the period since 1880 the global radiative forcing $F(t)$ of the Earth's climate has been estimated with annual resolution, and is routinely used as input in climate models [*IPCC*, 2007]. The evolution of global climatic variables, like the global mean surface temperature (GMST), on decadal to centennial time scales can be modeled as the integrated response of the atmosphere-ocean system to $F(t)$ in addition to stochastic forcing of GMST from internal synoptic-scale dynamics. To illustrate this point, let us consider a simple one-box energy balance model for the GMST anomaly $T(t)$ resulting from an external forcing $F(t)$ and an internal stochastic forcing $\sigma w(t)$, where $w(t)$ is a Gaussian white-noise process of unit variance,

$$\frac{\mathrm{d}T}{\mathrm{d}t} + \frac{1}{\tau_c} T = \frac{1}{C}[F(t) + \sigma w(t)]. \tag{1}$$

Here $C$ is the effective heat capacity of the climate system and $\tau_c$ is the time constant for the climate response. An elementary explanation of the one-box model can be found in *Vallis* [2012], and a derivation is given in *Rypdal* [2012]. The stationary solution of this equation in presence of a constant forcing $F$ and zero stochastic forcing is $T = S_{eq}F$, where $S_{eq} = \tau_c/C$ is the equilibrium climate sensitivity. Since the equation is linear, the general solution can be separated into a response to the deterministic and stochastic forcing,

$$T(t) = \underbrace{\frac{1}{C}\int G(t-s)F(s)\,\mathrm{d}s}_{\text{deterministic solution}} + \underbrace{\frac{\sigma}{C}\int G(t-s)\,\mathrm{d}w(s)}_{\text{Ornstein-Uhlenbeck}}. \tag{2}$$

where $G(t) = e^{-t/\tau_c}\,\vartheta(t)$ is the impulse response, and $\vartheta(t)$ is the Heaviside step function. The response to the stochastic forcing is the well-known Ornstein-Uhlenbeck stochastic process, which has the character of a Brownian motion on time scales shorter than $\tau_c$ and of a white noise on scales longer than $\tau_c$. This stochastic process is the continuous-time analog to the discrete-time AR(1) process. Equation (1) can be generalized to yield an LRM-process (a fractional Gaussian noise (fGn) or a fractional Brownian motion (fBm)) as solutions to the stochastic forcing problem. Formally this is done by replacing the left-hand side of the equation with a Liouville fractional derivative operator $\mathcal{D}^{\beta/2}$ [*Herrmann*, 2011], such that the equation takes the form,

$$\frac{1}{\Gamma(\beta/2)}(\mathcal{D}^{\beta/2}T)(t) = \frac{1}{C}[F(t) + \sigma w(t)]. \tag{3}$$

In practice, it is not essential to know the definition of the fractional derivative, since the equation is uniquely defined by its solution, which is far more instructive;

$$T(t) = \frac{1}{C}\left[\underbrace{\int (t-s)_+^{\beta/2-1}F(s)\mathrm{d}s}_{\text{deterministic solution}} + \sigma\underbrace{\int (t-s)_+^{\beta/2-1}\mathrm{d}w(s)}_{1/f^{\beta}\text{ noise}}\right]. \tag{4}$$

The stochastic part of this solution (the term to the right) has a power spectral density of the form $S(f) \sim f^{-\beta}$, and is an fGn (a stationary process) if $-1 < \beta < 1$ and an fBm (nonstationary) if $1 < \beta < 3$. The physical rationale behind replacing the exponential climate response with a power-law response is discussed in *Rypdal* [2012]. It is argued that the climate response involves more than one single time constant (which has also been noted by several other authors), and that the main features of the GMST record can be better reproduced by the LRM response than by the exponential response. An LRM-like response can also be constructed from multi-box energy balance models involving a hierarchy of interacting subsystems with increasing time constants, such as the atmosphere, ocean mixed layer, sea ice, deep ocean, and so on.

[8] In a forthcoming paper, we employ equations (2) and (4) as parameterized stochastic-dynamic models with the known forcing function $F(t)$ as input and observed and reconstructed global temperature time series as output. The unknown parameters $\{C, \sigma, \tau_c\}$ in equation (2) and $\{C, \sigma, \beta\}$ in equation (4) are then determined by maximum-likelihood estimations (MLE). The MLE method is described in *Beran* [1994] and in most intermediate or advanced textbooks on time series analysis. By modeling the response rather than the signal, the trends are represented as the response to the deterministic component of the forcing. The stochastic component of the signal is uniquely defined as the response to the stochastic forcing, and hence no explicit detrending is needed. This is an obvious advantage compared to those methods where $\beta$ is inferred from the temperature records alone.

[9] When forcing information is not available, or available but not used, trends must be modeled along with the stochastic component of the signal. This can be done within a fully parametric model, e.g., by modeling the signal as an fGn superposed on a polynomial trend of a given order, leaving the memory exponent and the polynomial coefficients to be estimated by MLE. But trends can also be modeled or eliminated in a semiparametric approach where the trend is determined by some smoothing procedure, sometimes guided by physical insight or assumptions, or eliminated by techniques designed to remove polynomial components in the signal up to a given order. We call this approach semiparametric because the methods do not estimate trend parameters. The fully parametric models represent a more systematic approach, but the results are more sensitive to the selection of model, e.g., the selection of the polynomial order of the trend.

[10] In the present paper, we shall not use information about forcing, and hence we will have to separate trends from noise. A major goal is to establish sound physical intuition on this issue by applying a number of semiparametric methods, which derive directly from the scaling properties

of the LRM noise. For this reason, we shall also avoid the more abstract MLE methods in this paper.

## 3. Detrending Methods

[11] These methods have to be implemented if reliable data about the deterministic forcing component are unavailable, or can be used as a complement to the methods described in the previous section even when forcing records exist. Let us assume that the stochastic component of the observed record is a discrete-time stationary stochastic process $x_1, x_2, \ldots$ (a "noise") and let $y_0, y_1, \ldots$ be the cumulative sum (also called the "profile" of the sequence $\{x_k\}$);

$$y_0 = 0, \text{ and } y_t = \sum_{k=1}^{t} x_k, \text{ for all } t = 1, 2 \ldots . \tag{5}$$

In other words, $x_t = y_t - y_{t-1}$ is the differenced profile time series. For a self-similar process $\{y_t\}$ the second-order structure function is a power law [*Beran*, 1994],

$$S_2(\tau) \equiv E[(y_{t+\tau} - y_t)^2] = E[y_\tau^2] \propto \tau^{2H}. \tag{6}$$

Here $H$ is the self-similarity exponent for the profile $\{y_t\}$ and the Hurst exponent for the differenced noise process $\{x_t\}$. If the probability density function is Gaussian, the process $\{y_t\}$ is called a fractional Brownian motion (fBm), and $\{x_t\}$ is a fractional Gaussian noise (fGn). Strictly, self-similarity on all scales is defined only for a continuous-time stochastic process, but the results above are still valid for discrete processes which are self-similar (scale invariant) on scales larger than the time step of the discrete process. The importance of the Hurst exponent is its relation to correlations in the noise $\{x_t\}$. If it is an fGn then the autocorrelation function (ACF) takes the form [*Beran*, 1994],

$$C(\tau) \equiv E[x_t x_{t+\tau}] \sim (2-\gamma)(1-\gamma)\tau^{-\gamma}. \tag{7}$$

where $\gamma = 2 - 2H$. Equation (7) implies that the correlation function of $\{x_t\}$ has algebraic decay for all $H \in (0, 1)$ except for $H = 1/2$, for which $\{x_t\}$ is an uncorrelated noise. For $1/2 < H < 1$, the integral over the correlation function $\int_0^\infty C(\tau)\, d\tau$ is infinite, and this property is what defines long-range memory (or long-range persistence). By taking the Fourier transform of equation (7) it is easy to show that the power spectral density (PSD) also has a power-law form [*Beran*, 1994],

$$S(f) \propto f^{-\beta}, \tag{8}$$

where $\beta = 2H - 1$ is the spectral index. Thus, $H = 1/2$ corresponds to a "flat" PSD (white noise) and $H = 1$ to $S \sim 1/f$ (pink noise). In this paper we shall mainly be concerned with persistent, fractional Gaussian noises (or LRM noises). These are processes characterized by spectral indices in the range $0 < \beta < 1$, or equivalently; Hurst exponents in the range $1/2 < H < 1$, or autocorrelation exponents in the range $0 < \gamma < 1$. The instruments to estimate these exponents are then the instruments to estimate power spectral densities $S(f)$, second-order structure functions $S_2(\tau)$, and autocorrelation functions $C(\tau)$. For $S(f)$ we shall invoke the Fourier transform technique, also known as the periodogram, due to its conceptual simplicity, but for actual computation of the spectral index, we shall employ the Wavelet Variance Analysis (WVA) [*Flandrin*, 1992; *Malamud and Turcotte*, 1999] because of its ability to eliminate the effect of trends. For $S_2(\tau)$, we will perform fluctuation analysis (FA), supplemented by detrended fluctuation analysis (DFA). For ACF, we will use a standard moving-window averaging technique for estimation:

$$C(\tau) = \frac{1}{(N-\tau)\sigma^2} \sum_{k=1}^{N-\tau} (x_{k+\tau} - \mu)(x_k - \mu), \tag{9}$$

where $\mu$ and $\sigma^2$ are the true mean and variance for the stationary process, respectively. In numerical realizations (samples) of stochastic processes, the true mean and variance are known, but in observed time records, they usually are not. In those cases they have to be replaced by the sample mean and variance, and this gives rise to a biased estimate when records are short. In this paper the purpose of generating numerical samples of specified processes is to subject them to the same analysis as applied to observed time records. Since the ACF of the observed record can only be estimated using the sample mean and variance (the biased estimate), we have to do the same with the numerical samples. Analytic expressions for the ACF bias, and methods for corrections, have been obtained by *Lennartz and Bunde* [2009b]. The bias of the ACF estimate is one reason for not using it to estimate the Hurst exponent. On the other hand, the ACF is the most intuitive and direct measure of LRM, and is why we shall use it to test if an observed record is consistent with an fGn model for which the Hurst exponent has already been estimated by other methods.

[12] The PSD is estimated with the periodogram, which for the evenly sampled time series $x_1, x_2, \ldots, x_N$ is defined in terms of the discrete Fourier transform $H_m$ as

$$S(m) = \frac{2|H_m|^2}{N}, \; m = 1, 2, \ldots, N/2.$$

Since our time unit here is the sampling time, the frequency measured in cycles per time unit is $f_m = m/N$. The smallest frequency which can be represented in the spectrum (and the frequency resolution) is $1/N$, and the highest frequency that can be resolved (the Nyquist frequency) is $f_{N/2} = 1/2$.

[13] If we want to eliminate the effect of a linear trend on the estimate of $H$, an elegant approach is to use Wavelet Variance Analysis (WVA) [*Flandrin*, 1992]. Suppose we have chosen a mother wavelet $\psi(t)$. Common choices of $\psi(t)$ are $n$th order derivatives of the Gaussian function, among which the second-order derivative (the Mexican-hat wavelet) is most frequently used. For an fGn characterized by $\beta = 2H - 1$, the variance of the wavelet coefficient $W(t, s) = (1/\sqrt{s}) \int \psi[(t' - t)/s]\, dt'$, i.e.,

$$V_w(s) = \frac{1}{N} \sum_{t=1}^{N} |W(t, s)|^2$$

depends on the wavelet scale $s$ like

$$V_w(s) \sim s^\beta.$$

The WVA method, with $n$th order derivatives of the Gaussian wavelet, filters out oscillations on the scale $s$ and has much in common with the local Fourier transform. Because it is a local filter, it reduces the effects of trends on longer scales than the scale $s$ and eliminates exactly polynomial trends of order $n - 1$ and lower. In this paper we

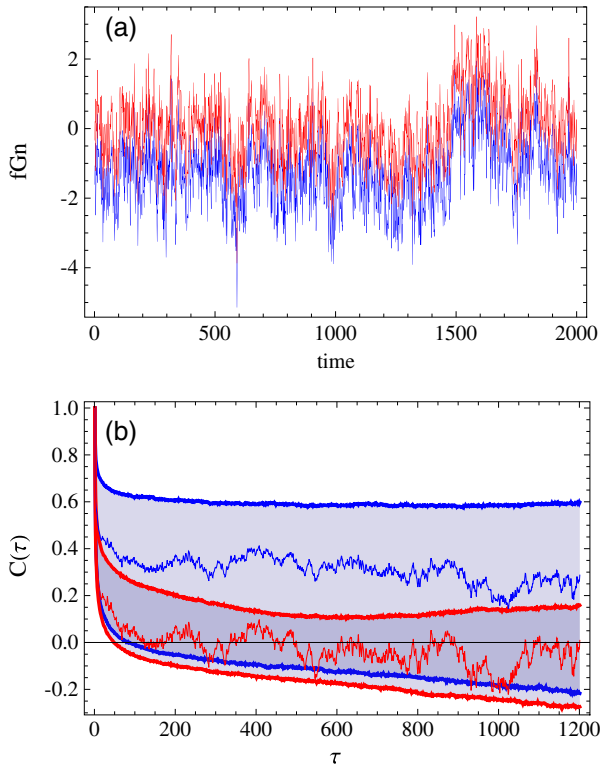


**Figure 1.** (a) Blue: a numerical realization of an fGn with $\beta = 0.8$ ($H = 0.9$). Red: the same signal, but with zero sample mean. (b) The blue, irregular curve is the unbiased ACF estimate from the blue signal in Figure 1a. The red, irregular curve is the biased ACF estimate from the red signal in Figure 1a. The thick, blue curves mark the border of the 95% confidence region for the unbiased ACF estimate, based on an ensemble of 5000 realizations of the fGn. The red, thick curves mark the border of the confidence region for the biased estimate.

will restrict ourselves to the Mexican-hat wavelet, which completely eliminates linear trends and reduces the effect of higher-order trends. In the figures, we shall plot the wavelet variance as a function of the stretched scale $\tau = (10/3)s$ which for this wavelet is approximately the period of oscillation in the wavelet. Using this as the scale parameter allows direct comparison with Fourier methods like the periodogram.

[14] Because of the stationarity of the increments of the profile $y_t$, the square root of the second-order structure function $\sqrt{S_2(\tau)}$ can be estimated by the fluctuation function;

$$F(\tau) \equiv \sqrt{\frac{1}{N-\tau}\sum_{l=1}^{N-\tau}|y_{l+\tau}-y_l|^2}. \tag{10}$$

According to equation (6), the fluctuation function of an LRM process with Hurst exponent $H$ scales with $\tau$ as

$$F(\tau) = k\tau^H, \tag{11}$$

where $k$ is a constant, and hence $\log F(\tau) = H \log \tau + \log k$. The plot of $\log F$ versus $\log \tau$ is a straight line with slope $H$ if $\{x_t\}$ is an LRM noise.

[15] Like the FA method, the detrended fluctuation analysis (DFA) is performed on a fluctuation function based on the profile $\{y_t\}$ [*Koscielny-Bunde et al.*, 1996, 1998]. The profile is divided into $N_\tau = N/\tau$ non-overlapping segments of equal length $\tau$ and enumerated by the index $\nu = 1, \ldots, N_\tau$. In each segment an $n$th order polynomial fit is computed and subtracted from $y_t$ for each segment, thus producing a locally detrended signal. In the final step, the variance $F^2(\nu, \tau)$ for the detrended signal in each segment is computed, and the fluctuation function is found as the square root of the average over all the segments;

$$F(\tau) = \left[\frac{1}{N_\tau}\sum_{\nu=1}^{N_\tau} F^2(\nu,\tau)\right]^{\frac{1}{2}}, \tag{12}$$

The Hurst exponent is then estimated from the asymptotic relation $F(\tau) \sim \tau^H$ by plotting $\log F(\tau)$ against $\log \tau$ and computing the slope of the linear regression line. The fluctuation function depends on the order of the detrending polynomial, hence, for polynomial order $n$, we denote the method as DFA$n$. For a time series with no trends, our detrending function is a zeroth-order polynomial, i.e., we subtract the segmental mean from $y_t$ in every segment. We shall adopt the convention of denoting this method DFA0, in accordance with *Eichner et al.* [2003]. This is not identical to FA, where the record mean is subtracted in every segment.

[16] The fact that all measures of LRM have their uncertainties and biases is not an unsurmountable problem if one explores the opportunity to clarify these through Monte Carlo simulation of the specified LRM processes. When we know the biases through analysis of large ensembles of simulated realizations of the processes, we can correct our analysis results, and we can obtain confidence estimates. In this paper we shall show some examples on how this can be done, which will give us an idea about how accurate the estimates we can obtain from the relatively short climate records that we have at hand. The method we employ to generate an LRM process with a given Hurst exponent is described in *McLeod et al.* [2007]. The resulting signal has the desired correlation structure, a PSD on the form $S(f) \sim f^{-\beta}$, and is a realization of an fGn with $H = (\beta + 1)/2$.

[17] In Figure 1a, the blue curve is a realization of an fGn with $H = 0.9$ ($\beta = 0.8$, $\gamma = 0.2$) containing 2000 data points. The true mean (ensemble mean) of the process is zero, but the sample mean is not. The red curve is the same signal with zero sample mean. In Figure 1b, we have plotted the ACF estimate for this realization (blue irregular curve). The theoretical ACF for this process decays as $1/\tau^\gamma$, but due to the finite length of the sample, the estimate is very noisy. The red irregular curve is the biased ACF estimate obtained from the red signal with zero sample mean in Figure 1a. By computing these unbiased and biased estimates for an ensemble of 5000 realizations of the fGn process, and computing the ensemble mean, we obtain one smooth curve for the unbiased estimate and another for the biased estimate. The former is a $1/\tau^{0.2}$-function, but the latter will attain negative values for large $\tau$ due to the bias [*Lennartz and Bunde*, 2009b]. This negative bias is more pronounced for $H$ approaching unity. At any given $\tau$ we compute the 95% confidence interval for the distribution of ACF estimates. The border of these intervals are shown as the blue thick curves in Figure 1b for the unbiased estimates, and as the red, thick curves for the biased estimate. If an observed record has a biased ACF estimate within the confidence limits marked by

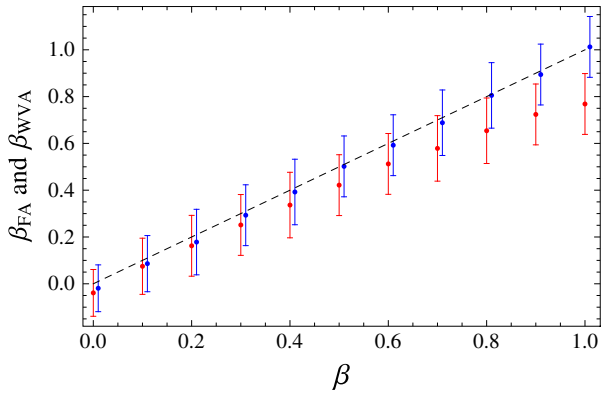


**Figure 2.** $\beta_{FA}$ (red) and $\beta_{WVA}$ (blue) plotted against $\beta$. Every point has been computed from an ensemble of numerically generated fGn records of only 2000 data points. The error bars are 95% confidence intervals and are approximately $\pm 0.14$ for both estimation methods.

the red curves, the variability of the record can be described as a natural fluctuation within the LRM process, and hence needs not be explained as trends imposed by external forcing. On the other hand, if the estimate extends way beyond these confidence limits, one has to conclude the existence of trends, provided the null hypothesis is an LRM process with the prescribed Hurst exponent. A great advantage of this simple test is that it is capable of detecting both slow and fast signal components violating the null hypothesis, not only the slow trends. For instance, climate oscillations which are incompatible with the LRM hypothesis could be detected, irrespective of their characteristic period.

[18] The estimates which yield equation (11) are unbiased only if $\{x_k\}$ are samples of a process with true mean $\mu = 0$. If $\{x_k\}$ are samples with zero sample mean, the variogram is a strongly biased estimate for $H$ close to 1, where it returns too small values for short records. This was observed in Monte Carlo simulations by *Malamud and Turcotte* [1999], and computed analytically by *Lennartz and Bunde* [2009b]. Such biases is one of many reasons to use several different estimators when one investigates data for long-range memory. DFA does not have this bias problem because it is inherent in the method to subtract the segment mean, and the same is the case with power spectra and wavelets. Finite length of the records also introduce large uncertainties in the estimates, and this is a problem with all methods, although some are worse than others [*Franzke et al.*, 2012]. This is shown for FA and WVA in Figure 2, where estimates have been made based on ensembles of 1000 realizations with record length 2000 data points, which is the typical length of the climatic data records we analyze in this paper. Here we observe that while the typical bias for $\beta_{WVA}$ is negligible, the bias for $\beta_{FA}$ when $\beta$ approaches 1 is close to –0.2. The $\pm 2\sigma$ error for both estimates over the entire $\beta$-interval is approximately $\pm 0.14$. Since $H = (\beta + 1)/2$, the corresponding figures for $H$ is $\pm 0.07$.

[19] In Figure 3, we investigate the detrending capability of FA and WVA for records of 2000 data points. We generate a numerical realization $x_t^{(\beta)}$ of an fGn with $\beta = 0.5$ and unit variance and analyze this record and another record $X_t^{(\beta)} = x_t^{(\beta)} + 0.001(t - 1000)$. The growth in $X_t^{(\beta)}$ due to the added linear trend over the record is twice the standard deviation of the noise, which is not more than what is obtained by linear regression of the instrumental temperature records, which we will analyze in the next section. The chosen value of $\beta$ is also in the range found in these records, so this signal exhibits roughly the LRM- and trend-properties of the instrumental records. Figure 3a shows the variogram for the fGn signal (red) and the signal with trend (black). The former has slope $H = 0.75$, corresponding to $\beta = 0.50$, while the latter has slope $H = 0.82$, corresponding to $\beta = 0.64$. Hence, FA for the signal with trend gives a clear overestimate of the true exponent $\beta = 0.5$. For a stronger trend, FA on the signal with trend will return $\beta \approx 1$, i.e., the FA is totally overwhelmed by the trend. How such analyses have lead to misinterpretations were discussed by *Rypdal and Rypdal* [2010]. In Figure 3b, we show the corresponding results from the WVA. The fluctuation function for both signals look very similar up to a certain scale; in this case $\tau_+ \approx 60$. The effect of the trend appears in the black curve for $\tau > \tau_+$ as a cross-over to a scaling dominated by the trend. The value of $\tau_+$ is reduced for stronger trend. The curvature for $\tau \leq \tau_- \approx 3$ is inherent in the wavelet method. Hence,

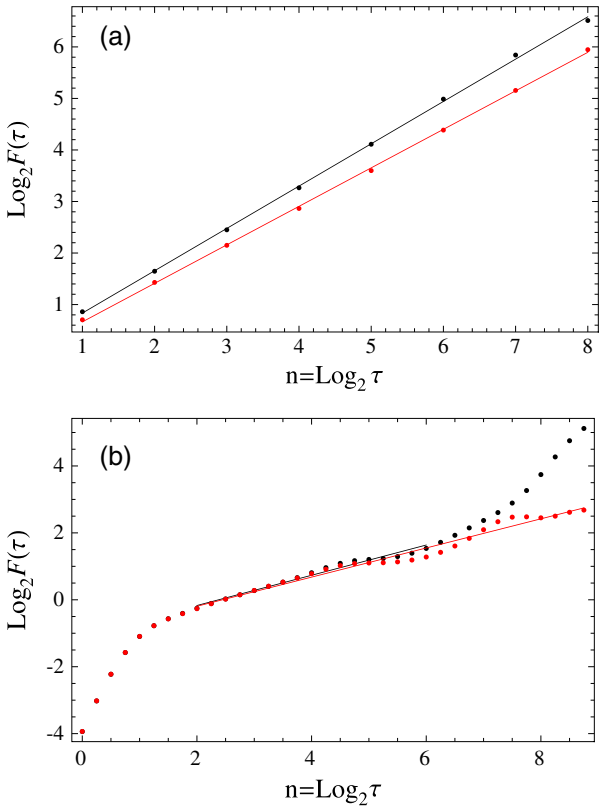


**Figure 3.** (a) FA of synthetic fGn record with $\beta = 0.5$ ($H = 0.75$), $\sigma = 1$ and length 2000 data points (red) and FA of the same record with the linear trend with slope 0.001 (black). Over the entire record, the trend implies an increase of $2\sigma$. The slope of the red curve is $H = 0.75$ and for the black curve $H = 0.82$. (b) The same as in Figure 3a, but for WVA. The slope of the red curve corresponds to $H = 0.72$ and for the black curve to $H = 0.73$. The scale $\tau$ used in the WVA is $\tau = (10/3)\,s$, where $s$ is the wavelet scale parameter. This convention is used in all WVA plots throughout the paper.

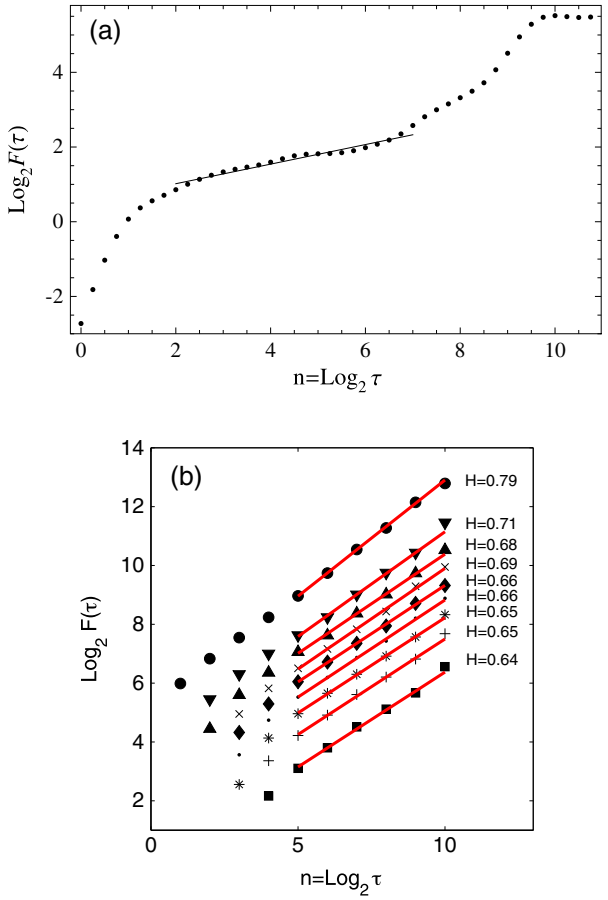


**Figure 4.** (a) WVA of the CET record. The slope of the black line is $\beta = 0.26$, corresponding to $H = 0.63$. (b) DFA0–8 of CET. The upper curve is the fluctuation function for DFA0, the ones below are DFA$n$, $n = 1, \ldots 8$, with DFA$n+1$ coming as the curve right below DFA$n$.

we conclude that if the trend is weak enough to provide a segment $\tau_- < \tau < \tau_+$, which is long enough to fit a straight line, the WVA will allow us to obtain a good estimate of $\beta$ for the underlying noise process. However, the method is not fool proof. One has to examine the fluctuation function to find the best fitting interval (if possible), and one has to take into consideration the uncertainties that were demonstrated in Figure 2.

[20] The detrending properties of DFA with respect to a linear trend can be studied the same way as we did for FA and WVA in Figures 2 and 3. We find that the bias for DFA and WVA are both negligible, while the errors are somewhat larger for DFA. In our WVA analysis we have used the Mecixan-hat wavelet. Higher-derivative wavelets will have effects similar to higher-order DFA. They will in principle have better detrending capabilities, but for short records higher-order wavelets will give rise to stronger oscillations in the fluctuation function and high-order DFA has spuriously steep fluctuation function for small $\tau$. Thus, the general performance and detrending capabilities of the two methods are quite similar. Curiously, DFA has completely dominated the literature on LRM in climate records.

[21] Throughout this section we have for conceptual simplicity used the fGn as our paradigmatic model, and it could be objected that this is also a parametric model, requiring self-similarity and Gaussianity. However, all results described above, except for those based on the Monte Carlo simulations of fGns, are valid for a much broader class of processes. The power-law dependence of the ACF and PSD depends only on the power-law dependence of the second-order structure function $S_2(\tau)$ [*Rypdal and Rypdal*, 2012]. Hence, the PDFs do not have to be Gaussian, the only requirement is that the second moment is finite. Moreover, the process does not have to be self-similar. It could be multifractal, i.e., the $q$th structure function $S_q(\tau) = E[y_\tau^2] \propto \tau^{\zeta(q)}$ does not need to have scaling exponent $\zeta(q)$ which is linear in $q$. And it does not even have to belong to the class of multifractals, since we don't require that $S_q(\tau)$ are power laws in $\tau$, except for $q = 2$. Hence, the techniques of periodogram, FA, DFA, and VWA all estimate the scaling exponent $H$ for a wide class of stationary processes with finite second-order structure function, which scales like $S_2(\tau) \propto \tau^{2H}$. This is the strength of these techniques, which make them worthwhile to pursue in spite of weaknesses as estimators.

## 4. Analysis of Instrumental Temperature Records

[22] In this section we analyze three different instrumental temperature records with detrending methods. We start the analysis of each record by WVA and DFA0–8 to obtain a first assessment of the scaling properties and an "automatic" estimate of the Hurst exponent. This is followed by estimation for different degrees of polynomial detrending of periodograms, FA, and ACF estimates with confidence limits determined from Monte Carlo simulations. The purpose of applying these simple estimators on the polynomially detrended signals is to establish which degree of polynomial detrending we can undertake before we destroy the fGn scaling at long time scales. The physical significance of the results are discussed on the way.

### 4.1. The Central England Temperature Record

[23] The Central England temperature record (HadCET) is the longest continuous instrumental record in the world. The monthly mean temperatures are recorded from 1659 to date and are representative of a roughly triangular area of the United Kingdom enclosed by Lancashire, London, and Bristol [*Manley*, 1974; *Parker et al.*, 1992]. The data set can be downloaded from the Hadley Center Met Office web site. It is assumed to be representative of the monthly mean temperature variations over a region with spatial extent of a few hundred kilometers, and hence is somewhat less influenced by weather noise than records from a single station, but much more than hemispheric or global records. The seasonal variation of this record is obtained by computing the climatology, which is the mean temperature of a given month averaged over the record. The climatology curve over the year is very close to a sine function with peak-to-peak amplitude of approximately 12 K. The deseasonalized record is obtained by subtracting the climatology.

[24] The WVA and DFA estimates of the deseasonalized CET record are shown in Figure 4. In this case it is a bit difficult to determine the exact position of the crossover $\tau_+$ in the WVA fluctuation function. This is because of a wave-like structure on the fluctuation function, which is an effect of the finite record length. These waves are present also in

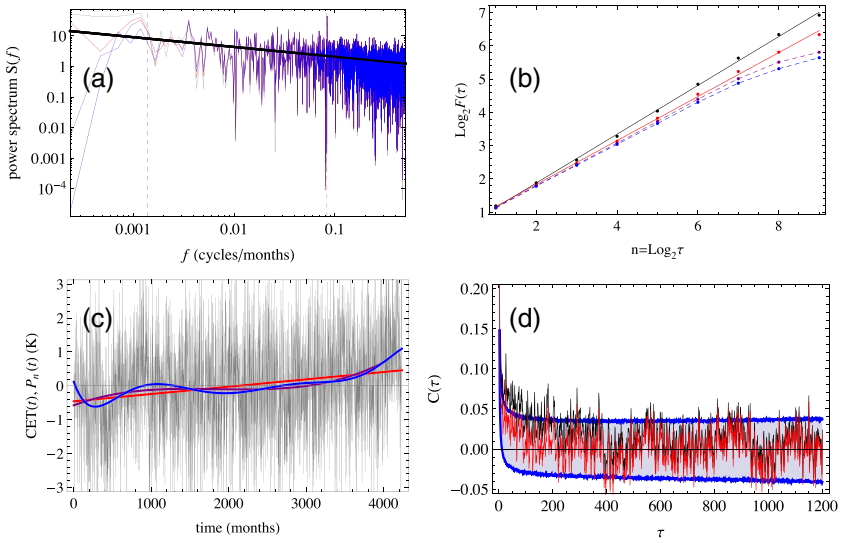


**Figure 5.** (a) PSDs of deseasonalized monthly CET record 1659–2011 A.D. with variable degree of detrending. Gray: undetrended. Red: $P_1$ detrended. Purple: $P_3$ detrended. Blue: $P_7$ detrended. Thick line has slope $-\beta = -0.30$, corresponding to $H = 0.65$. Vertical dashed lines mark the 60 year period (blue), and the 1 year period (red). (b) FA of the CET record with variable degree of detrending. Black: after no detrending. Red: after $P_1$-detrending. Purple: after $P_3$-detrending. Blue: after $P_7$-detrending. The slope of the black line is $H = 0.73$ and of the red line is $H = 0.66$. (c) Gray: Deseasonalized monthly CET record in degrees Kelvin (time origin starts 1659 A.D.). Colored: Polynomial fits. Red curve: $P_1$-fit. Purple: $P_3$-fit. Blue: $P_7$ fit. (d): Black: ACF estimate from undetrended, deseasonalized CET record. Red: ACF estimate from $P_1$-detrended, deseasonalized CET record. The shaded area represents the 95% confidence interval for the ACF computed from an ensemble fGns of the same length as the GMLT record and with $H = 0.65$.

numerical realizations of fGns (see e.g., the red curve in Figure 3 b), but are reduced for longer records. For the CET record, the positive phase of this wave incidentally coincides with $\tau_+$ and makes the crossover less evident. However, with this insight, we estimate that $\tau_+ \approx 2^7$ and use this as the upper border of the fitting region, giving the estimate $H = 0.63$. Interestingly, an analysis of a shorter record from a single station (Durham, UK, 1880–2012) yields the same $H$, suggesting that local and regional Central England temperatures on time scales longer than a month exhibits the same scaling properties. The uncertainty of the estimate shown in Figure 4a is greater than usual, because of the uncertainty in estimating $\tau_+$. Figure 4b shows an DFA0-slope of $H = 0.79$ converging toward $H \approx 0.64$ for DFA8. We made assessments of bias and uncertainty of these estimates in section 3 (Figure 2), and found negligible bias and uncertainty of $\pm 0.07$ for both WVA and DFA8. This suggests an fGn process with $H = 0.64 \pm 0.07$ superposed on a linear trend, which is significant enough to influence the DFA0 analysis.

[25] We shall supplement these estimates with a more intuitive heuristic analysis based on the periodogram and the FA. The deseasonalized record is shown in Figure 5c, along with linear ($P_1$), third-order ($P_3$), and seventh-order ($P_7$) polynomial least-square fits to this record. When a $P_n$ polynomial fit is subtracted from the deseasonalized record, we shall refer to the result as a $P_n$-detrended record. The $P_1$-detrended, deseasonalized CET record has standard deviation 1.39 K.

[26] Since the most intuitive measure of LRM is the estimated PSD (we use the periodogram estimator) in a log-log plot, we have made such plots for varying degrees of detrending. In general, detrending reduces the low-frequency components in the spectrum, and more so for higher degree of the detrending polynomial. For the CET record, the power in the lowest frequencies is not very much above a linear fit to the log-log spectrum, but a $P_1$-detrending seems to give a better power-law behavior of the PSD. The black line in Figure 5c is not a fit to any of the spectra, but a line of slope $-\beta = -0.30$ ($H = 0.65$). The reason for plotting this line derives from the results of WVA and DFA shown in Figure 4, but also from the FA curves, as will be explained in the following. Fluctuation functions for $P_n$-detrended records are shown in Figure 5b. Both the undetrended and $P_1$-detrended record exhibit good scaling (straight log-log variograms) on scales up to $2^9$ months (about 40 years), but higher-order detrending destroys the scaling for $\tau > 2^7$ months (about 10 years). This means that $P_n$-detrending with $n > 1$ removes low-frequency components in the record, which are consistent with the LRM-scaling, and hence, periodograms for such detrended records will show power in the low frequencies below the straight line in Figure 5a. The slope of the variogram for the $P_1$-detrended record is $H = 0.66$, which corresponds to a spectral index $\beta = 2H-1 = 0.32$. Recalling that the FA bias is negligible for this small $\beta$, this result is consistent with those found from WVA and DFA. In summary, WVA, DFA, and FA yield $H$ estimates of 0.63, 0.64, and 0.66, respectively, suggesting the best estimate $H = 0.65 \pm 0.07$ for the CET record.

[27] We can also use Monte Carlo simulations to check that our estimates are consistent with the conjecture that the $P_1$-detrended record is a realization of this fGn process. What we want to demonstrate is that the $P_1$-detrended observed record falls well within the ensemble of simulated fGns with our estimated $H$, or more precisely, that the

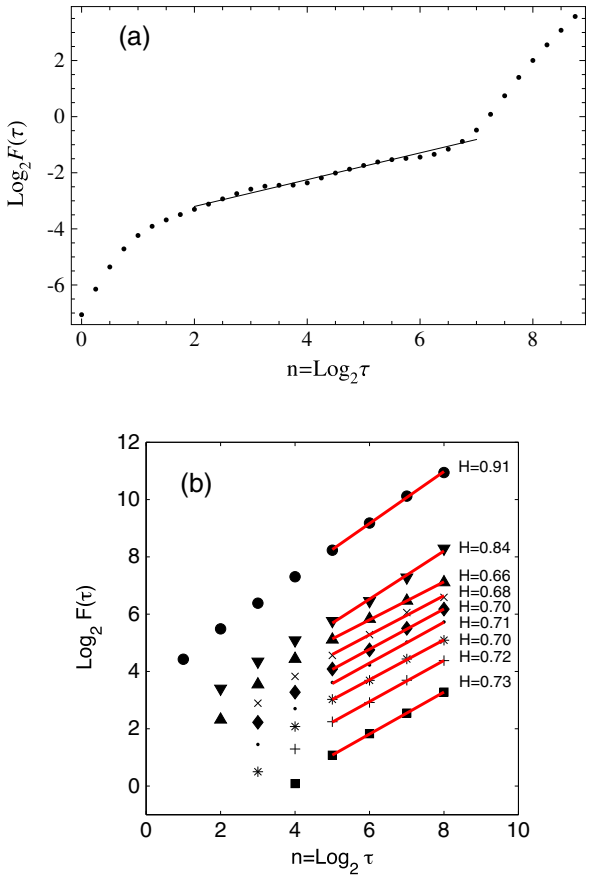


**Figure 6.** (a) WVA of the GMLT record. The slope of the black line is $\beta = 0.48$, corresponding to $H = 0.74$. (b) DFA0–8 of GMLT. The upper curve is the fluctuation function for DFA0, the ones below are DFA$n$, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFA$n$.

long time-scale variability of this record (which alternatively could be interpreted as trends) lies within the statistical spread of the simulated ensemble. One measure we can use to estimate this spread is the biased ACF. In Figure 5d, we have estimated the ACFs for the undetrended (black) and $P_1$-detrended (red), deseasonalized CET record. Since the record has a finite length, this estimate is too noisy to be used to assess whether the ACF has a power-law asymptotic dependence, and if so, to estimate the exponent $\gamma$. A problem in this context is that estimating an exponent would require to look at the ACF estimate in a log-log plot, but this is impossible since the noisy estimate is not always positive. This, in addition to the known bias, is an obvious reason for using the PSD estimate (which is positive definite), rather than the ACF, for estimating memory exponents. What we can do, however, is to estimate the statistical spread of the biased ACF estimates in the simulated ensemble. This spread is shown as the shaded area in Figure 5d. The fact that the estimated ACF for the $P_1$-detrended record mostly falls within the $\pm 2\sigma$ confidence interval ($\sigma$ is the standard deviation of the distribution of the simulated ACFs) shows that the fluctuations on all time scales of the $P_1$-detrended record are within the limits of fluctuations that can be expected in realizations of an fGn with $H = 0.65$.

[28] In the present example, this result is quite obvious and trivial, since the noise dominates the "trends" and the long-range correlations appear to be rather weak. The results are consistent with the findings of *Bunde et al.* [2001] for records from individual continental stations. The situation changes, however, when observations from individual stations are synthesized into a global temperature record. We have analyzed the global, monthly mean temperature records from land, ocean, and combined land-ocean from 1850 A.D. to present [*Brohan et al.*, 2006]. These data sets are freely downloadable from the Hadley Center. The ocean sea surface temperatures are so dominating in the combined data set, that the analysis of the two data sets (ocean and combined land-ocean) yields virtually identical results. In this paper we therefore only present the results from the combined data set.

### 4.2. The Global Land Temperature Record

[29] Figures 6 and 7 show the results of the analysis of the global mean land temperature record (GMLT), the CRUTEM4 global data set. The result of WVA is shown in Figure 6a and yields $H = 0.74$. Figure 6b presents the result of the DFA. Here, DFA0 yields a slope of $H = 0.91$, which converges to $H \approx 0.73$ for DFA1-8. The temperature record itself is shown in Figure 7c along with polynomial fits $P_1$, $P_3$, and $P_7$. The standard deviation of the $P_3$-detrended record is 0.35 K. This is four times less than the standard deviation for the CET data set, and demonstrates the dramatic reduction global spatial averaging introduces on monthly fluctuation levels. Fluctuation levels are reduced for both data sets if one performs a moving average with a time window $\tau$, and the reduction is larger for larger $\tau$. The rate at which the fluctuations change with window size is exactly what fluctuation analysis measures, i.e., the standard deviation of the moving average with window $\tau$ is $F(\tau)/\tau$. Hence, if the data set exhibits scaling with Hurst exponent $H$, the fluctuations of the moving average scales as $\tau^{H-1}$. It may not come as a surprise that the fluctuation level of the CET data set decreases faster with increasing $\tau$ than the corresponding fluctuation level for the GMLT data, since the monthly fluctuations for the former is so much higher, and after averaging over several decades, the fluctuations of two data sets are both dominated by global variability and are of similar magnitude. A faster decrease of the moving average with increasing $\tau$ implies a smaller $H$, hence, we should expect that the FA yields smaller $H$ for CET than for GMLT. This is exactly what is found in Figure 7b. Here the dotted line has a slope $H = 0.77$ (and corresponds to the thick line in Figure 7a), and is the representative scaling exponent after $P_3$-detrending. In this range of Hurst exponents ($H = 0.77$, $\beta = 0.54$), the bias of $H$ for FA estimates is close to –0.03, and the error approximately $\pm 0.07$. This should yield $H = 0.80 \pm 0.07$. $H$ for WVA has negligible bias and error $\pm 0.07$ and yields $H = 0.74 \pm 0.07$. DFA converges to something near $H = 0.73$, with negligible bias and error (for DFA8) $\pm 0.07$. A value that is consistent with all these constraints must be close to $H = 0.75$.

[30] The different scaling for CET and GMLT is physically very important because it illustrates that local variability is dominated by the horizontal spatial structure of the atmospheric circulation systems, while the global variability is dominated by different dynamical mechanisms (as

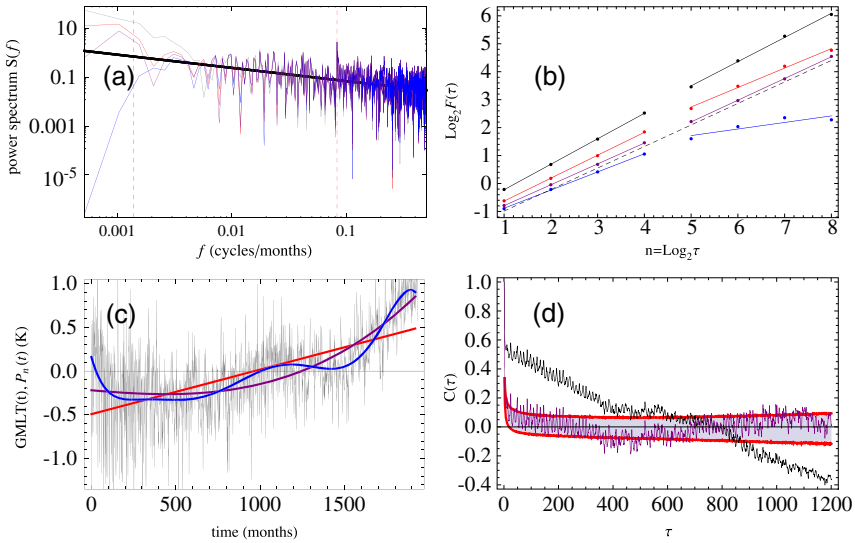


**Figure 7.** (a) PSD of monthly GMLT record 1850–2010 A.D. Gray: undetrended. Red: $P_1$ detrended. Purple: $P_3$ detrended. Blue: $P_7$ detrended. Thick line has slope $-\beta = -0.54$, corresponding to $H = 0.77$. Vertical dashed lines mark the 60 year period (blue), and the 1 year period (red). (b) FA of the GMLT record with variable degree of detrending. Black: after no detrending. Red: after $P_1$-detrending. Purple: after $P_3$-detrending. Blue: after $P_7$-detrending. The slopes for $n = 1, \ldots, 4$ correspond to the following: black: $H = 0.91$, red: $H = 0.82$, purple: $H = 0.78$, and blue: $H = 0.65$. The slopes for $n = 5, \ldots, 8$ are as follows: black: $H = 0.87$, red: $H = 0.70$, purple: $H = 0.77$, and blue: $H = 0.23$. Dotted line has slope 0.77. (c) Gray curve: Monthly GMLT anomaly record 1850–2010 in degrees Kelvin (time origin starts 1850 A.D.). Red curve: $P_1$- fit. Purple: $P_3$-fit. Blue: $P_7$ fit. (d) Black: ACF estimate from undetrended GMLT record. Purple: ACF estimate from $P_3$-detrended record. The shaded areas represent the 95% confidence interval for the ACF computed from ensembles of fGns of the same length as the GMLT record and with $H = 0.75$.

will become evident in the next subsection) influenced by the ocean-atmosphere interaction.

[31] We have not yet explained why we have chosen a $P_3$ detrended signal for estimation of $H$ for the GMLT by the FA method. Again, this is based on an assessment of the result of the analysis using several different methods. The periodograms for undetrended and $P_1$-detrended records show more power in the lowest frequencies than consistent with a straight-line fit to the corresponding log-log periodogram, and for detrending higher than $P_3$ there is too little power in these frequencies. The corresponding signatures in the FA plots in Figure 7b is that scaling is lost for $\tau > 3$ years for higher polynomial detrending. This picture is supported by the ACFs in Figure 7d. The undetrended ACF estimates is outside the confidence limits for a Monte Carlo ensemble with $H = 0.75$, while the $P_3$ detrended ACF is within these limits. This implies that the monotonic trend is inconsistent with an fGn with $H = 0.75$, while the apparent 60 year oscillation, which is prominent in the $P_7$-detrended record, can consistently be described as an fGn-fluctuation with this Hurst exponent.

### 4.3. The Combined Global Ocean and Land Record

[32] This is the HadCRUT3 global data set for global mean surface temperature (GMST), for which results are shown in Figures 8 and 9. As mentioned earlier, this data set is very similar to the HadSST global ocean sea surface temperature data. The main difference from the analysis of the GMLT is that the WVA and all DFA3-8 curves have slopes corresponding to $H \approx 1.0$ (Figure 8), and the log-log periodogram of the $P_3$-detrended record is well fitted by a line with slope $-\beta = -1$ (Figure 9a). For $H \approx 1.0$ the bias of $H$ for WVA and DFA3 is negligible, while the error for WVA is $\pm 0.07$ and for DFA3 is $\pm 0.14$. However, the variograms of the $P_n$-detrended records have slopes similar to that of the GMLT record (Figure 9b), suggesting a lower Hurst exponent in the range $0.8 < H < 0.9$. One reason why the slopes of the $P_n$-detrended variograms is lower than $H = (\beta + 1)/2 = 1$ suggested by the periodogram, WVA, and DFA is the large bias of the variogram when $H$ is close to unity, as shown in Figure 2. This bias of $H$ is –0.12 and the error is $\pm 0.07$. We must also take into account that the estimates of bias and errors for FA in Figure 2 are done for fGns without trends. As is apparent from Figure 9b, it is difficult to assess accurately the appropriate degree of polynomial detrending, and the appropriate fitting interval for the FA, although of the variograms displayed, the one for $P_3$ detrending gives the most constant scaling over the entire range of $\tau$. That $P_3$ detrending is the most appropriate is confirmed by the results shown in Figure 9d. Here the ACF estimates for the undetrended record are outside the confidence intervals for ACFs for an $H = 0.99$-ensemble, whereas the ACF estimate for the $P_3$-detrended record is within these confidence limits.

[33] The $P_7$ polynomial fits for GMLT og GMST highlight the existence of an oscillation in the instrumental records with period of about 60 years. It has been suggested that this oscillation is of astronomical origin [*Scafetta*, 2010, 2011a, 2011b], while the mainstream view is that it is of internal origin and associated with the Atlantic Multidecadal

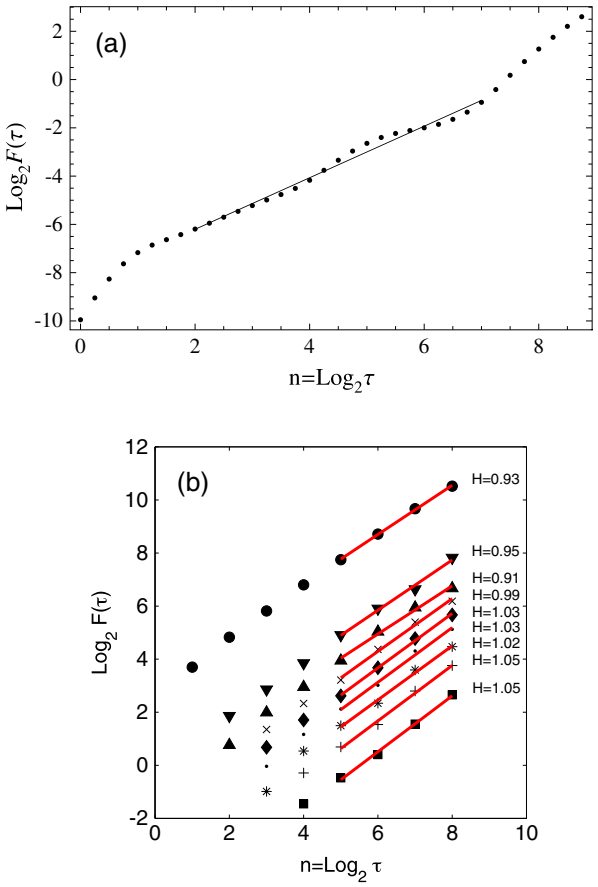


**Figure 8.** (a) WVA of the GMST record. The slope of the black line is $\beta = 1.07$, corresponding to $H = 1.03$. (b) DFA0–8 of GMST. The upper curve is the fluctuation function for DFA0, the ones below are DFA$n$, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFA$n$.

Oscillation (AMO). Our analysis shows that removal of this oscillation by $P_7$-detrending destroys the scaling properties of the record on time scales longer than a decade (the blue variograms in Figures 7c and 9c, and that these oscillations (which are present after $P_3$-detrending) are within the confidence limits for fGns with the estimated Hurst exponents. In other words, these oscillations are explicable as natural LRM-fluctuations superposed on a $P_3$ growing trend, which most likely is of anthropogenic origin. This is not inconsistent with the AMO interpretation, since the AMO is not a coherent oscillation, but rather one of many natural oscillations of the climate system whose totality might be well represented in the global temperature record as a pink-noise LRM process.

[34] Since the GMST record is very similar to the global sea surface temperature (SST) time series, it is reasonable to assume that the higher memory in the GMST data, compared to the GMLT, is due to the thermal inertia of the oceans. However, the time constant of the thermal interaction between the atmosphere and the ocean mixed layer is estimated to be at most a few years [*Padilla et al.*, 2011], while we find LRM extending at least over several decades. This high inertia must involve heat exchange between the mixed layer and the deep ocean which involves the thermohaline overturning circulation [*Vallis*, 2012].

[35] The standard deviation of the monthly $P_3$-detrended GMST is only 0.17 K, which is half of that of the GMLT. On multidecadal time scales, the two records have similar variability, so, this is consistent with the higher Hurst exponent for the GMST. It may also suggest that the physical source of the LRM in land temperatures is really associated with ocean dynamics and ocean-atmosphere interaction, and not within the atmosphere itself.

## 5. Analysis of Hemispheric Reconstructions

[36] A rule of the thumb is that scaling in a time record of length $N$ can be verified by FA or DFA only for time scales up to $\tau \sim N/10$. For longer time scales, the number of independent samples (the number of independent windows of length $\tau$) is so low that the tail of the distribution of the fluctuations is not well represented and the variance is underestimated. The result is that the log-log curve of the fluctuation function bends over for these large $\tau$. For the WVA, the fluctuation function develops a wavy structure on these time scales. This is the reason why we have only fitted straight lines to the log-log fluctuation function for the global records up to $\log_2 \tau = 8$, corresponding to 256 months or about 20 years. From the instrumental records, it is hard to verify if the LRM scaling holds for longer scales than this, and is a major motivation for analyzing longer records of reconstructed temperatures based on paleo proxies. Unfortunately, the multitude of published northern-hemisphere temperature reconstructions differ in the timing of fluctuations on decadal and multidecadal time scales, and also in the amplitude of the long oscillation of period approximately a millennium, encompassing the Medieval Warm Period (MWP) and the Little Ice Age (LIA). This ambiguity turns out to be a serious problem for establishing a reliable assessment of the LRM properties of the records on centennial time scales.

[37] We shall illustrate the issue by analyzing the longest existing paleo reconstruction of northern hemisphere temperatures [*Moberg et al.*, 2005]. This reconstruction spans the last two millennia (0–1979 A.D.) and is given with annual resolution, although it appears smooth on time scales less than a decade. We shall also give some consideration to another recent reconstruction [*Mann et al.*, 2009], which spans the somewhat shorter period 500–1850 A.D. The latter, however, is more heavily low-pass filtered so that the record appears smooth on time scales up to a few decades. This makes the range of scales available for scaling analysis smaller than for the Moberg record. Among the published reconstructions, the Moberg record has one of the largest amplitudes of the millennium oscillation, while the Mann record is in the lower end; the difference being roughly a factor two.

[38] The Moberg record itself is shown in Figure 10c along with the seventh order ($P_7$) polynomial fit. Lower-order polynomial fits give insignificant trends as shown by the variograms in Figure 10b. The undetrended variogram suggests a Hurst exponent of $H = 0.90$ and the corresponding line with slope $-\beta = -0.80$ is plotted in Figure 10a along with the periodogram of the undetrended and $P_7$-detrended signals. The periodogram of the $P_7$-detrended record (i.e., the millennium oscillation is subtracted from the record) displays a reduced power in the low-frequency part of the

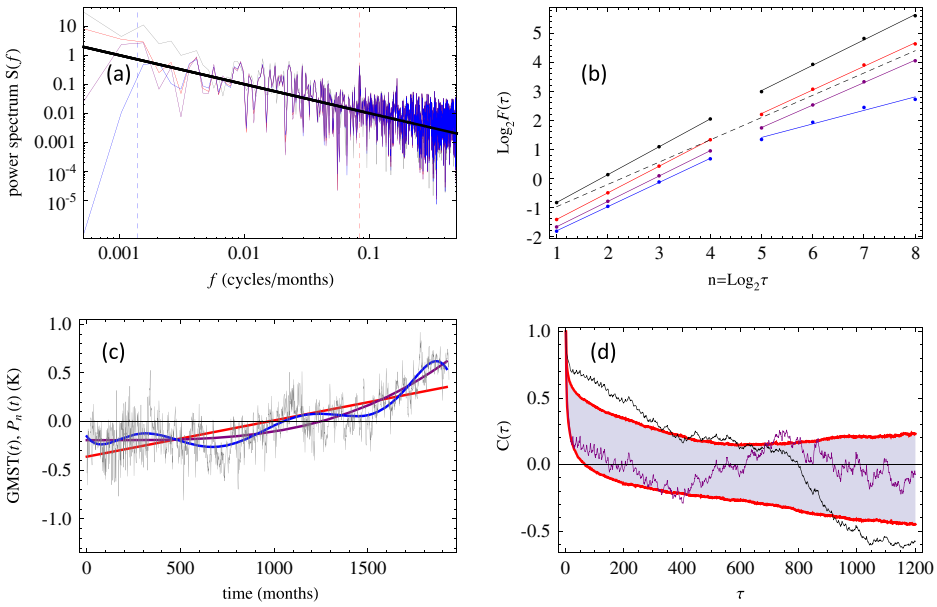


**Figure 9.** (a) PSD of monthly GMST record 1850–2010 A.D. Gray: undetrended. Red: $P_1$ detrended. Purple: $P_3$ detrended. Blue: $P_7$ detrended. Thick line has slope $-\beta = -1.0$, corresponding to $H = 1.0$. Vertical dashed lines mark the 60 year period (blue), and the 1 year period (red). (b) FA of the GMST record with variable degree of detrending. Black: after no detrending. Red: after $P_1$-detrending. Purple: after $P_3$-detrending. Blue: after $P_7$-detrending. The slopes for $n = 1, \ldots, 4$ correspond to the following: black: $H = 0.96$, red: $H = 0.92$, purple: $H = 0.88$, and blue: $H = 0.83$. The slopes for $n = 5, \ldots, 8$ are as follows: black: $H = 0.88$, red: $H = 0.82$, purple: $H = 0.78$, and blue: $H = 0.48$. Dotted line has slope 0.77 and is the same as the dotted line in Figure 7b. (c) Gray curve: Monthly GMST anomaly record 1850–2010 A.D. Red curve: $P_1$ fit. Purple: $P_3$-fit. Blue: $P_7$ fit. (d): Black: ACF estimate from undetrended GMLT record. Purple: ACF estimate from $P_3$-detrended record. The shaded area represents the 95% confidence interval for the ACF computed from an ensemble of 5000 realizations of fBms of the same length as the GMST record and with $H = 0.99$.

spectrum (i.e., for periods above 250 years). This creates a flat variogram for $\tau > 2^8$ years and makes it difficult to fit a straight line to any extended range of time scales $\tau$ in the variogram as shown by the blue, dotted curve in Figure 10b. This detrending obviously does not remove only the actual trend, but also the low-frequency part of the fGn noise background, and this may be the cause of apparent lack of scaling of the detrended signal. On the other hand, it may not be obvious from this analysis to which extent the millennium oscillation should be interpreted as a trend, an inherent part of the noise, or a combination of the two. The fact that the power in the frequency corresponding to the 1000 year period in the PSD of the undetrended signal is considerably above the fit-line in Figure 10a suggests that all of the power in this mode cannot be a part of the noise. This is confirmed by the observation in Figure 10d that the biased ACF estimated from the undetrended record is outside the confidence limits for the biased ACF estimates for samples from the Monte Carlo ensemble of fGns with $H = 0.9$ (and this will also be the case for records with detrending lower than seventh order). This means that the millennium oscillation has too large amplitude to be consistent with a $H = 0.9$ noise process.

[39] If the Moberg record can be modeled as a millennium-oscillation trend similar to the blue curve in Figure 10c superposed on an fGn with Hurst exponent of the magnitude derived from WVA, FA, or DFA, it should be possible to subject Monte Carlo realizations of such a model to the same analysis as the observation data. The result of these analyses should then agree within the established confidence limits for the respective methods. We have done this analysis as follows. First we produce a wavelet-filtering of the Moberg signal which is similar to the $P_7$ polynomial fit shown in Figure 10c, but believed to be a somewhat better representation of the millennium oscillation trend. The detrended signal is obtained from subtracting this filtered signal from the original record. We then produce a synthetic Moberg signal consisting of this trend superposed on a realization of an fGn with $H = 0.87$ and variance equal to that of the detrended Moberg record. The results of the WVA and FA applied to this signal and the observed Moberg record are shown as the black and red curves in Figures 11a and 11b. The same analyses have been applied to the fGn realization and the detrended Moberg record in Figures 11c and 11d. There is an overall good agreement between the analysis results for the synthetic records and the observed ones. Since the red curves in Figures 11c and 11d are results from analysis of a synthetic fGn, it is clear that most of the apparent loss of scaling for large $\tau$ in the variogram for the detrended record is a feature of the FA method applied to a short record and not a loss of LRM on these scales.

[40] Figure 12a shows results of the DFA applied to the Moberg reconstruction, and Figure 12b to the detrended signal. This should be compared to the DFA of the synthetic signal with the trend superposed on the fGn with $H = 0.87$ in Figure 12c and of the fGn itself in Figure 12d. For the

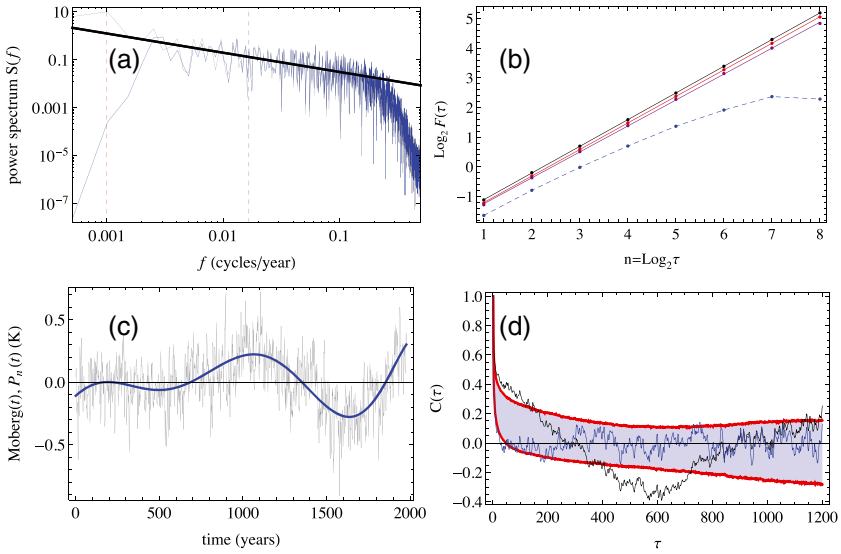


**Figure 10.** (a) PSD of undetrended Moberg record 0–1979 A.D. (gray) and of $P_7$-detrended record (blue). Thick line has slope $-\beta = -0.80$, corresponding to $H = 0.90$. Vertical dashed lines mark the 60 year period (blue), and the 1000 year period (red). (b) Variogram of the "profile" $y(t)$ of the Moberg record with variable degree of detrending. Black: after no detrending. Red: after $P_1$-detrending. Purple: after $P_3$-detrending. Blue: after $P_7$-detrending. The slope of the black line is: $H = 0.90$, red line: $H = 0.89$, and purple line: $H = 0.88$. (c) Gray: Moberg record 0–1979 A.D. in degrees Kelvin. Blue: $P_7$ fit . (d) Black: ACF estimate from undetrended Moberg record. Blue: ACF estimate from $P_7$-detrended record. The shaded area represents the 95% confidence interval for the biased ACF computed from an ensemble fGns of the same length as the Moberg record and with $H = 0.90$.

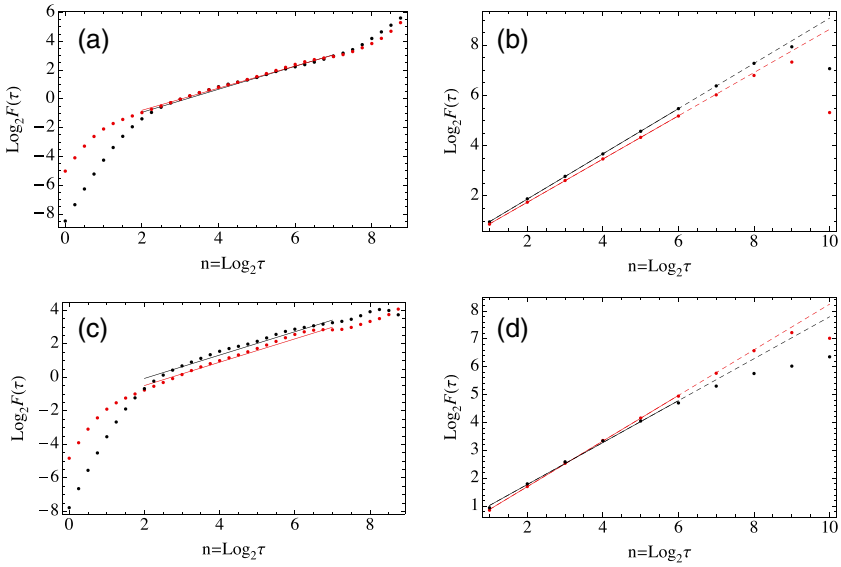


**Figure 11.** (a) WVA of undetrended Moberg record (black) and of synthetic record consisting of wavelet-filtered signal plus fGn with $H = 0.87$ (red). The slope of the black line is $\beta = 0.80$ ($H = 0.90$) and of the red line $\beta = 0.76$ ($H = 0.88$). (b) FA of the same signals as in Figure 11a. The slope of the black line is $H = 0.90$ and of the red line $H = 0.86$. (c) WVA of Moberg record detrended by subtraction of wavelet-filtered signal (black) and of synthetic record consisting of fGn with $H = 0.87$ (red). The slopes of the black and red lines are both $\beta = 0.70$ ($H = 0.85$). (d) FA of the same signals as in Figure 11c. The slopes of the black and red lines are $H = 0.75$ and $H = 0.82$, respectively. In the FA-curves the fit has been made in the range of the full lines. The dashed lines are continuation of these lines to help visualize the departure from scaling for large $\tau$.

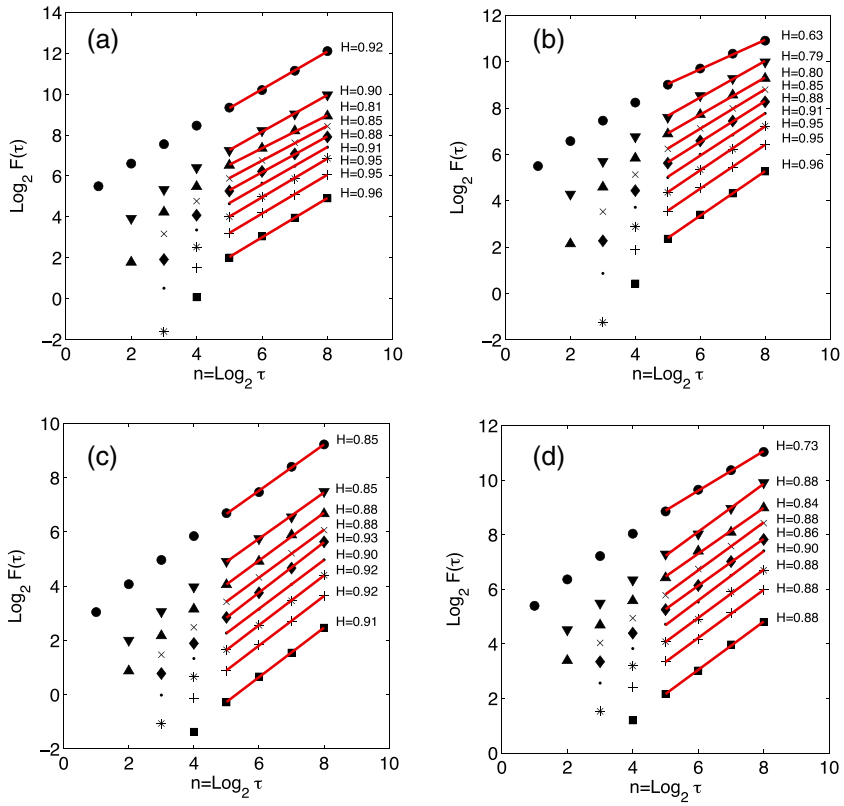


**Figure 12.** (a) DFA0–8 of undetrended Moberg record. The upper curve is the fluctuation function for DFA0, the ones below are DFA$n$, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFA$n$. (b) DFA0–8 of Moberg record detrended by subtraction of wavelet-filtered signal. (c) DFA0–8 of synthetic record of wavelet-filtered Moberg record superposed on synthetic fGn with $H = 0.87$. (d) DFA0–8 of synthetic fGn with $H = 0.87$.

observed signals in Figures 12a and 12b, there is an increasing slope for increasing order of the DFA, which is related to a downward curving of the fluctuation function for small $\tau$. This anomaly is obviously caused by the smoother character of the Moberg record for time scales less than a few decades, as is clearly observed in the PSD shown in Figure 10a. It is the same feature of the Moberg record that yields the discrepancy for small $\tau$ between the observed and synthetic signals in the WVA shown in Figures 11a and 11c. Another clear anomaly is the large slope of DFA0 for the signals with trend, and the corresponding smaller slope of DFA0 for the detrended signals. The former is due to the effect of the trend on DFA0, the latter is due to a negative bias on estimates like FA and DFA0. This bias will be discussed in the next section. For DFA1–4 the results in Figures 12a–12d are similar and consistent with the value $H = 0.87$, but with some random scatter in the estimated $H$-values due to the previously discussed errors associated with short records.

[41] We should bear in mind that the analysis on synthetic records here has been done on one arbitrary realization of the fGn. Estimates of $H$ from other realizations will give somewhat different results. In section 3, we estimated biases and error bars for WVA, FA, and DFA applied to ensembles of realizations of fGns containing 2000 data points. The length of 2000 data points in the record was chosen because the global instrumental time series contain approximately 2000 data points with monthly resolution. The Moberg record also contains nearly 2000 data points with annual resolution, but from the periodogram in Figure 10a, we observe a strong depletion of the spectrum for high frequencies, indicating that the record is smooth on scales less than 4 years. This is also the reason why the fluctuation function for WVA of the Moberg signal in Figure 11a (black dots) is depleted for $\tau < 2^2$. This means that the meaningful sampling interval for the study of the scaling properties of this time record is 4 years, and hence that the "real" length of the record is about 500 data points. Hence, in a Monte Carlo study relevant for this record, we should generate fGns of this length. The result of such a study for FA and WVA yields results similar to those shown in Figure 2, but with larger errors and larger negative bias for FA. The negative bias of estimated $\beta$ for FA is now –0.30 when $\beta$ approaches unity, and the error for both FA and DFA is approximately $\pm 0.20$. Since $H = (\beta + 1)/2$, the corresponding figures for the Hurst exponent is a bias of –0.15 for FA and error for $H$ of $\pm 0.10$. From these bias and error estimates it makes little sense to give the estimate of $H$ for the detrended Moberg record with more than one decimal, i.e., our best estimate is $H = 0.9 \pm 0.1$ on scales up to $\tau \approx 250$ years.

[42] A more rigorous approach to this problem is to make Monte Carlo simulations with fGns properly filtered to yield a PSD similar to that of the Moberg record. We do this by wavelet filtering and show the WVA fluctuation function of the filtered signal as the red dots in Figure 13a. The dots are the mean values computed from an ensemble of filtered fGns with $\beta = 0.75$ ($H = 0.875$). The error bars are the 95%

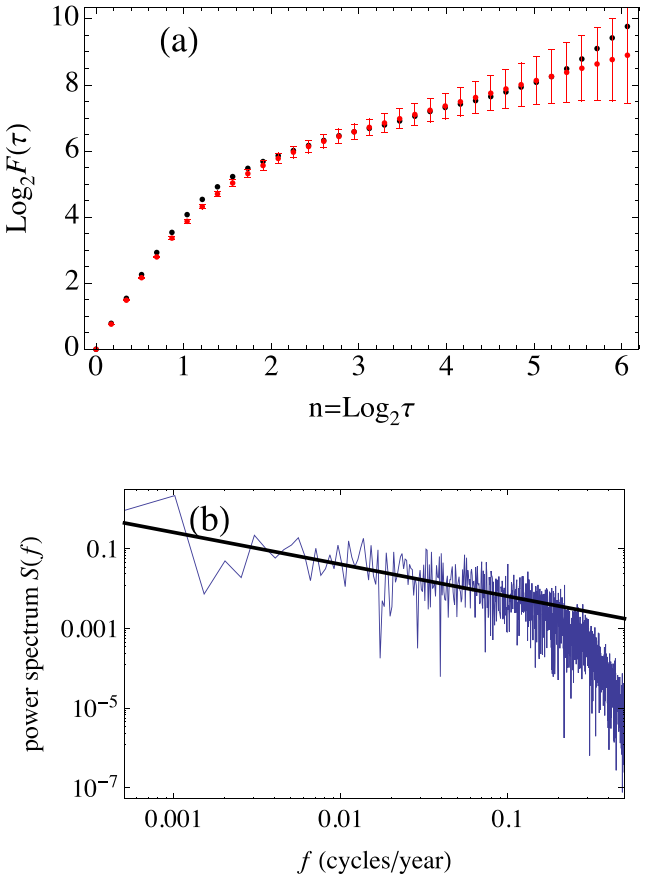


**Figure 13.** (a) Black dots: WVA fluctuation function for the Moberg record. Red dots: Mean WVA fluctuation function for an ensemble of filtered fGns with $\beta = 0.75$ ($H = 0.875$). The error bars indicate the 95% confidence intervals estimated from the ensemble. (b) Periodogram for one realization of the filtered fGn. The black straight line has slope –0.8, corresponding to $H = 0.9$.

confidence intervals computed from this ensemble. The black dots is the fluctuation function computed from the Moberg signal, and is the same as the black dots in Figure 11a. Recall that the red dots in Figure 11c represent the fluctuation function for the unfiltered fGns and note how the filtering makes the fluctuation function coincide with that of the Moberg record for small time lags. The same is seen in the PSD of a realization of the filtered fGn shown in Figure 13b, which should be compared to the PSD of the Moberg record in Figure 10a.

[43] We can now repeat the estimates of biases and errors with a full length record of nearly 2000 data points, of filtered fGns. The result of the WVA for such an ensemble agrees very well with those of unfiltered fGns of length 500 data points, both in the small bias and the estimated error bars. The results of the ACF estimates in Figure 10d are not influenced noticeably by the filtering since the fluctuations on the shortest time scales have little impact on the correlations on longer time scales. A weakness of that analysis, however, is that it is done for one specific value of $\beta$, which is the $\beta$ estimated from the WVA. But as our analysis shows, this estimate has an uncertainty of $\pm 0.2$, so we cannot be certain that we have used the right $\beta$ to test if the millennium oscillation in the Moberg record can be reconciled with the fGn hypothesis. Below we shall show that we can do better.

[44] The procedure is as follows: for the observed record $x_{\rm obs}(t)$, we estimate $\beta_{\rm obs}$ from the WVA method. The estimate is $\hat{\beta}_{\rm obs}$. By low-pass wavelet filtering, we find an estimated millennium-oscillation trend $x^T_{\rm obs}(t)$ and we characterize the strength of the trend by means of the range $\hat{r}_{\rm obs} \equiv \max(x^T_{\rm obs}) - \min(x^T_{\rm obs})$. Then we generate numerically an ensemble of appropriately filtered fGns of length equal to that of the Moberg record (1978 data points) and repeat this procedure for each realization in the ensemble. The true $\beta$ values for the synthetic fGns are drawn at random from a prior probability density distribution $p(\beta)$. From this ensemble, we can establish a conditional joint PDF $p(\hat{r}, \hat{\beta}|\beta)$ and the joint distribution of estimated $\hat{r}, \hat{\beta}$ is

$$p(\hat{r}, \hat{\beta}) = \int (p(\hat{r}, \hat{\beta}|\beta) p(\beta) {\rm d}\beta. \qquad (13)$$

Our knowledge prior to the analysis in this section is that the observed record can be described by an fGn, possibly superposed on an oscillatory trend. We also know an estimate $\hat{\beta}_{\rm obs}$ and the PDF for this estimate derived from an ensemble of fGns generated with $\beta = \hat{\beta}_{\rm obs}$. If the prior distribution $p(\beta)$ is chosen to be this PDF, the joint PDF given by equation (13) should be interpreted as the likelihood of observing the pair $(\hat{r}, \hat{\beta})$ provided the null hypothesis; that the signal is an fGn without a trend, is true. In Figure 14, we have drawn the contour of constant $p(\hat{r}, \hat{\beta})$ that separates an internal region for which the total probability is 0.95 from an external region for which it is 0.05. It shows that if the WVA yields a small estimated memory exponent $\hat{\beta}$ it is unlikely that the low-pass wavelet filtering will estimate a spurious trend with a high range parameter $\hat{r}$, but at high $\hat{\beta}$ it is more likely that the estimates return a spurious strong trend. If the WVA and wavelet detrending of the observed record yield $(\hat{r}_{\rm obs}, \hat{\beta}_{\rm obs})$ lying in the region where $p(\hat{r}, \hat{\beta})$ is large, it is not possible to conclude that there is a real trend, i.e., we cannot falsify the null hypothesis that no trend exists. On the other hand, if $(\hat{r}_{\rm obs}, \hat{\beta}_{\rm obs})$ is well outside this region, the null hypothesis

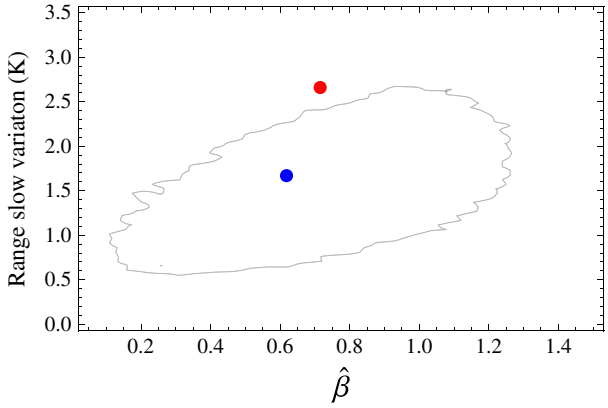


**Figure 14.** The closed contour indicates the line of constant $p(\hat{r}, \hat{\beta})$ inside which the integrated probability is 0.95. The red dot is the estimated $(\hat{r}_{\rm obs}, \hat{\beta}_{\rm obs})$ for the Moberg record. The blue dot is the same estimate for the modified Moberg record where the millennium-oscillation amplitude is reduced by a factor two.

is falsified, and we have to conclude that there exists a significant trend that goes beyond the fGn model. The red dot in the plot is $(\hat{r}_{\rm obs}, \hat{\beta}_{\rm obs})$ for the Moberg record, and since it is located outside the confidence region, it confirms our result from previous sections that millennium oscillation in this record is incompatible with the fGn null hypothesis, and hence is a significant trend. But the red dot is not very far outside the 95% confidence region, and it is therefore imperative to investigate how sensitive this result is to the estimated amplitude of the millennium oscillation.

[45] As mentioned in section 3, the millennium oscillation in the Moberg record has larger amplitude $r$ than in most other reconstructions. For instance, it is about twice the amplitude of the more recent reconstruction by *Mann et al.* [2009]. A crucial question is then if the significance of this millennium trend will survive if the range $r$ of the oscillation is reduced by a factor two. We produce such a signal by adding a signal corresponding to the wavelet-filtered trend-signal, but with half its amplitude, to the "detrended" signal. The WVA fluctuation function of this modified Moberg record coincides with the ensemble mean of the WVA fluctuation function of the filtered fGns shown by the red dots in Figure 11a, which already indicates that reducing the amplitude in the slow oscillation by a factor two makes the result consistent with an fGn. This conclusion is enforced by computing $\hat{r}, \hat{\beta}$ for the modified record. The result is marked as the blue dot on Figure 14b and falls close to the center of the joint distribution.

## 6. Discussion and Conclusions

[46] In this paper we have employed non-parametric detrending techniques on regional and global surface temperature records. These techniques should be considered complementary to the more model-dependent parametric statistical methods. The results obtained confirm the existence of strong ($H \approx 1$) long-range memory in the global temperature records on time scale from months and at least up to several centuries obtained by non-parametric methods [*Rybski et al.*, 2006], and on scales from months to decades by parametric methods [*Gil-Alana*, 2005]. The error bars ($\pm 0.07$) obtained on these estimates are due to the short lengths of the records and not strongly dependent on the analysis technique. Ensembles of numerical realizations of the same fGn process with 2000 data points shows considerable diversity and is an unsurmountable source of uncertainty when it comes to estimating the memory parameter from a single realization. The results further suggest that the LRM is more pronounced in global than in local records, and more pronounced in ocean records than in land records. They also suggest that the LRM is associated with the thermal inertia of the oceans, and not only the inertia of the ocean mixed layer. Response times longer than a decade must involve overturning circulations that couple the mixed layer to the deep ocean [*Delworth et al.*, 1993].

[47] It is well known that aggregation of AR(1) processes with a wide distribution of lag-one autocorrelations $\phi$ can give rise to a long-memory process [*Granger*, 1980]. This can be the case even if the individual processes are independent. In principle, this could explain the emergence of LRM as local temperature records are merged into a global record. However, as we will demonstrate in a forthcoming paper, there is strong evidence that even local temperature records exhibit LRM, so the problem to deal with is rather aggregation of relatively weakly persistent LRM-processes to produce a strongly persistent fGn. On the other hand, the various subsystems of the climate system (atmosphere, ocean mixed layer, deep ocean, sea ice, etc.) may exhibit exponential response functions with varying time constants, whose aggregation may produce an LRM-response on the global scale. These are challenging issues for future research.

[48] Our analysis confirms that the rising temperature trend over the last century is too strong to be consistently described as part of the LRM process associated with undriven climate variability [*Schlesinger and Ramankutty*, 1994]. However, the 60 year oscillation that is observed in these records, and especially strong in the ocean SST record, is explicable as a natural LRM fluctuation, and does not have to be externally driven.

[49] In *Rypdal* [2012], the deterministic version of equations (1) and (3) were studied for a prescribed forcing record $F(\tau)$, but without any stochastic forcing. The result can be interpreted as the non-stochastic response to this forcing, i.e., in one specific meaning of the word, as a trend. The deterministic response signal shown in that paper appears rather "noisy" in the sense that it contains some sawtooth-like spikes. These are the responses to forcing from volcanic eruptions, which are present in the deterministic parts of the forcing. Hence, with this definition, trends do not have to be slow, and this makes the detection problem more difficult. However, it helps a lot if we have knowledge about the forcing $F(t)$ that gives rise to the trend. The traditional approaches to detecting long-range memory in climate records is to disregard the available information about the deterministic forcing function $F(t)$ and analyze the signal as if it is the response to the stochastic forcing superposed on some hypothesized slow trend. One such approach is to assume that the response to the deterministic forcing can be described by a low-order polynomial, and that the stochastic second term of the solution is an LRM process. In the present paper we have employed some of these techniques to regional and global instrumental temperature records with emphasis on establishing proper confidence limits on the estimates of memory exponents.

[50] The results of our analysis of the Moberg reconstruction are consistent with those obtained by *Rybski et al.* [2006] and provide proper error bars which imply that the actual Hurst exponent for the Moberg record is in the interval $0.8 < H < 1.0$, with the most probable value $H = 0.9$. The millennium-oscillation trend consistent with this estimate is given as the blue curve in Figure 10a. The last half-period of this oscillation coincides approximately with the period and phase of a number of reconstructions of total solar irradiance based un sunspot number observations, which do not go further back than to the early 17th century, and hence may incorporate the Maunder minimum and the LIA, but do not extend back to the MWP [*Gray et al.*, 2010]. However, a number of more recent multiproxy reconstructions, which extend back to 850 A.D., show much higher amplitudes of an oscillation with period of roughly 200 years than of the millennium-period oscillation [*Schmidt*, 2011]. This period is not very prominent in the Moberg record, so it may be difficult to explain the millennium oscillation exclusively as an effect of solar variability on the basis of these TSI

reconstructions. It is not our ambition in this paper to provide a physical explanation of the millennium oscillation in the Moberg reconstruction of northern hemisphere temperature, but one cannot disregard the possibility that this reconstruction overestimates its amplitude. Reducing this amplitude by a factor of two will bring it in more in line with the majority of other reconstructions, and then the null hypothesis; that the millennium oscillation is an inherent part of the LRM noise, and cannot be rejected. This means that, unless we use information about the forcing record, it will not be possible to settle with any certainty the issue of whether this oscillation is an LRM fluctuation or a forced variation of the global climate. Fortunately, forcing reconstructions for the last millennium exists, and using it to settle this issue will be addressed in a forthcoming paper.

[51] **Acknowledgments.** The authors are grateful to Ola Løvsletten for illuminating discussions and for contributing to a numerical routine for accurate generation of fractional Gaussian noises.

## References


Beran, J. (1994), Statistics for Long-Memory Processes*, Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Boca Raton.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *Geophys. Res. Lett.*, *111*, D12106, doi:10.1029/2005JD006548.

Bunde, A., S. Havlin, E. Koscielny-Bunde, and H.-J. Schellnuber (2001), Long term persistence in the atmosphere: Global laws and tests of climate models, *Physica A*, *302*, 255–267.

Bunde, A., and S. Havlin (2002), Power-law persistence in the atmosphere and in the oceans, *Physica A*, *314*, 15–24.

Delworth, T., S. Manabe, and R. J. Stouffer (1993), Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model, *J. Climate*, *6*, 1993–2011.

Eichner, J. F., E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber (2003), Power-law persistence and trends in the atmosphere: A detailed study of long temperature records, *Phys. Rev. E*, *68*, 046133, doi:10.1103/PhysRevE.68.046133.

Efstathiou, M. N., C. Tzanis, A. P. Cracknell, and C. A. Varotsos (2011), New features of land and sea surface temperature anomalies, *Int. J. Remote Sensing*, *32*, 3231–3238.

Flandrin, P. (1992), Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. Inform. Theory*, *48*, 910–917.

Franzke, C., T. Graves, N. W. Watkins, R. B. Gramacy, and C. Hughes (2012), Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, *Phil. Trans. R. Soc. A*, *370*, 1250–11267.

Gil-Alana, L. A. (2005), Statistical modeling of the temperatures in the northern hemisphere using fractional integration techniques, *J. Climate*, *18*, 5357–5369.

Granger, C. W. J. (1980), Long memory relationships and the aggregation of dynamic models, *J. Econometrics*, *14*, 227–238.

Gray, L. J., et al. (2010), Solar influences on climate, *Rev. Geophys.*, *48*, RG4001, doi:10.1029/2009RG000282.

Govindan, R. B., D. Vyushin, A. Bunde, S. Brenner, S. Havlin, and H.-J. Schellnuber (2002), Global climate models violate scaling of the observed atmospheric variability, *Phys. Rev. Lett.*, *89*, 028501, doi:10.1029/2009RG000282.

Govindan, R. B., A. Bunde, and S. Havlin (2003), Volatility in atmospheric temperature variability, *Physica A*, *318*, 529–536.

Herrmann, R. (2011), *Fractional Calculus. An Introduction for Physicists*, World Scientific Publishing Company, Singapore.

Intergovernmental Panel of Climate Change (2007), *Climate Change 2007: The Physical Science Basis,* Solomon, S., D. Quin, and M. Manning (eds), Geneva. (Available at http://ipcc-wg1.ucar.edu/wg1/wg1-report.html).

Király, A., I. Bartos, and I. M. Jánosi (2006), Correlation properties of daily temperature anomalies over land, *Tellus, Ser. A*, *58*(5), 593–600.

Koscielny-Bunde, A. B., S. Havlin, and Y. Goldreich (1996), Analysis of daily temperature fluctuations, *Physica A*, *231*, 393–396.

Koscielny-Bunde, A. B., S. Havlin, H. E. Roman, Y. Goldreich, and H-J. Schellnhuber (1998), Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.*, *81*, 729–732.

Lennartz, S., and A. Bunde (2009a), Trend evaluation in records with long-term memory. Application to global warming, *Geophys. Res. Lett.*, *36*, L16706, doi:10.1029/2009GL039516.

Lennartz, S., and A. Bunde (2009b), Eliminating finite-size effects and detecting the amount of white noise in short records with long-term memory, *Phys. Rev. E*, *79*, 066101, doi:10.1103/PhysRevE.79.066101.

Malamud, B. L., and D. Turcotte (1999), Self-affine, time series: I. Generation and analyses, *Adv. Geophys.*, *10*, 1–90.

Manley, G. (1974), Central England temperatures: Monthly means 1659 to 1973, *Q. J. R. Meteorol. Soc.*, *100*, 389–405.

Mann, M. E., and J. M. Lees (1996), Robust estimation of background noise and signal detection in climatic time series, *Clim. Change*, *33*, 409–445.

Mann, M. E., and J. Park (1994), Global-scale modes of surface temperature variability on interannual to century timescales, *J. Geophys. Res.*, *99*, 819–825.

Mann, M. E., et al. (2009), Global signatures and dynamical origins of the little ice age and medieval climate anomaly, *Science*, *326*, 1256–1260.

Mann, M. E., and J. Park (1999), Oscillatory spatiotemporal signal detection in climate studies: A multitaper spectral domain approach, *Adv. Geophys.*, *41*, 1–131.

McLeod, A. I., H. Yu, and Z. L. Krougly (2007), Algorithms for linear time-series analysis, *J. Stat. Softw.*, *23*, 1–25.

Moberg, A., et al. (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, *433*, 613–617.

Monetti, R. A., S. Havlin, and A. Bunde (2003), Long-term persistence in the sea surface temperature fluctuations, *Physica A*, *320*, 581–589.

Padilla, L. E., G. K. Vallis, and G. K. Rowley (2011), Probabilistic estimates of transient climate sensitivity subject to uncertainty in forcing and natural variability, *J. Climate*, *24*, 5522.

Parker, D. E., T. P. Legg, and C. K. Folland (1992), A new daily, Central England Temperature series, 1772–1991, *Int. J. Clim.*, *12*, 317–342.

Pelletier, J. D., and D. Turcotte (1999), Self-affine time series: II. Generation and analyses, *Adv. Geophys.*, *10*, 91–166.

Rybski, D., A. Bunde, S. Havlin, and H. von Storch (2006), Long-term persistence in climate and the detection problem, *Geophysical. Res. Lett.*, *33*, L06718, doi:10.1029/2005GL025591.

Rypdal, M., and K. Rypdal (2010), Testing hypotheses about sun-climate complexity inking, *Phys. Rev. Lett.*, *104*, 128501.

Rypdal, K. (2012), Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions, *J. Geophys. Res.*, *117*, D06115, doi:10.1029/2011JD017283.

Rypdal, M., and K. Rypdal (2012), Is there long-range memory in solar activity on time scales shorter than the sunspot period? *J. Geophys. Res.*, *117*, A04103, doi:10.1029/2011JA017283.

Scafetta, N. (2010), Empirical evidence for a celestial origin of the climate oscillations and its implications, *J. Atmos. Sol.-Terr. Phy.*, *72*, 951–970, doi:10.1016/j.jastp.2010.04.015.

Scafetta, N. (2011a), A shared frequency set between the historical mid-lattitude aurora records and the global surface temperature, *J. Atmos. Sol.-Terr. Phy.*, *74*, 145–163, doi:10.1016/j.jastp.2011.10.013.

Scafetta, N. (2011b), Testing an astronomically based decadal-scale empirical harmonic climate model versus the IPCC (2007) general circulation models, *J. Atmos. Sol.-Terr. Phy.*, *80*, 124–137, doi:10.1016/j.jastp.2011.12.005.

Schlesinger, M. E., and N. Ramankutty (1994), An oscillation in the global climate system of period 65–70 years, *Nature*, *367*, 723–726.

Schmidt, G. A., et al. (2011), Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, *4*, 33–45, doi:10.5194/gmd-4-33-2011.

Vallis, G. K. (2012), Climate and the Oceans*, Princeton Primers in Climate*, Princeton University Press, Princeton.

Varotsos, C., and D. Kirk-Davidoff (2006), Long-memory processes in ozone and temperature variations at the region 60° S–60° N, *Atmos. Chem. Phys.*, *6*, 4093–4100.

Vjushin, D, R. B. Govindan, S. Brenner, A. Bunde, S. Havlin, and H.-J. Schellnhuber (2002), Lack of scaling in global climate models, *J. Phys.:Condens. Matter*, *14*, 2275–2282.

Vyushin, D. I., P. J. Kushner, and F. Zwiers (2012), Modeling and understanding persistence of climate variability, *J. Geophys. Res.*, *117*, D21106, doi:10.1029/2012JD018240.

Weber, R. O., and P. Talkner (2001), Spectra and correlations of climate data from days to decades, *J. Geophys. Res.*, *106*, 20131–20144.

Wilk, M. B., and R. Gnanadesikan (1968), Probability plotting methods for the analysis of data, *Biometrika*, *55*, 117.