UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

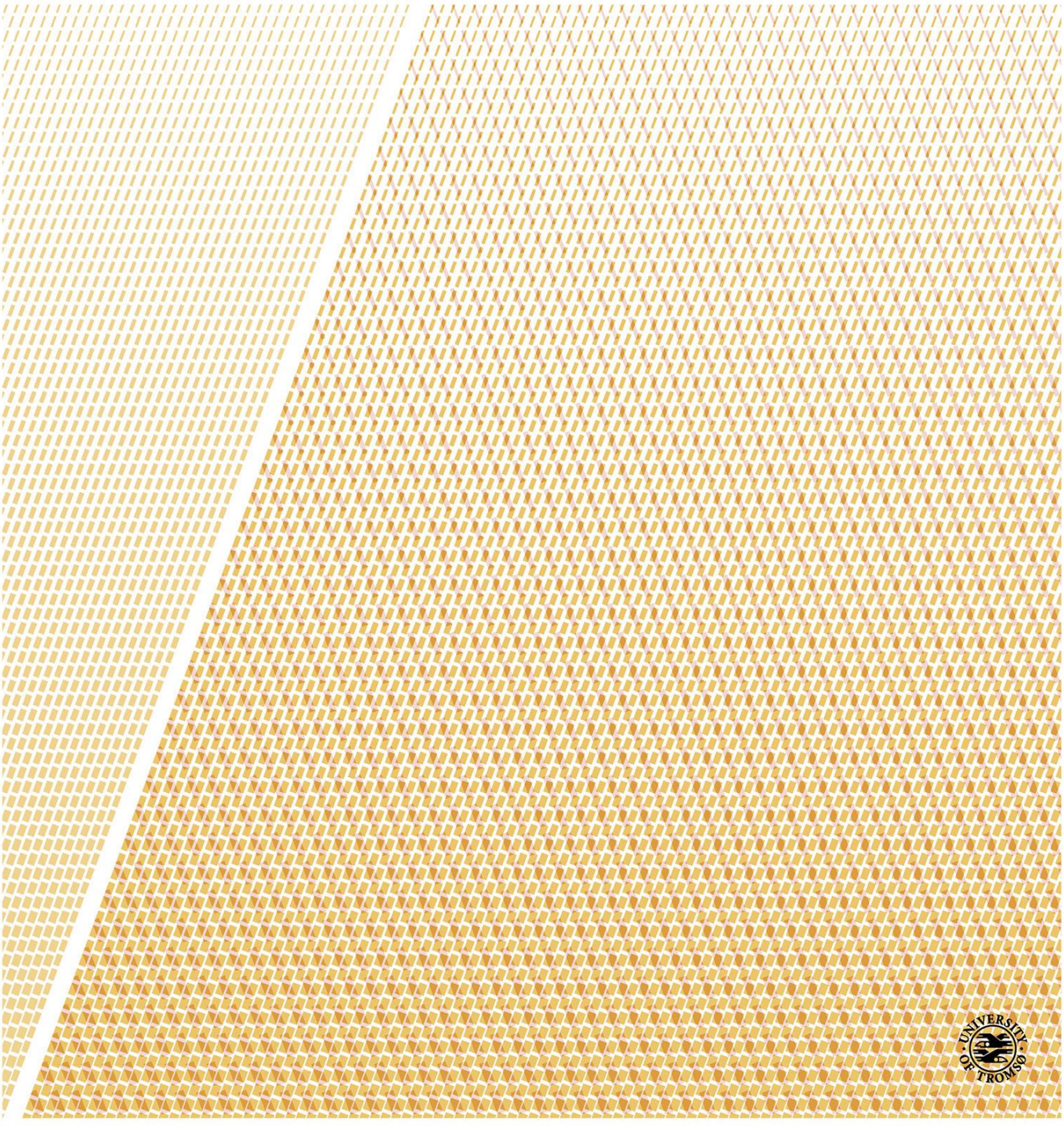
Faculty of Science and Technology

Long Range Memory in Time Series of Earth Surface Temperature

—

Lene Østvand

A dissertation for the degree of Philosophiae Doctor – February 2014



Abstract

Long-range memory (LRM) has been found in numerous natural data records, both in geophysics and other fields. In this thesis LRM in surface temperature time series is studied. Short-range memory (SRM) models, especially the first order auto-regressive model AR(1), have been widely used to describe geophysical data, and in the more recent years, SRM and LRM models have been compared. SRM is therefore also included in this thesis.

Trends are important in climate studies, but the trend definition is ambiguous. Two different approaches are included here: the trend described as a function with parameters estimated from the data, and a response model to external forcing. The most commonly used trend function is the linear trend, often used as a measure of anthropogenic effects on global warming. This approach is illustrated by the application to a local instrumental temperature record from Cheyenne (Wyoming, USA). The significance of the trend is dependent on the noise model assumed to describe the data, and here AR(1) and fractional Gaussian noise (fGn) are used. In the example of the Cheyenne record, the trend could not be explained as natural variations in any of the two models. Since the forcing data available are global, the response model is applied to global land temperature, with AR(1) and fGn used as models for the stochastic response.

Some methods for investigating SRM and LRM are described, and the Cheyenne record used as an example of applications, showing how a linear trend affects the analysis. In this section, comparison of SRM and LRM is also included, and applied to local temperature records from Cheyenne and Prague (the Czech Republic) and the Niño3 index (area averaged sea surface temperature over Niño region 3) to find which model best describes the data. The results suggest that temperature is best described as an fGn on large time scales. The Niño3 index is not perfectly described by any of the noise models, but AR(1) is a better statistical model than fGn. Application to the response model approach shows that fGn is a far better model than AR(1) for the stochastic response in the case of the global land temperature.

The thesis also includes a literature review. In the scientific literature mostly local temperature records have been analysed with regards to LRM. Global and hemispheric temperature means are far less studied, so this has been our main focus. In Paper I, the LRM properties of local and global instrumental records and a Northern Hemisphere temperature record were studied after detrending with different polynomial trend models. LRM was found on a wide range of time scales, but different trend models were needed for the different records to yield the best scaling

properties. Proper error bars for LRM scaling exponents were an important part of this study and represents improvement of previous work. In Paper II, the significance of trends in global ocean and land temperature was investigated under three null models for noise, i.e., AR(1), fGn and fBm. A linear trend was clearly significant in the land temperature, and incorporating this in the null model showed that an oscillation also stands out from the natural variations that could be explained by the noise models. The significance of trends was not so apparent for the ocean temperature, but an AR(1) noise model could be rejected. Temperature from different climate model experiments was studied in Paper III, including control runs and experiments with full dynamic forcing. Two temperature reconstructions were also analysed for comparison with the simulated temperatures. Scaling properties in agreement with persistent LRM noise was found for a wide range of scales for most of the simulated temperatures. The temperature from the control runs and the runs with dynamic forcing showed similar scaling exponents. Only the HadCM3 control run differed from the other climate model experiments, yielding a temperature with a clear cross-over from a motion to a persistent noise.

The overall conclusion that can be drawn from the present work is that long-range persistence on time scales from years to centuries is ubiquitous in observed Earth surface temperature records, and that similar persistence is present in the most advanced climate models to date. This persistence weakens the significance of observed temperature trends, but not enough to render the rising temperature trends throughout the last century statistically insignificant.

Acknowledgements

I would like to acknowledge UiT The Arctic University of Norway and both the Department of Physics and Technology and the Department of Mathematics and Statistics for giving me the opportunity to study an interesting topic.

My warmest thanks go to my supervisor Kristoffer Rypdal. I am grateful for the opportunity to go to interesting workshops, summer schools and conferences abroad during my PhD, but most of all for all inspiration, encouragement and support. It is truly motivating to have a supervisor who is always eager to learn, discuss and work on new problems in science, regardless of place and time of the day.

I also thank my co-supervisor Chris Hall for his contributions early in my PhD.

I thank my whole research group for interesting discussions, travels and general support, and especially
Tanja for being a role model, office mate and friend, and making conferences in San Francisco and Vienna more fun.
Martin and Tine for their contributions to my journal articles.
Tine and Hege for the good company on conference trips.
Ola for helping me learn R and making me feel welcome at the Department of Mathematics and Statistics.

Happy thoughts are sent to everyone involved in TODOS for improving the PhD life. Keep up the good work!

I am grateful to my friends and family for their support and for reminding me that there is more to life than a PhD.

Contents

Chapter 1

Introduction

1.1 Motivation

The first studies of long-range memory (LRM) include *Hurst* (1951), who studied the hydrology of the Nile river. *Hurst et al.* (1965) developed the rescaled range (R/S) analysis where the scaling parameter H, known as the Hurst exponent, was found through a power-law relation $\frac{R}{S} = \left(\frac{T}{2}\right)^H$. T is the time scale, R is the range and S is the standard deviation during this time. This power-law relation was first found in river flow, but also in lake levels, thicknesses of tree rings and varves, atmospheric temperature and pressure, and sunspot numbers. *Mandelbrot and Wallis* (1969) suggested adjustments to the (R/S) analysis with applications to a large number of geophysical data, discarding some of the values for the scaling exponent obtained by *Hurst et al.* (1965), but confirming the presence of LRM. Since then, various methods have been developed to investigate scaling properties and estimate a scaling parameter. Many studies focus on the performance of the methods when applied to processes which are known to be scale invariant with a known scaling exponent (*Heneghan and McDarby*, 2000, *Weron*, 2002, *Delignieres et al.*, 2006, *Mielniczuk and Wojdyłło*, 2007, *Franzke et al.*, 2012). Geophysical time records, however, often exhibit deviations from pure scale-invariance which influence the estimation of the scaling parameter. Trend estimation and detrending have therefore become important in LRM studies. The time series is then modelled as a trend superposed on an LRM noise process, but inherent in such modelling are ambiguities concerning how to separate the trend and the noise. The statistical significance of the trend depends on how one models the noise process against which the trend is tested, as some noise processes naturally have slow variations which may be falsely regarded as trends, while others do not. The separation of the trend from the noise is an issue that needs to be resolved, and one needs to clarify whether time series which appear to have long-range memory really do so, or if they can be better described as time series with short-range memory with superposed trends.

A number of studies suggest that atmospheric temperature have long-range memory, and a review of these is given in Chapter 5. Although there is increasing evidence of the presence of LRM in such time series, a precise physical explanation of the phenomenon has been elusive. Studies

of instrumental data indicate that sea surface temperature is more persistent than air temperature over land (*Pelletier*, 1997, *Eichner et al.*, 2003, *Monetti et al.*, 2003, *Lennartz and Bunde*, 2009), so ocean dynamics seem to be an important component. Spatial averaging also influences the persistence, as temperature averages over larger regions are more persistent than local data (*Lennartz and Bunde*, 2009). Some studies also indicate that persistence is large close to the equator, and is reduced with location closer to the poles (e.g., *Pattantyús-Ábrahám et al.*, 2004, *Huybers and Curry*, 2006, *Vyushin and Kushner*, 2009). There are some studies of temperature from model experiments that suggest that several types of dynamic forcing must be included to find scaling exponents in agreement with those from instrumental records. Dynamic CO_2 forcing alone is not enough, but adding dynamic solar and volcanic forcing has been claimed to be necessary to produce scaling properties in better agreement with those of observed records (*Vyushin et al.*, 2004, *Rybski et al.*, 2008). There are also indications that in some data records the scaling properties may differ in different regimes of time scales (*Pelletier*, 1997), suggesting that different physics govern the different regimes. The goal of the present thesis is to shed light on some of these issues.

1.2 Thesis Outline

This thesis is organized as follows: Chapter 2 provides a brief introduction to the concepts of short- and long-range memory. Trend estimation is a major issue in climate science, and involves analysis of time series with short-range and long-range memory. The approaches used here are described in Chapter 3. Chapter 4 explains methods for estimating parameters for different noise processes. A wide range of methods is used in the literature regarding LRM, and there are also several methods for analysing SRD series. In addition there are variations of some of the methods. The methods described in Chapter 4 are limited to the methods used in this thesis. A literature review is given in Chapter 5. The papers are summarized in Chapter 6, and some concluding remarks are given in Chapter 7.

1.3 List of Publications

Papers

Paper I

Rypdal, R., L. Østvand, and M. Rypdal, **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, 2013.

Paper II

Østvand, L., R. Rypdal, and M. Rypdal, **Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years**. Submitted to *Earth System*

Dynamics Discussions, 2014.

Paper III

Østvand, L., T. Nilsen, R. Rypdal, D. Divine, and M. Rypdal, **Long-Range Memory in Millennium-Long ESM and AOGCM Experiments**. Submitted to *Earth System Dynamics Discussions*, 2014.

Other Publications and Presentations

As first author

Østvand, L., T. Nilsen, K. Rypdal, and M. Rypdal, **Long range memory and trends in model data**. Poster presentation at *American Geosciences Union Fall Meeting*, San Francisco, December, 2013.

Østvand, L., M. Rypdal, and K. Rypdal, **The performance of wavelet-variance analysis as a method for estimating long-range memory in climatic temperature record**. Poster presentation at *European Geosciences Union General Assembly*, Vienna, April, 2013.

Østvand, L., O. Løvsletten, M. Rypdal, and K. Rypdal, **Maximum Likelihood Estimates of trend- and memory-coefficients in climatic time series**. Poster presentation at *European Geosciences Union General Assembly*, Vienna, April, 2012.

Østvand, L., O. Løvsletten, M. Rypdal, and K. Rypdal, **Maximum Likelihood Estimates of trend- and memory-coefficients in climatic time series**. Oral presentation at *Workshop on Complexity and Climates*, Tromsø, March, 2012.

Østvand, L., K. Rypdal, and M. Rypdal, **Universal Hurst exponent of local an global Earth temperature records?**. Poster presentation at *European Geosciences Union General Assembly*, Vienna, April, 2011.

As co-author

Rypdal, M., K. Rypdal, L. Østvand, and O. Løvsletten, **Stochastic modelling of global temperature**. Oral presentation at *Workshop on Complexity and Climates*, Tromsø, March, 2012.

Zivkovic, T, L. Østvand and K. Rypdal , **On the connection between the multifractality and the predictability from the auroral index time series**. Poster presentation at *24rd Summer School and International Symposium on the Physics of Ionized Gases*, Novi Sad, Serbia, August 2008. Published in *Publications of the Astronomical Observatory of Belgrade*, **84**, 511-514, 2008.

Chapter 2

Short- and Long-Range Memory

2.1 Short-Range Memory

Although the main theme in this thesis is long-range memory, a brief discussion of short-range memory cannot be avoided. SRM processes have been widely used to model climate series, and in the more recent years both SRD and LRD processes have been used for statistical modelling of climate time series (*Percival et al.*, 2001, *Zorita et al.*, 2008, *Vyushin et al.*, 2012). SRD processes are characterized by an autocorrelation function (ACF), $\rho(t)$, for which the integral $\int_0^\infty \rho(t)\mathrm{d}t$ is finite. One of the simplest and most commonly used SRD processes is the first order auto-regressive process AR(1), given by (e.g., *Box and Jenkins*, 1970)

$$x(t) = \phi x(t-1) + \sigma w(t),\ t = 1,2,\ldots, \tag{2.1}$$

where $w(t)$ is a discrete Gaussian white noise process of unit variance. AR(1) is part of the wider AR(p) family,

$$x(t) = \sum_{l=1}^{p} \phi_l x(t-l) + \sigma w(t). \tag{2.2}$$

AR(p) has an autocorrelation function

$$\rho(k) = \sum_{l=1}^{p} \phi_l \rho(l-k), \tag{2.3}$$

and power spectral density (PSD)

$$S(f) = \frac{\sigma_x^2}{|1 - \sum_{l=1}^{p} \phi_l \exp(-2\pi i l f)|^2}, \tag{2.4}$$

where f is the frequency. The PSD is defined in the interval $-1/2 < f < 1/2$. For AR(1), the ACF and PSD becomes

$$\rho(k) = \phi^{|k|}, \tag{2.5}$$

$$S(f) = \frac{\sigma_x^2}{1+\phi^2-2\phi\cos(2\pi f)}. \tag{2.6}$$

The ACF can be written as an exponentially decaying function,

$$\begin{aligned}\rho(k) &= \phi^{|k|} \\ &= \exp(\log\phi^{|k|}) = \exp(-|k|/\tau_c).\end{aligned}$$

The decorrelation time τ_c is then determined by ϕ through

$$\tau_c = -\frac{1}{\log\phi}. \tag{2.7}$$

If $\phi \approx 1$, $\tau_c \approx \frac{1}{1-\phi}$. If $2\pi f \ll 1$ and $\phi \approx 1$, then

$$\begin{aligned}S(f) &\sim \frac{1}{1+\phi^2-2\phi(1-(2\pi f)^2/2)} \\ &\sim \frac{1}{(1-\phi)^2+\phi(2\pi f)^2} \\ &\sim \frac{1}{\tau_c^{-2}+(2\pi f)^2}.\end{aligned}$$

The spectrum has the form of a Lorenzian, showing that the process behaves as Brownian motion with $\beta = 2$ on scales $t < \tau_c$ and as white noise on scales $t > \tau_c$.

AR(1) is the discrete equivalent of an Ornstein-Uhlenbeck process, which is the solution of the Langevin stochastic equation:

$$\mathrm{d}x(t) + \frac{1}{\tau_c}x(t)\mathrm{d}t = \sigma\mathrm{d}B(t), \tag{2.8}$$

where $B(t)$ is the Wiener process (Brownian motion).

2.2 Long-Range Memory

For stationary processes, long-range memory (LRM) is characterized by a slowly decaying autocorrelation function, $\rho(t) \sim t^{-\gamma}$ as $t \to \infty$, with $0 \leq \gamma < 1$. In this range of γ we have that

$$\int_0^\infty \rho(t)\mathrm{d}t = \infty. \tag{2.9}$$

Equivalently, the power spectral density (PSD) of a long-range memory time series follows a power law,

$$S(f) \sim f^{-\beta}. \tag{2.10}$$

As different methods for studying the phenomenon have been introduced, so are different scaling parameters. The Hurst exponent H, after *Hurst* (1951), is widely used, and so is the power spectral density parameter β. The relation between the parameters is

$$H = 1 - \frac{\gamma}{2} = \frac{\beta+1}{2}. \tag{2.11}$$

The LRM ranges corresponding to stationary processes for H and β are $1/2 < H \leq 1$ and $0 < \beta \leq 1$. Nonstationary self-similar processes (with stationary increments) are characterized by $1 < \beta \leq 3$. LRM time series are often called persistent processes. Stationary processes for which $-1 < \beta < 0$ are anti-persistent, while $\beta = 0$ represents a completely random process (white noise).

Two generic processes with LRM properties are fractional Gaussian noise (fGn) and fractional auto-regressive integrated moving average (FARIMA). An fGn can be cumulatively summed to yield a fractional Brownian motion (fBm), commonly denoted $B_H(t)$, where H is the Hurst exponent for the increments. An fBm, $B_H(t)$, exhibits the properties (i)-(iv):

(i) $B_H(t)$ is Gaussian.

(ii) $B_H(0) = 0$ almost surely.

(iii) $E[B_H(t) - B_H(s)] = 0$.

(iv) $\mathrm{var}[B_H(t) - B_H(s)] = \sigma^2|t-s|^{2H}$.

An fBm is self-similar, i.e., $B_H(at) \overset{d}{=} a^H B_H(t)$, where $\overset{d}{=}$ means equal in distribution. Using this property, it can be shown that an fGn has the following autocorrelation function,

$$\rho(k) = \frac{1}{2}\sigma[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]. \tag{2.12}$$

The asymptotic behaviour follows by Taylor expansion, where $\rho(k)$ is first rewritten:

$$\rho(k) = \frac{1}{2}\sigma k^{2H} g(k^{-1}), \quad g(x) = (1+x)^{2H} - 2 + (1-x)^{2H}$$

$$\begin{aligned} g(x) &= g(0) + g'(0)x + g''(0)x^2 + O(x^3) \\ &= 4H(2H-1)x^2 + O(x^3). \end{aligned}$$

Then

$$\lim_{k\to\infty} \rho(k) = 2H(2H-1)\sigma k^{2H-2} \sim k^{-\gamma}, \tag{2.13}$$

explaining the relation $H = 1 - \gamma/2$ in eq. (2.11).

An ARMA(p,q) model is the stationary solution of

$$\phi(B)x(t) = \psi(B)w(t), \tag{2.14}$$

where B is the backshift operator $Bx(t) = x(t-1)$ and $w(t)$ is Gaussian white noise. ϕ and ψ are given by

$$\begin{aligned} \phi(x) &= 1 - \sum_{j=1}^{p} \phi_j x^j \quad \text{and} \\ \psi(x) &= 1 + \sum_{j=1}^{q} \psi_j x^j. \end{aligned}$$

A FARIMA(p,d,q) process is an ARMA(p,q) process which holds for the dth difference $(1-B)^d X_t$,

$$\phi(B)(1-B)^d X_t = \psi(B)w(t). \tag{2.15}$$

$(1-B)^d$ can be defined for any real number d by

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k. \tag{2.16}$$

The FARIMA$(0,d,0)$ process, where $d = H - 1/2$, is often preferred due to its simple autocorrelation function,

$$\rho(k) = \sigma \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)}, \tag{2.17}$$

where Γ is the gamma function. Similar to an fGn, this process is stationary for $-1/2 < d < 1/2$ (*Beran*, 1994).

Chapter 3

Trends

3.1 Trends and noise

In climate studies, it is common to separate the time series into a trend component and a random component, often called the "climate noise". The notion of climate noise may be slightly misleading in climate studies, since the noise contains interesting information about the climate system. Another source of confusion is that the trend definition is ambiguous. In some studies, the trend refers to a linear increase or decrease, while in others it is the slow variation of the observed record. This slow variation is often characterized through a function which may be, e.g., a polynomial, an oscillation or a combination of both.

The simplest and most commonly used noise model is white noise, which is completely uncorrelated. Correlated noise may be a short-range memory (SRM) process or long-range memory (LRM) process. Common examples are AR(1) (SRM, Section 2.1) and fGn (LRM, Section 2.2). When a model for the trend has been selected, the trend could be subtracted from the record, and the correlation structure of the residual studied. Different methods may be applied to estimate parameters for the noise model. For an LRM model these parameters could be the scaling parameter β and standard deviation σ. Some methods have detrending incorporated in the estimator for the scaling parameter, e.g., DFA which removes polynomial trends. If a polynomial trend model is chosen, prior detrending is unnecessary using this estimator. However, the residual may not be completely described by a simple noise model. Hence, to to call it a noise may therefore be misleading, and the broad term residual is preferred.

Detrending is important, because trends may influence both the estimation of noise parameters and scaling properties of the time series. However, too much detrending or choosing a poor trend model may also generate spurious correlations in the residual. The significance of trends is therefore assessed more convincingly by exploring different noise models and search for a combination of trend and noise where the trend represents a good fit to the slow variations of the record and the noise model represents a good description of the statistical properties of the residual. Generally, noise models with no or only short-range memory do not exhibit strong slow

variations, while persistent long-range memory processes do.

3.2 Trend estimation

A trend model is based on the hypothesis that observed record can be modelled as a realization of a stochastic process of the form

$$x(t) = T(A;t) + \sigma w(t), \tag{3.1}$$

where $T(A;t)$ is the trend model with parameters $A = (A_1, \ldots, A_m)$ and $\sigma w(t)$ is some noise process. The model in eq. (3.1), with parameters A estimated from the observed record, is the hypothesis whose significance we will test, and is denoted *the alternative hypothesis*. If our trend model is properly selected, it captures a great fraction of the variance of the record, and hence the noise part has a small variance. In that case a maximum likelihood estimation of the trend parameters A under different noise models (white or "coloured") will yield results similar to what is found by a least-square fit. For the estimation of trend parameters it does not matter much what the correct model for the climate noise is. We shall see, however, that it matters a lot when we formulate the *null hypothesis* against which the trend model is tested.

Some typical trend models are:

$$\begin{aligned}
&T(a_0, a_1; t) = a_0 + a_1 t && \text{linear trend} \\
&T(a;t) = \sum_{k=0}^{m} a_k t^k && \text{polynomial trend} \\
&T(A, \omega, \varphi; t) = A \sin(\omega t + \varphi) && \text{oscillatory trend} \\
&T(a, A, \omega, \varphi; t) = \sum_{k=0}^{m} a_k t^k + A \sin(\omega t + \varphi) && \text{combination of polynomial and oscillatory trend}
\end{aligned} \tag{3.2}$$

An aid in formulating the alternative hypothesis could be to do a low-pass filtering to capture the slow variations, and denote this the "trend". This can be done through, e.g., Fourier analysis, wavelet analysis or principal component analysis. Everything in the record not captured by the trend is called noise or residual from the trend, and is typically considered to be the fast variation in the record. One should keep in mind that the trend model is of interest only if it reflects some hypothesized physical reality, e.g., a rising trend due to anthropogenic forcing and/or a distinct oscillation of natural origin that stands out of the background climate noise continuum.

In *the null model*, all the slow variability of the observed record is assumed to be captured by a particular noise model $\varepsilon(\theta;t)$. Here $\theta = (\theta_1, \ldots, \theta_n)$ are the parameters which characterize this noise process. Monte Carlo studies can be made to assess the probability that the estimated trend may be explained as a natural fluctuation produced by the null model. The work flow is as follows:

(i) Select a trend model and a null model.

(ii) Estimate the parameters of both models from the observed record.

(iii) Construct a Monte Carlo ensemble of realizations of the null model noises with the estimated noise parameters.

(iv) Estimate "pseudotrend" parameters for each realization by fitting the trend model to each realization, and estimate a probability distribution for these parameters.

(v) Test statistical significance of the trend by observing whether the observed trend parameters are outside the 95 percentile of the pseudotrend distribution. If they are, the null hypothesis is rejected and the trend is significant. If they are not, the trend is deemed insignificant.

If the null hypothesis is rejected, a new null model is formed where the trend is included, i.e., $x(t) = T(A;t) + \varepsilon(\theta;t)$. The new null model can then be tested against new alternative models, until a model that describes the record in a satisfactory way is found, as shown in Figure 3.1.

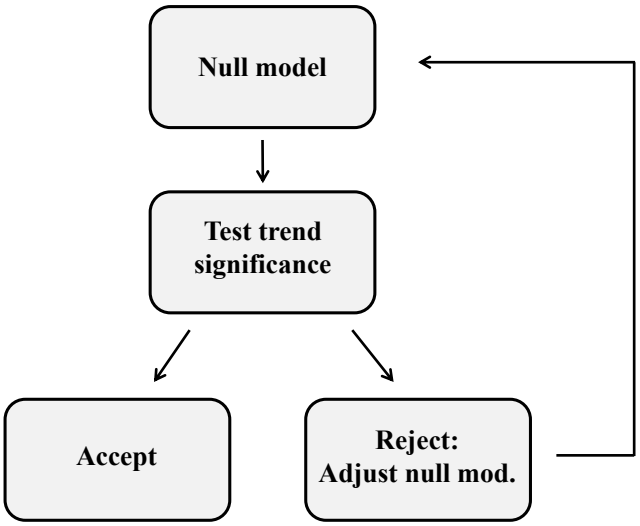


Figure 3.1: Flowchart of hypothesis testing

In Paper II we consider SRM as well as LRM null models. This does not imply that any null model is appropriate for a given data set. A proper null model should be consistent with the correlation structure of the observed data record. For instance, if the null model is an LRM noise, a correlation measure derived from the record should be consistent with an LRM model, and inconsistent with an SRM model. Methods for selecting the proper null model are discussed in Chapter 4.

Example

Figure 3.2 illustrates fitting a linear trend to monthly temperature at Cheyenne, Wyoming, USA (*Brohan et al.*, 2006, *Jones and Moberg*, 2003). The temperature is the anomaly from the temperature mean from 1961 to 1990. This record was chosen because it is one of the local continental time series with the longest record without any missing data. It covers the period 1871-2010 AD, and thus consists of 1680 data points. Cumulative distribution functions (CDFs) for the slope of

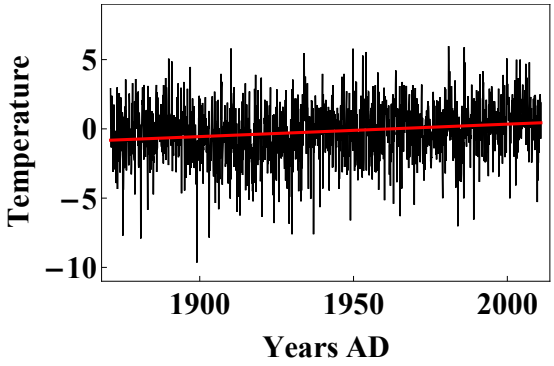


Figure 3.2: Temperature anomaly at Cheyenne. The red line is the linear trend fitted to the record.

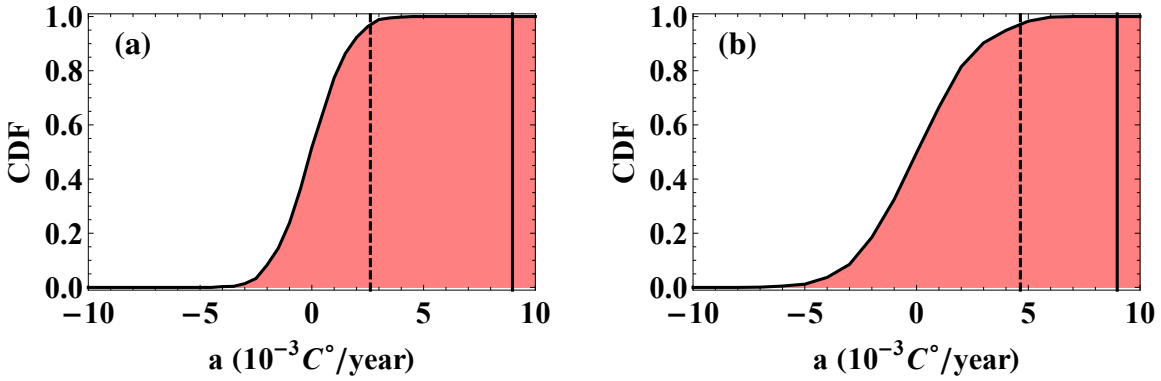


Figure 3.3: Significance of trends under (a) AR(1) null hypothesis and (b) fGn null hypothesis. Ensembles of synthetic noises are produced with the same parameters as those estimated under each hypothesis by maximum likelihood estimation (MLE). The pseudotrend parameters are then estimated by least-square fitting to each realization, and the CDF for the slope parameter is found. The dashed line marks the 95 percentile for the CDF and the solid line marks the trend slope found for the observed temperature.

the linear trend are found by generating ensembles of realizations of AR(1) and fGn processes with parameters found by maximum likelihood estimation (MLE) from the temperature record. This means that the parameters (ϕ, σ) are estimated from the AR(1) model and (β, σ) from the fGn model. The pseudotrend parameters are then estimated by least-square fitting to each realization, and the CDF for the slope parameter is found, as shown in Figure 3.3. The CDF for the

synthetic AR(1) is narrower than that for the fGn, which is expected since long-range correlations may produce more slow variations in the synthetic realizations. The dashed line in each figure indicates the 95% confidence and the solid line is the slope estimated from the temperature record. The trend is found to be significant for AR(1) as well as fGn noise models. Which noise process that best describes the noise is discussed in Chapter 4.

3.3 Response Model

A different approach is to consider a linear response model of the surface temperature which incorporates deterministic and stochastic forcing. The deterministic forcing describes known, external climate forcing components and the stochastic forcing represents the internal dynamics on unresolved spatiotemporal scales. The response to the deterministic forcing is the counterpart of the trend in the trend models, but differs from this in the sense that the deterministic forcing can contain both fast and slow variations. In the trend models the fast, forced variations are relegated to the residual noise, whereas the response model is capable of separating this fast response from the internal, stochastic variability driven by the stochastic forcing. The separation of externally driven from internal, natural variability is one of the central problems in climate science. The starting point for this approach is the linearized energy-balance equation (e.g., *Hansen et al.*, 2011, *Rypdal*, 2012),

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} + \frac{1}{S_{\mathrm{eq}}} T(t) = F(t). \tag{3.3}$$

Q is the total energy content of the climate system, and F and T are perturbations of radiative influx and surface temperature relative to a reference state in radiative equilibrium. S_{eq} is the climate sensitivity. Using the effective heat capacity C ($\mathrm{d}Q = C\mathrm{d}T$) and the time constant $\tau_c = CS_{\mathrm{eq}}$, this can be rewritten to

$$\mathcal{L}T(t) \equiv C\left(\frac{\mathrm{d}}{\mathrm{d}t} + \frac{1}{\tau_c}\right) T(t) = F(t), \tag{3.4}$$

where the linear operator $\mathcal{L}$ has the Green's function $G(t) = C^{-1}\exp(-t/\tau_c)$. The solution of eq. (3.4) is the deterministic response to the forcing,

$$T(t) = \int_{-\infty}^{t} G(t-s)F(s)\mathrm{d}s. \tag{3.5}$$

An equilibrium reference state is defined such that T is the temperature relative to the initial temperature $\hat{T}_0$, i.e., $T = \hat{T} - \hat{T}_0$. The forcing $F(0)$ at $t = 0$ is not necessarily 0, and also usually not known a priori. The forcing data is given as $F(t) = F(0) + F_G(t)$, where $F_G(t)$ is the total "given" forcing and $F(0)$ is one of the parameters to be estimated. A perfect match to the observed record cannot be obtained because the forcing should also have a stochastic component corresponding to the random forcing of the ocean-land heat content from the atmospheric weather system. This

can be introduced by rewriting eq. (3.5) to

$$T(t) = \underbrace{\int_{-\infty}^{t} G(t-s)F(s)\mathrm{d}s}_{\text{deterministic response}} + \underbrace{\sigma \int_{-\infty}^{t} G(t-s)\mathrm{d}B(s)}_{\text{stochastic response}}, \tag{3.6}$$

where $B(t)$ is the Wiener process. The stochastic response yields an Ornstein-Uhlenbeck process, corresponding in the discrete case to an AR(1). The linear operator can be replaced with a fractional derivative operator to obtain a scale-free response model with a Green's function with a power-law, $G(t) = (t/\mu)^{\beta/2-1}\xi$, where μ is a scaling factor in the units of time characterizing the strength of the response and $\xi \equiv 1$ Km^2/J is a factor needed to give $G(t)$ the right physical dimension. The stochastic response will then be an fGn when $-1 < \beta \leq 1$ and an fBm when $1 < \beta \leq 3$. The σ in eq. (3.6) is the standard deviation of the noise process. Different Green's functions can be used, and the correlation structure of the residual from the deterministic response can be analysed by methods (e.g. DFA) that will distinguish LRM processes from SRM processes (*Rypdal and Rypdal*, 2013). The parameters $(F(0), C, \sigma, \tau_c)$ for the exponential model, and $(F(0), \mu, \sigma, \beta)$ for the scale-free model are estimated using MLE, as will be described in Chapter 4.

Example

The forcing data available are global, so in the following example, global land temperature (*Jones et al.*, 2012) is used. Figure 3.4 shows the global land temperature record in black, and deterministic response in cyan and red using the exponential and scale-free response model respectively. The two models yield quite similar deterministic responses, and it is not clear which one is the better fit. The difference between the temperature and the deterministic response will therefore be analysed further with the methods given in Chapter 4.

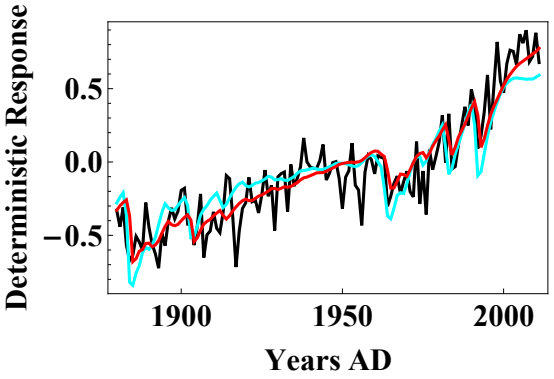


Figure 3.4: Yearly averaged global land temperature (black), and deterministic response using the exponential response model (cyan) and scale-free response model (red).

Chapter 4

Methods

4.1 Short-Range Memory

There are several methods for estimating the parameters ϕ_k in an AR(p) process. In our studies we have limited the use of SRM processes to AR(1), and have used the following methods to estimate ϕ.

Autocorrelation Function

For an AR(1) the autocorrelation function (ACF) is given by $\rho(t) = \phi^{|t|}$. This means that the lag-one correlation $\rho(t = 1) = \phi$. Although estimators of ACF are noisy and inaccurate, most of them are unbiased and with low uncertainty for the smallest lag. For the purpose of comparing AR(1) to fGn as models for a times series it works well enough for estimating ϕ.

Maximum Likelihood Estimation

In maximum likelihood estimation (MLE) the log-likelihood function of a process is optimized. For AR(1) we can rewrite eq. (2.1);

$$\delta(t) = x(t) - \phi x(t-1) = \sigma w(t), \tag{4.1}$$

and use $\delta(t)$ as input time series in the log-likelihood function of a white noise with standard deviation σ, and then optimize with respect to (ϕ, σ).

A slightly different approach is to use the log-likelihood function of an Ornstein-Uhlenbeck process,

$$\log L(\lambda, \sigma) = -\frac{N+1}{2}\log(\pi\sigma^2) - \frac{x(0)^2}{\sigma^2} - \frac{1}{2}\sum_{k=1}^{N}\log(1-\exp(-2\lambda\Delta)) \tag{4.2}$$

$$-\sum_{k=1}^{N}\frac{x(k)-\exp(-\lambda\Delta)x(k-1))^2}{\sigma^2(1-\exp(-2\lambda\Delta)}, \tag{4.3}$$

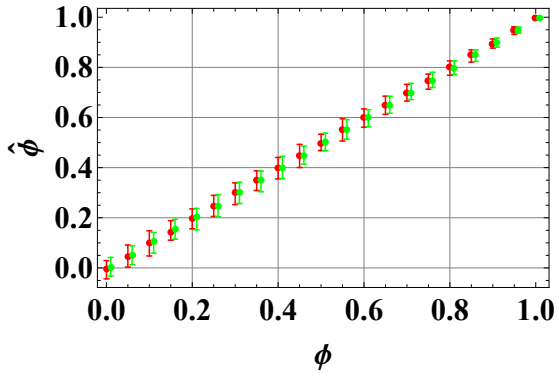


Figure 4.1: Bias and error bars for ACF (red) and MLE (green).

where Δ is the sampling rate, N is the length of the time series and $\lambda = 1/\tau_c = -\log\phi$.

Figure 4.1 shows bias and error bars when applying the ACF method and MLE to estimate ϕ in an AR(1) process. The two methods seem to perform very similar. They are unbiased, with decreasing error bars as ϕ increases.

4.2 Long-Range Memory

The toolkit of methods to estimate scaling parameters for long-range memory includes rescaled range R/S analysis, variations of the variogram, variations of spectral analysis, wavelet variance analysis (WVA), detrended fluctuation analysis (DFA) and maximum likelihood estimation (MLE). MLE is in this case used to optimize the log-likelihood function of an LRM process, e.g., an fGn with respect to the scaling parameter H. Only the methods used in this thesis are described here. There are several studies comparing different estimators of LRM parameters (e.g., *Heneghan and McDarby*, 2000, *Weron*, 2002, *Delignieres et al.*, 2006, *Mielniczuk and Wojdyłło*, 2007, *Franzke et al.*, 2012). All methods have their advantages and disadvantages, and their performance depends on the data at hand and the purpose of the analysis. Trends may be present, generally leading to overestimation of the scaling parameter. Methods like DFA and WVA has polynomial detrending incorporated, but are in some cases biased. Correct estimation also presumes choosing correct detrending order. If the order is to low, β is in general overestimated, but choosing a high order leads to a smaller scaling regime. MLE is unbiased for most β and has the smallest error bars, but depends on choosing the correct noise model and a good representation of trends if they are present. MLE also tends to emphasize the short scales, which may be problematic, e.g., if several scaling regimes are present or there are trends affecting the short scales to a large degree. The periodogram is noisy and does not include any detrending, but may give good indications of whether scaling is at all present, whether there are several scaling regions and whether there are oscillations in the investigated record. Application of several methods to the same data sets therefore gives a clearer picture of scaling regimes, possible trends and the value and uncertainty of the estimated scaling parameter.

Semivariogram

The semivariogram (*Matheron*, 1963) is given by

$$\gamma[k] = \frac{1}{2(N-k)} \sum_{n=1}^{N-k} (y[n+k] - y[n])^2, \tag{4.4}$$

where k is the time lag between two values of the cumulatively summed time series $y[n] = \sum_{i=1}^{n} x[i]$. The semivariogram scales with k (*Mandelbrot and Van Ness*, 1968),

$$\gamma[k] \sim k^{2H} \tag{4.5}$$

Periodogram

The periodogram is a simple estimator for the power spectral density (PSD),

$$S(f) = \frac{2}{N\Delta t} |X(f)|^2, \tag{4.6}$$

where $X(f)$ is the Fourier transform of the time series to be analysed, $x(t)$. Since the PSD is symmetric, the frequencies of interest is $f = m/N$, $m = 1, 2, \ldots N/2$. For these frequencies, a self-affine time series scales as a power-law (e.g., *Voss*, 1986)

$$S(f) \sim f^{-\beta}. \tag{4.7}$$

To put equal emphasis on all scales, log-binning is often used before fitting a straight line to the PSD in a log-log plot. The periodogram is known to have variance problems, but this is reduced by the log-binning, and the scaling behaviour is still easily seen. The periodogram can be used as a first analysis to look for power-law scaling. For an accurate estimate of β, other methods are recommended.

Wavelet Variance Analysis

The wavelet transform was introduced by *Grossmann and Morlet* (1984). The continuous version is given by

$$W(t, \tau; x(t), \Psi(t)) = \int_{-\infty}^{\infty} x(t') \frac{1}{\sqrt{\tau}} \Psi\left(\frac{t'-t}{\tau}\right) \mathrm{d}t', \tag{4.8}$$

i.e., the convolution between a time series $x(t)$ and the wavelet $\Psi(t)$. The mother wavelet $\Psi(t)$ and all rescaled versions of it must fulfil the criteria

$$\int_{-\infty}^{\infty} \Psi(t') \mathrm{d}t' = 0. \tag{4.9}$$

For self-similar time series, the variance $F(\tau) = \sum_{t=1}^{N} W(t, \tau)$ scales as a power-law (*Flandrin*, 1992, *Malamud and Turcotte*, 1999).

$$F(\tau) \sim \tau^{\beta}. \tag{4.10}$$

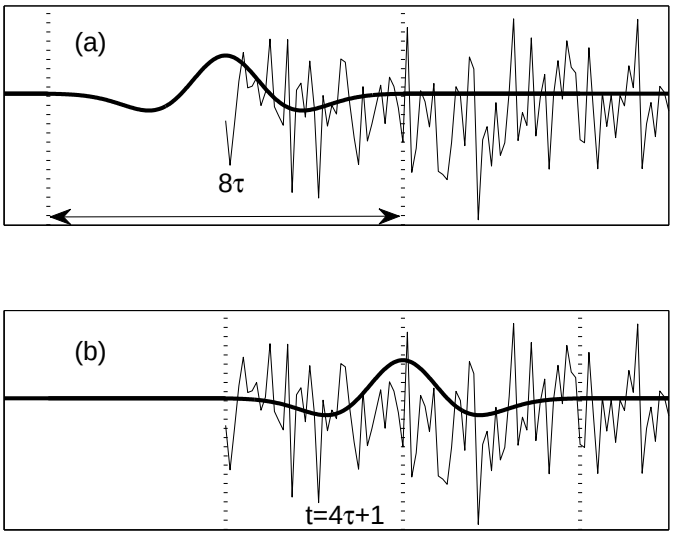


Figure 4.2: The time series is covered by the wavelet (a) for only the last half of the time when the wavelet is centred in $t = 1$ and (b) at all times when it is centred in $t = 4\tau + 1$.

The method is therefore known as the Wavelet Variance Analysis (WVA). Eq. (4.10) is also valid for the discrete wavelet transform. Any wavelet can be used, but not all wavelets exist for both the continuous and discrete transform. If trends are present in the time series to be analysed, it is preferable with a wavelet with a high number of vanishing moments. For polynomial trends of order p, the wavelet must have $p+1$ vanishing moments to detrend the data properly and give the correct value for β (*Abry and Veitch*, 1998). In our studies we have used the nth derivative of Gaussian wavelet,

$$\frac{(-1)^{n+1}}{\sqrt{\Gamma(n+1/2)}}\frac{\partial^n}{\partial t^n}\exp\left(-\frac{t^2}{2}\right), \tag{4.11}$$

and use the notation WVAn to indicate which order of the wavelet that has been used. The number of vanishing moments corresponds to the order of the derivative.

When performing the wavelet transform, the times near the beginning and the end of the time series will not be covered by the wavelet. This will influence the wavelet coefficients at these times, and these are therefore deleted before computing the variance. The derivative of Gaussian wavelet has a width 8τ, and the time series will not be completely covered by the wavelet until $t = 4\tau + 1$, see figure 4.2.

Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis (DFA) (*Peng et al.*, 1994, *Kantelhardt et al.*, 2001) was explicitly designed to remove polynomial trends. The method can be summarized in four steps. First, the cumulative sum (the profile) is computed,

$$Y(i) = \sum_{t=1}^{i} x(t) - \langle x \rangle, \tag{4.12}$$

where $\langle x \rangle$ denotes the mean. In the second step the profile is divided into $N_\tau = N/\tau$ non-overlapping segments of equal length τ. This is done starting both at the beginning and at the end of the profile, so $2N_\tau$ segments are obtained altogether. In the third step, an nth order polynomial is computed and subtracted for each segment,

$$Y_\tau(i) = Y(i) - p_\nu(i), \tag{4.13}$$

where $p_\nu(i)$ is the polynomial fitted to the νth segment. The notation DFAn is used to indicate the order of the polynomial. In the final step, the variance of each segment is computed,

$$F^2(\nu,\tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} Y_\tau^2[(\nu-1)\tau + i]. \tag{4.14}$$

The fluctuation function is given by the square root of the average over all the segments,

$$F(\tau) = \left[\frac{1}{2N_\tau} \sum_{\nu=1}^{2N_\tau} F^2(\nu,\tau) \right]^{\frac{1}{2}}. \tag{4.15}$$

The scaling exponent is defined by the relation

$$F(\tau) \propto \tau^\alpha. \tag{4.16}$$

α corresponds to the Hurst exponent H when $0 < H < 1$. DFA may also yield $\alpha > 1$, and is related to β through $\beta = 2\alpha - 1$. For a time series with no trends, the detrending in the third step is unnecessary, and the standard fluctuation analysis, FA, can be used. Then a simplified definition of the variance for each segment, $F_{\mathrm{FA}}^2(\nu,\tau) = [Y(\nu\tau) - Y((\nu-1)\tau)]^2$, replaces $F^2(\nu,\tau)$ in eq. (4.15) (*Bunde et al.*, 2001).

By definition the scales must be $\tau > n+2$ (*Kantelhardt et al.*, 2001). The effect of trends was studied in *Hu et al.* (2001), where an upper limit of $\tau < N/10$ for the scaling region was suggested, where N is the record length. The practical implication is that scaling properties can only be accurately assessed up to time scales one tenth of the time record analysed.

Maximum Likelihood Estimation

In Maximum Likelihood Estimation of the Hurst exponent H, a log-likelihood function using the autocorrelations for the LRM process is optimized with respect to H. This function is given by

$$\log L(\mu,\sigma,H) = -\frac{1}{2}\log|C_N(H)| - (2\sigma)^{-1}S(\mu,H) - (N/2)\log\sigma, \tag{4.17}$$

where $S(\mu,H) = (\mathbf{x}-\mu\mathbf{1})^\top[C_N(H)]^{-1}(\mathbf{x}-\mu\mathbf{1})$. $\mathbf{x}$ is the time series to be analysed expressed as a column vector, and $C_N(H)$ is the correlation matrix (*McLeod and Hipel*, 1978). The autocorrelations are given by eq. (2.12) for fGn and eq. (2.17) for FARIMA$(0,d,0)$. The inversion of $C_N(H)$ has a high computational cost, so the implementation of the MLE method was done by using the R package FGN (*McLeod et al.*, 2007), where the Durbin Levinson algorithm is used to compute the log-likelihood function.

The advantage of WVA and DFA is that the methods can be applied to both stationary and non-stationary time series, with scaling exponent in the range $0 < \beta < 3$ (although for some β there is a bias). One problem with both WVA and DFA is to find the best scaling regime to determine the scaling parameter. LRM processes do not always scale well on the smallest scales, as LRM defined by the asymptotic behaviour, but it is not straightforward to find when this behaviour starts (*Beran*, 1994). Furthermore, the time series may include different scaling properties for different scales, e.g., in the presence of trends or if the dynamics change with time. In DFA, this is often seen as a clear cross-over at a certain scale. However, if the time series is short, the range of scales available is small, and it may be hard to find the proper scaling regime.

For MLE, the problem of finding the proper scaling regime is shifted to selecting the most proper model, both with regards to noise process and trends. Trends could be included in the model such that trend coefficients for, e.g., a polynomial function is estimated together with H. Then $\mathbf{z} = \mathbf{x} - \mathbf{T}$, where $\mathbf{T}$ is the trend, is used instead of $\mathbf{x}$ in eq. (4.17). Observational data often contain trends, but the nature of these is usually unknown. A model with few parameters may not reproduce the observational data well. Using many parameters may fit the data better, but introduces higher uncertainty with a higher number of parameters to be estimated. There is also a chance of overfitting, and thus attributing properties of the noise to the trend. Another alternative is to apply a response model, using $\mathbf{z} = \mathbf{x} - G\mathbf{F}$ instead of $\mathbf{x}$ in eq. (4.17), where G is the Green's function and $\mathbf{F}$ is the forcing described in Chapter 3.

Comparing LRM Methods

To compare the performance of the methods, we can produce ensembles of synthetic fGn and fBm, and estimate β. The mean and 95% quantiles for each ensemble is computed, and shown as dots and error bars respectively in Figure 4.3. If the mean of the estimated β corresponds to the given β, there is no bias. Figure 4.3(a) shows bias and error bars for fGn/fBm analysed with DFA2, WVA2 and MLE for an fGn, which are the methods we have used most frequently to estimate β in our studies. DFA2 and WVA2 eliminate linear trends in the data. The MLE

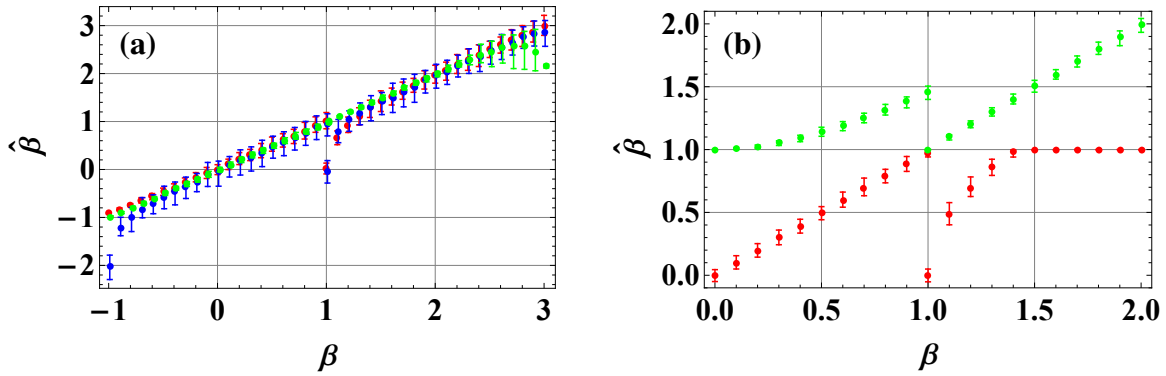


Figure 4.3: (a) Bias and error bars for DFA2 (red), WVA2 (blue) and MLE (green). (b) Bias and error bars for MLE. fGn and fBm with different β analysed as fGn (red) and fBm (green). analysing as fBm means that the increments are analysed as fGn and $\hat{\beta} = \hat{\beta}_{\text{incr}} + 2$.

estimates of β are unbiased given that the correct noise model (fGn/fBm) is chosen, except when the given β is close to 3. Both DFA2 and WVA2 have a negative bias when $\beta \to 1^+$, and WVA2 has a similar bias when $\beta \to -1^+$ as well. DFA2 is slightly overestimating β in the same range. Figure 4.3(b) illustrates the problem with choosing the wrong noise model when applying MLE. In the method used here, the autocorrelation function of fGn is used in the estimation of β, shown in red. This means that the estimation can only return $\hat{\beta}$ in the range $-1 < \hat{\beta} < 1$, even when the synthetic data set is a realization of an fBm with β in the range $1 < \beta < 3$. The result of such an estimation for synthetic records with $0 < \beta < 3$ is shown as the red dots in Figure 4.3(b). To provide correct estimated $\hat{\beta}$ for synthetic fBms in the range $1 < \beta < 3$, the record must be differenced, and the method applied to the increments. The estimated $\hat{\beta}$ is then $\hat{\beta} = \hat{\beta}_{\text{incr}} + 2$. The result of this approach is shown in green. When the correct noise model is used, the estimates are unbiased. When fBm are analysed as fGn, $\hat{\beta} \to 1$ as β increases, but smaller values are found when $\beta \to 1^+$. When the increments of fGns are studied, $\hat{\beta} \to 1$ as β decreases, but larger values are found when $\beta \to 1^-$.

Example

The results of applying the semivariogram, periodogram, WVA, and DFA to the Cheyenne temperature record are illustrated in Figure 4.4, and show that the methods cannot be applied uncritically. The semivariogram yields a higher value for β ($\beta = 0.50$) than the other methods. The periodogram follows a fairly straight line corresponding to $\beta = 0.22$, but there are some deviations at the lower frequencies. These are not included in the estimation of β, but would lead to a higher estimate if they were. WVA and DFA are applied for order 1-4. For order 2-4 they both yield $\beta \approx 0.2$, while WVA1 and DFA1 show higher estimates when all scales up to about $1/10$ of the record length are used. This is due to the influence of a trend, apparent as a change in scaling regimes. In Figure 3.3(b) in Chapter 4, a linear trend was found to be clearly significant when adopting an fGn null model. This typically leads to an overestimation of the scaling exponent if not taken into account, and this is exactly what we observe. WVA1 and DFA1 do not detrend, but

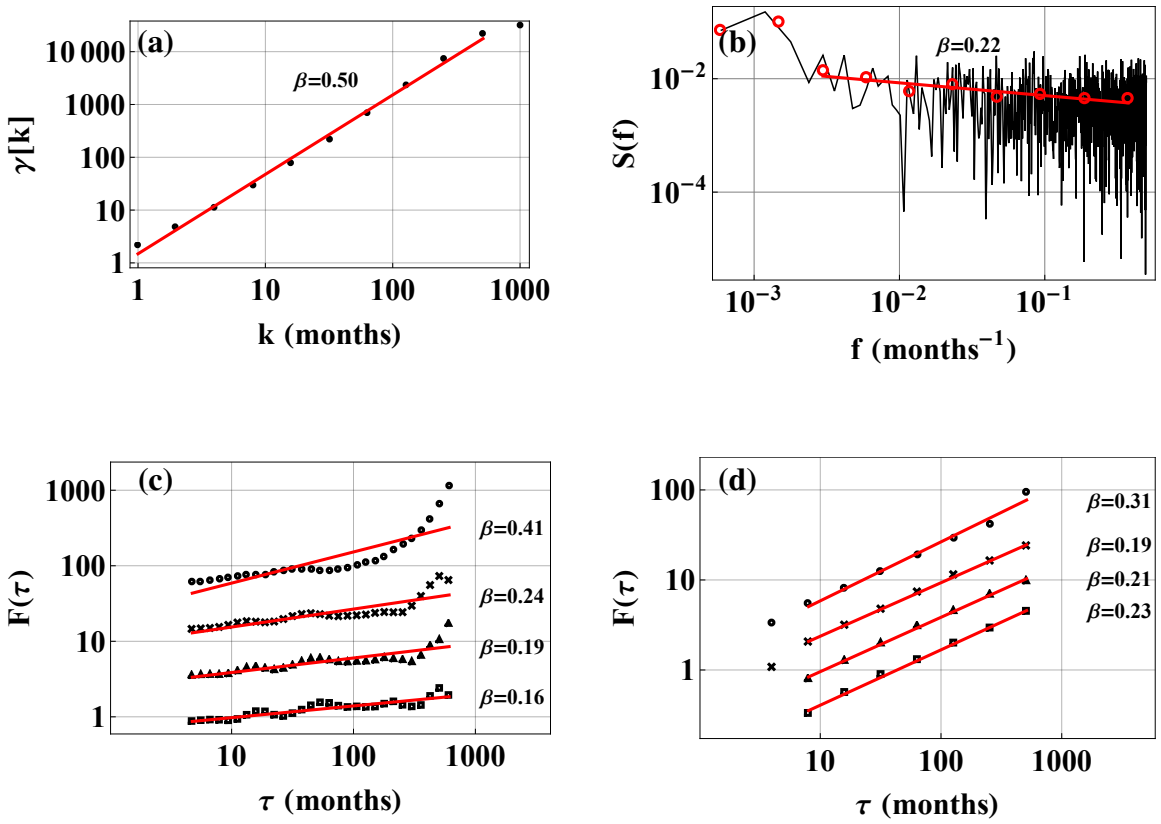


Figure 4.4: Temperature anomaly at Cheyenne analysed with (a) semivariogram, (b) periodogram, (c) WVA of order 1-4 from top to bottom and (d) DFA of order 1-4 from top to bottom. The red lines indicate the scaling range used to estimate β. The red circles in panel (b) are the result of log-binning.

order 2 and above eliminate linear trends. *Hu et al.* (2001) described how the fluctuation function of the noise is dominant on small scales and the fluctuation function of the trend is dominant on large scales for DFA. The change in regimes is not very obvious for DFA1 for the temperature at Cheyenne as the cross-over scale is fairly large. The effect is very visible for WVA1. If only the smaller scales are used in the estimate of β, overestimation is not a problem. In Figure 4.5 the scales used to estimate β are chosen by eye. The estimated β corresponds better to the ones found for the higher orders of WVA and DFA. There is, however, a problem not knowing at exactly what scale the cross-over takes place, as the transition is not very sharp. Including too high scales will make the estimate influenced by the trend, but only including the scales well below the cross-over leads to a smaller scaling regime, potentially leading to poor statistics. Using order 2 of DFA and WVA, or performing linear detrending prior to applying the estimation method, is therefore a safer choice. The higher orders of WVA do not follow a completely straight line. This could be an effect of trends (e.g. higher order polynomials or oscillations), but some waves in the wavelet variogram occur even for pure noises. WVA tends to enhance oscillations which appear as statistical fluctuations in realizations of a persistent noise, especially on the large scales. As

seen in Figure 4.3(a), WVA and DFA mostly have error bars of the same order, although DFA does not have the problem with wavy structure. For pure noises, the waviness in WVA mostly affects the error bars of the wavelet variance at each scale, and not so much the error bars of the estimated β.

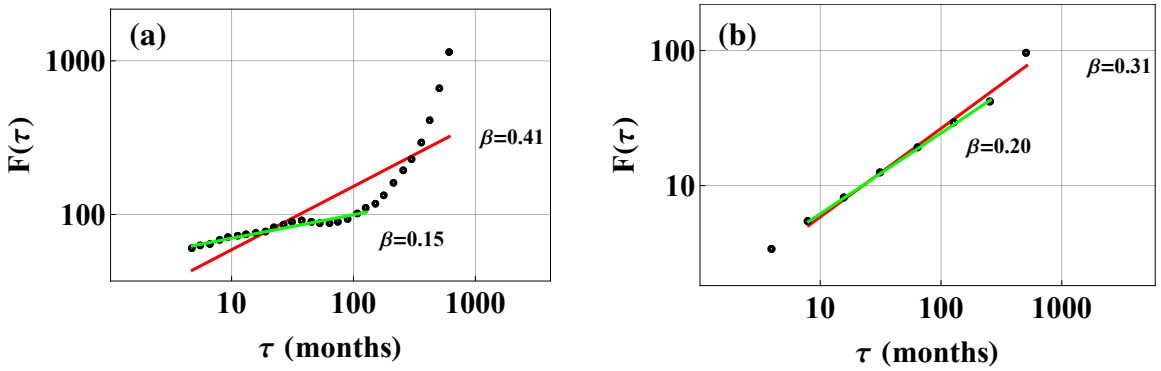


Figure 4.5: Temperature anomaly at Cheyenne analysed with (a) WVA1 and (b) DFA1. The red lines indicate the scaling region when all scales up to about $1/10$ of the record length are used to estimate β. The green lines show the scaling region when only the scales up to the assumed cross-over scale where the trend becomes dominant are used.

4.3 Comparing Short-Range and Long-Range Memory Processes

Selecting the correct null noise model is not straightforward, but one can select a few models and compare their scaling properties with those of the observational data. We have chosen to compare AR(1) and fGn/fBm, as these processes are widely used in the climate community. The theoretical spectra are well known, given in eq. (2.6) for AR(1) and eq. (2.10) for power-law scaling. WVA and DFA can also be used to distinguish between the two processes. In some cases it is difficult to draw firm conclusions on what process best fits the data. One example is local temperature records from continental interiors. These records show low persistence on time scales from months to decades; hence if they are sampled with monthly or longer sampling interval, they appear as white or very weakly persistent fractional noises. With higher sampling

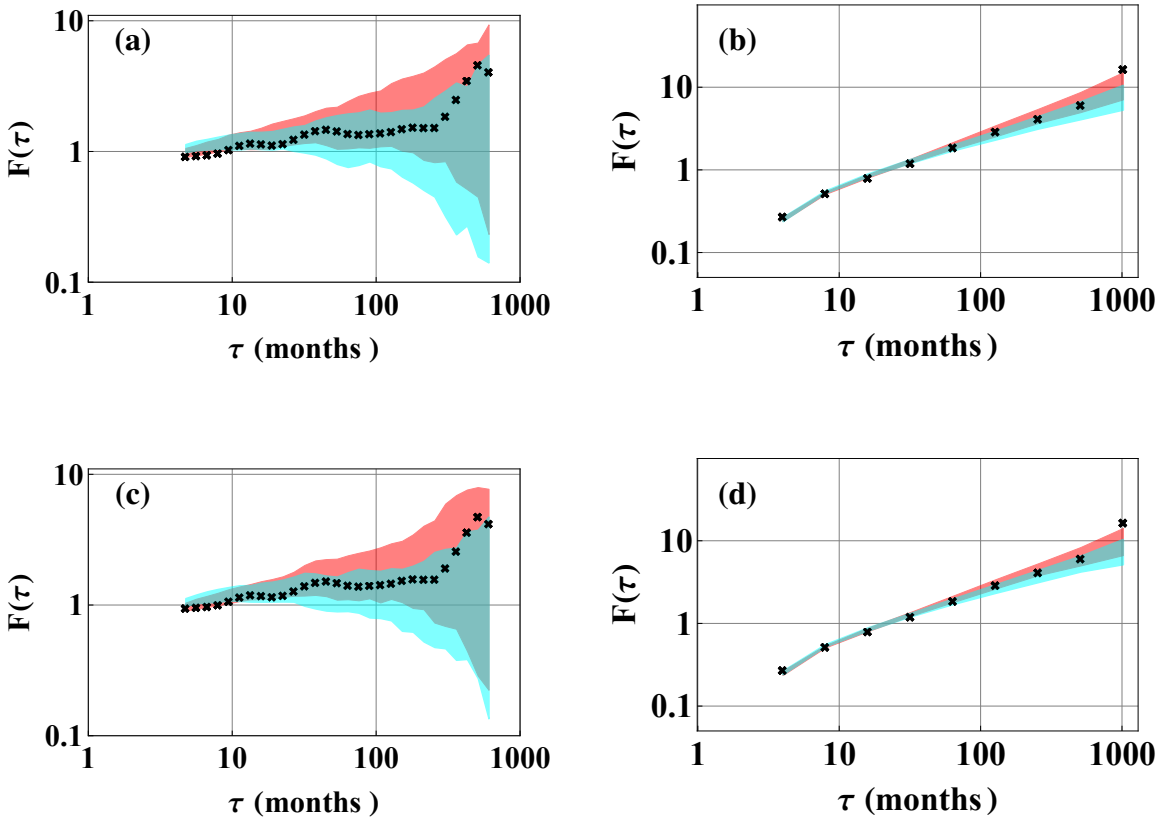


Figure 4.6: (a) WVA2 and (b) DFA2 applied to the temperature anomaly at Cheyenne (black crosses). Ensembles are generated of synthetic realizations of two different stochastic processes: An AR(1) process (cyan) and fGns (red). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the coloured areas are the 95% confidence regions for these estimates. Panel (c) and (d) show WVA2 and DFA2 applied to the linearly detrended temperature record and for the synthetic realizations of the processes generated with parameters estimated from the detrended record.

rates there will be correlations on scales shorter than a month that may be reminiscent of that of a Brownian motion. Hence the total correlation structure may be similar to that of an Ornstein-Uhlenbeck (OU) process with τ_c of the order of a month or less. The temperature from Cheyenne serves a good example. Figure 4.6 shows WVA2 and DFA2 applied to the temperature record before and after linear detrending. The cyan area is the 95% confidence area for an AR(1) process and the red area is the 95% confidence area for fGns. The noise processes have parameters estimated from the record with MLE. The results are fairly similar before and after the detrending. The estimate of τ_c of an AR(1) model from the monthly record yields $\tau_c \approx 0.5$ months and the estimate of β of an fGn model yields $\beta \approx 0.2$. This explains why the WVA fluctuation functions for the synthetic realizations of these two processes are very similar on time scales from months and up. Since the fluctuation function of the observed process is within the confidence areas for both models on these time scales we cannot select between AR(1) and fGn models on the basis of these monthly data.

In another method for distinguishing between noise models, we use the fact that a discrete-time sampling of the continuous-time OU process yields an AR(1) process, but that the lag-one correlation $\phi^{(\Delta t)}$ then will depend on the sampling time Δt. When we apply the relation

$$\tau_c^{(\Delta t)} = -\frac{\Delta t}{\log \phi^{(\Delta t)}} \tag{4.18}$$

for the decorrelation time, and estimate $\hat{\phi}^{(\Delta t)}$ from the AR(1) process resulting from sampling the OU process at time-lag Δt, we find that $\hat{\tau}_c \approx \tau_c$ as long as $\Delta t < \tau_c$, but when $\Delta t \gg \tau_c$ the AR(1) process cannot be distinguished from a white noise, resulting in $\hat{\tau}_c \propto \Delta t$. This feature is illustrated in Figure 4.7, which demonstrates explicitly that this method can be used to distinguish between AR(1) and weakly persistent fGn if the time resolution is better than τ_c, but otherwise not. In this figure the cyan area is the 95% confidence area for an ensemble of realizations of the Ornstein-Uhlenbeck (OU) process with $\tau_c = 10$. The grey area is the 95% confidence area for white noise. The application of this method to the Cheyenne monthly temperatures is shown in

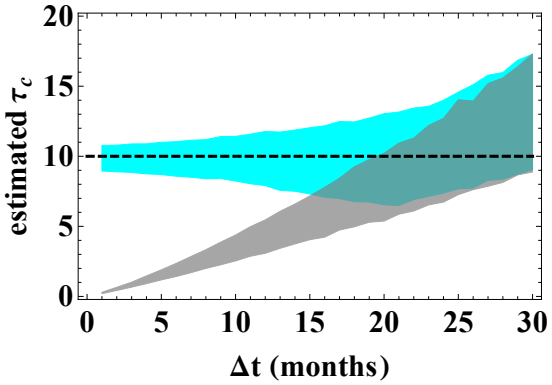


Figure 4.7: Estimated τ_c as a function of Δt for an Ornstein-Uhlenbeck process with $\tau = 10$ (cyan) and white noise (grey).

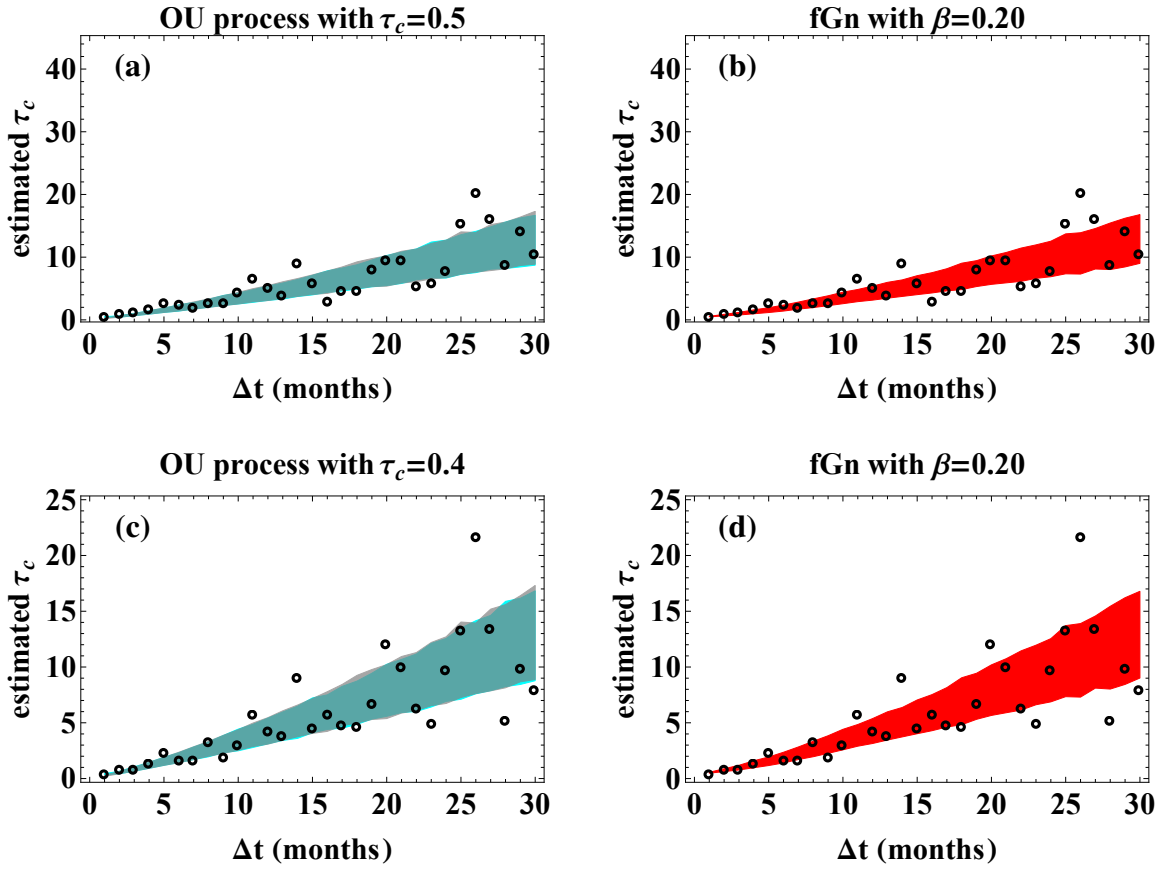


Figure 4.8: Panel (a) and (b) shows the estimated decorrelation time τ_c as a function of Δt for the temperature anomaly at Cheyenne as black circles. Ensembles are generated of synthetic realizations of two different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) in panel (b). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the coloured areas are the 95% confidence regions for these estimates. The grey area in panel (a) is the confidence region for τ_c from a white noise process. Panel (c) and (d) show the decorrelation time of the linearly detrended temperature record and for the synthetic realizations of the processes generated with parameters estimated from the detrended record.

Figure 4.8. In panel (a) and (c) the cyan areas are the 95% confidence areas for an OU process and the grey areas are the 95% confidence for white noise. They almost completely overlap. The red areas in panel (b) and (d) are the 95% confidence area for fGns. The parameters of the synthetic realizations are the same as those in Figure 4.6. The estimated $\hat{\tau}_c$ shows the behaviour of a white noise for almost all Δt when compared to synthetic realizations of an OU process, which is expected for OU processes with small τ_c. Since the white noise behaviour is dominant on most of the times scales, one would expect to see this when applying the standard LRM methods in Figure 4.4. In all cases the estimated $\beta > 0$, but error bars must also be taken into account. For WVA2, the lower error bar stretches below zero for given $\beta = 0.2$, indicating that

the temperature record could be white noise. However, for DFA2 the estimate is $\beta = 0.19 \pm 0.10$ and $\beta = 0.20 \pm 0.05$ using MLE after linear detrending. The results altogether indicate that the Cheyenne temperature may be described as an fGn with $\beta = 0.2$ superposed on a linear trend.

In Chapter 3 both the exponential and scale-free response models were applied to global land data, but from the deterministic response alone we could not really determine which response model that gives the best reproduction of the observed record. The clue to this assessment is found in the residual, i.e., the difference between the temperature record and the deterministic response. If this residual is analysed with DFA2, and the fluctuation function is compared with

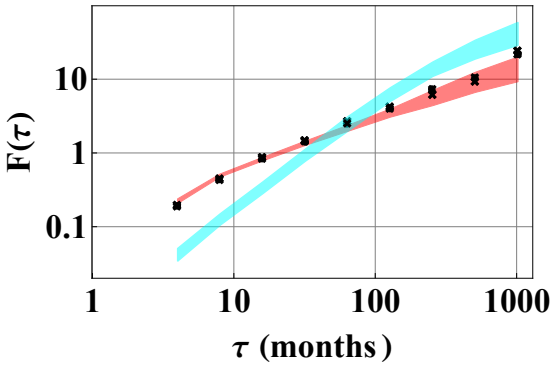


Figure 4.9: DFA2 applied to the residual from the deterministic response of the exponential response model (black circles) and scale-free response model (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn. The noises are produced with parameters estimated with the response models.

those produced from synthetic realizations of AR(1) and fGn, we obtain the results shown in Figure 4.9. The black circles and crosses are the fluctuation functions for the residuals from the exponential response model and scale-free response model, respectively. The coloured areas are the 95% confidence areas for synthetic noises with parameters estimated from the two models, where the cyan area is for AR(1) and the red area is for fGn. In Figure 3.4 the two deterministic responses were quite similar, and DFA2 applied to the residuals shows almost identical results. When comparing with synthetic noises, however, the scale-free response model is clearly favored. The fluctuation function falls mostly within the confidence area of the fGns, while it clearly deviates from the confidence area for the AR(1) model. The reason why we are able to select one model above the other for the global data set, but not for the Cheyenne record, is that the global data shows strong persistence even at time scales up to a century. When one tries to fit an AR(1) model to such data, the estimated τ_c exceeds a decade, and hence the fluctuation function of the synthetic AR(1) process has the steep slope $\alpha \approx 1.5$ ($\beta \approx 2$) corresponding to a Brownian motion for $\tau < 100$ months as shown by the cyan area in Figure 4.9. This is clearly distinguishable from the fluctuation function for the global record, for which $\alpha < 1$.

Two more examples are given to illustrate methods for distinguishing between AR(1) and fGn:

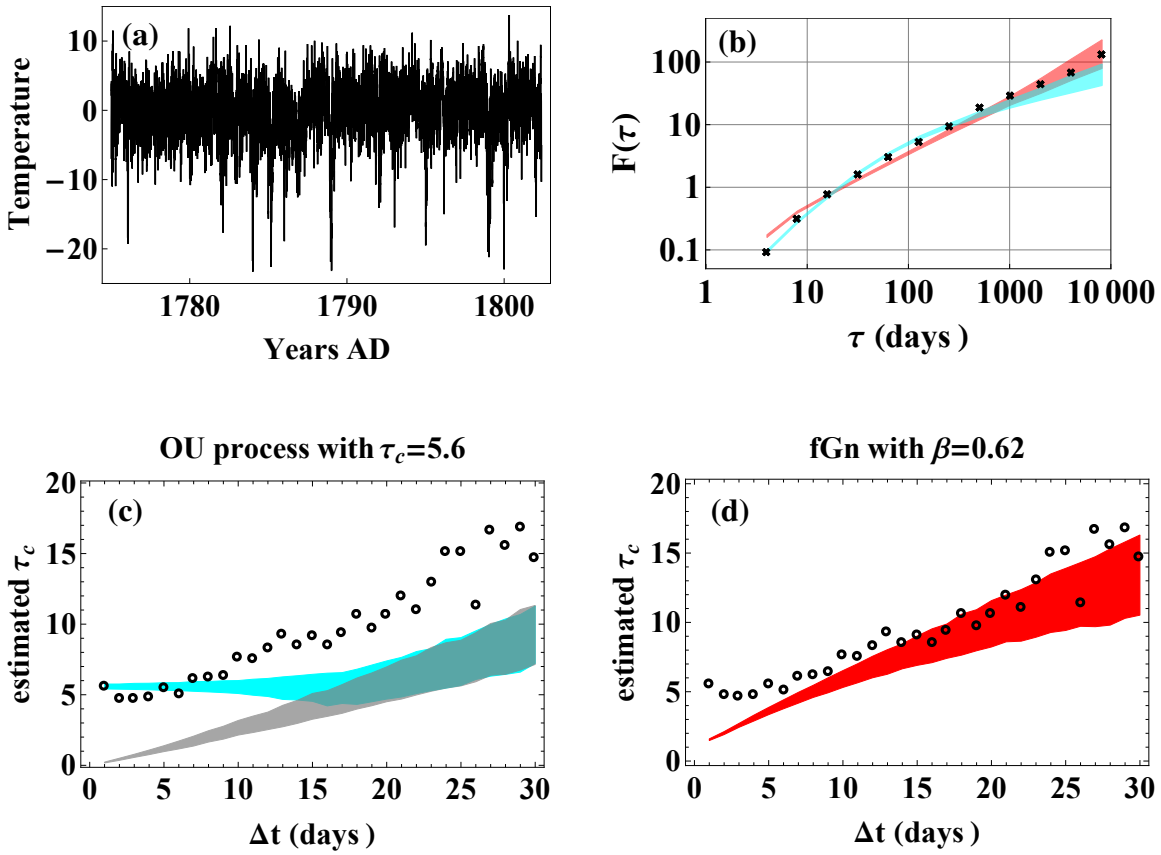


Figure 4.10: (a) 10 000 data points ($\sim$ 28 years) for Prague, daily temperature. (b) DFA2 applied to the record (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn. Panel (c) and (d) shows the estimated decorrelation time τ_c as a function of Δt for the record as black circles. Ensembles are generated of synthetic realizations of two different stochastic processes: An OU process (cyan) in panel (c), and fGns (red) in panel (d). The synthetic processes are generated with parameters estimated from the observed record by the MLE method for the OU processes and with DFA2 for the fGn, and the coloured areas are the 95% confidence regions for these estimates. The grey area in panel (c) is the confidence region for τ_c from a white noise process.

Daily mean temperature from Prague, the Czech Republic (*Klein Tank et al.*, 2002) and the Niño3 index (*Rayner et al.*, 2003). The first 10 000 data points from the Prague daily mean temperature record was used after removing the seasonality, with the results shown in Figure 4.10. The Prague temperature is widely used in temperature studies since it has a long record. In the analysis with DFA2 (Figure 4.10(b)), the AR(1) seems to be a better fit than fGn, at least on the smallest scales. The approach where the decorrelation time is estimated as a function of the sampling rate is shown in Figure 4.10(c) and (d), and $\tau_c = 5.6$ was found for the OU process. For the smallest Δt, there is a quite good fit for the estimated τ_c from the record to that of an OU process. For the larger sampling rates ($\Delta t > 10$), the estimated τ_c for the record are larger than

that of the OU process, which starts to behave as white noise. For these scales, the fGn seems to be a better fit. *Caballero et al.* (2002) found for daily mean temperature at three locations a good fit to the FARIMA$(1,d,1)$, which may capture both the AR(1) behaviour on small scales and LRM scaling on large scales. The results in Figure 4.10 suggest that this process also might well describe the daily temperature at Prague.

The Niño3 index is the area averaged monthly sea surface temperature from 5S-5N and 150W-90W. DFA2 (Figure 4.11(b)) does not show a perfect fit to neither noise process, but the record

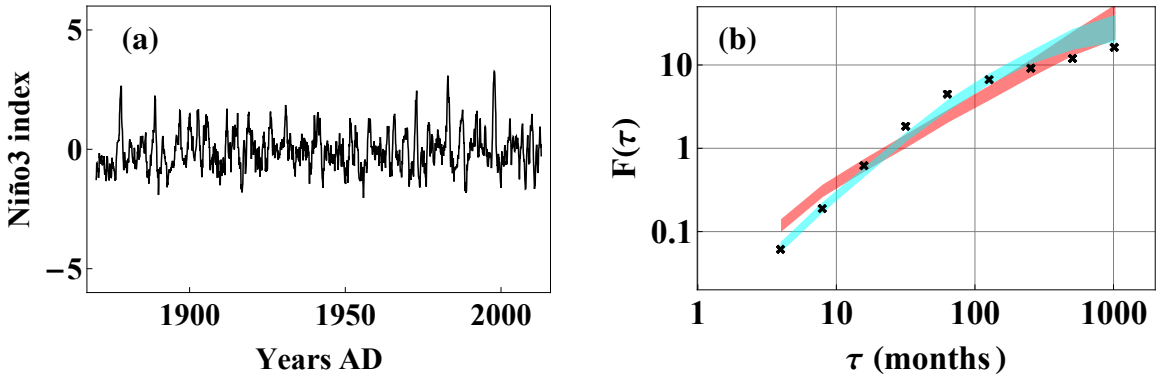


Figure 4.11: (a) The deseasonalized Niño3 index. (b) DFA2 applied to the record (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn. For AR(1), $\tau_c = 11.8$ was found with MLE, while $\beta = 0.62$ was found with DFA2.

has a closer fit to AR(1) than fGn. No trends are obvious from the record (Figure 4.11(a)), and DFA2 does not show influences typical for low-order polynomial functions or oscillations. It seems that the Niño3 index has a more complex underlying process than those described in this thesis.

Chapter 5

Literature Review

Earthquakes, rainfall and river flows all give rise to geophysical records with long-range memory (*Hurst et al.*, 1965, *Mandelbrot and Wallis*, 1969), but LRM has also been found in fields like medicine (*Goldberger and West*, 1987, *Stanley et al.*, 1992), finance (*Vandewalle and Ausloos*, 1997) and internet traffic (*Abry and Veitch*, 1998). Since LRM is so ubiquitous, this literature review is restrained to studies of Earth surface temperature records, which is the focus in this thesis. This includes observational temperature, temperature from model experiments, reconstructed temperature, and to some extent temperature proxies. It is common practice to remove daily and seasonal variations from the temperature records prior to analysis, if this is not already done in the record. Some analysis includes the removal of so called trends, often regarded as a slow variation or linear tendency in the time series, as explained in Chapter 3.

5.1 Instrumental Data

Universal Scaling Exponent?

Records of instrumental temperature from numerous stations have been available for a long time, and are widely used in studies of LRM. In the first approaches, the records are more or less randomly picked and studied to find if there is good scaling in agreement with LRM, and what the scaling exponent is if so. A variety of methods have been used. *Bodri* (1994) applied the rescaled range R/S analysis as defined in *Mandelbrot and Wallis* (1969) to annual mean temperature from 7 stations in Hungary. They found Hurst exponents between 0.72 and 0.81 with mean 0.77, corresponding to β between 0.44 and 0.62 with mean 0.54. *Bodri* (1995) applied the same approach to Central Europe annual mean temperature, yielding $H = 0.69$ ($\beta = 0.38$). This temperature was estimated using records from 224 stations by *Hansen and Lebedeff* (1987). The average periodogram for monthly mean temperature from 94 stations was estimated, yielding $\beta \approx 0.43$ in *Pelletier* (1997). *Pelletier and Turcotte* (1999) applied the periodogram to monthly averaged atmospheric temperature for Central England (CET) (*Parker et al.*, 1992) among other time series, resulting in $\beta \approx 0.47$. The CET is representative of a roughly triangular area of the United Kingdom enclosed by Lancashire, London and Bristol, and is widely studied since it is

the longest instrumental record in the world. In *Koscielny-Bunde et al.* (1998a) and *Koscielny-Bunde et al.* (1998b) DFA and a wavelet technique were applied to daily temperatures from 14 stations and 12 stations respectively. Roughly the same exponent of $\alpha \approx 0.65$, corresponding to $\beta \approx 0.3$, was found for all the temperature records. The result led to the question whether a universal scaling exponent was true for temperature. The exponent was found to be slightly smaller than in other studies of temperature from continental stations, and suggest that the R/S and periodogram estimates are influenced by trends that are eliminated in the DFA method.

Difference Between Air Temperature over Land and Oceans

The analyses mentioned so far, only regards records from continental stations. Several studies suggest that the temperature is more persistent at ocean sites than land sites. *Pelletier* (1997) analysed daily mean temperature from 90 maritime and 1000 continental stations. Average periodograms gave $\beta \approx 0.63$ for the maritime stations. For the continental stations $\beta \approx 0.37$ was found for frequencies less than $f \approx 1/(1 \text{ month})$ and $\beta \approx 1.37$ above this frequency. Both the periodogram and DFA were applied to daily temperature records from 20 continental and maritime stations in the USA in *Weber and Talkner* (2001). They found higher values of β for maritime stations ($0.30 < \beta < 0.36$) than for continental stations ($0.24 < \beta < 0.44$, but with most values around $\beta \approx 0.25$) in the low frequency range $f < 1/(10 \text{ days})$. These values are smaller than what *Pelletier* (1997) found, and they explained this by different ways of eliminating the annual cycle from the temperature records prior to the scaling analysis. *Eichner et al.* (2003) applied DFA to temperature from 95 stations all over the globe. They found that for continental stations, the scaling exponent is close to $\alpha = 0.65$ ($\beta = 0.3$). Temperature from island stations has a distribution between 0.65 and 0.85, with an average of 0.8, corresponding to β between 0.3 and 0.7 with mean 0.6. Their study confirms previous findings from DFA applied to continental temperature. For maritime stations, the result agrees well with *Pelletier* (1997), but the value of β is larger than what was found in *Weber and Talkner* (2001). *Monetti et al.* (2003) studied monthly and weekly sea surface temperature at different sites in the Atlantic and Pacific Oceans with DFA. A scaling exponent of $\alpha \approx 1.4$ ($\beta \approx 1.8$) was found for the North Atlantic sites and $\alpha \approx 1.2$ ($\beta \approx 1.4$) for the rest of the ocean sites for time scales below 10 months. In the region of the tropical Pacific where the El Niño-Southern Oscillation (ENSO) takes place, oscillations start to influence the fluctuation function above this time scale. Outside the ENSO region, $\alpha \approx 0.8$ ($\beta \approx 0.6$) for large time scales. The studies suggest that sea surface temperatures are *motions* ($\beta > 1$), while the temperatures at islands are persistent noises ($0 < \beta < 1$). The higher persistence at islands than at continents is probably due to the influence of the ocean. *Lennartz and Bunde* (2009) also applied DFA to a number of local temperature records, with results in agreement with, e.g., *Eichner et al.* (2003), *Monetti et al.* (2003).

Altitude Dependence

A few studies also investigate scaling differences for stations at low altitudes and stations located at mountains. *Talkner and Weber* (2000) and *Weber and Talkner* (2001) analysed daily minimum, maximum and mean temperatures with DFA and variations of spectral analysis. They found

lower values for the scaling exponent ($0.06 < \beta < 0.25$) at mountain sites, but these records did not show good scaling behaviour. These studies indicate that the scaling exponent is not the same all over the globe, but that air temperature at high altitudes is more random than other land air temperatures. In *Kurnaz* (2004a) monthly averages of maximum daily temperatures from 129 stations in the continental US were investigated with DFA to find scaling exponents $\alpha = 0.60 \pm 0.05$ ($\beta = 0.20 \pm 0.10$). The authors also looked for correlations between scaling exponents and elevation of weather stations, and between scaling exponents and distance from the stations to the ocean, without finding any clear patterns. However, they used the standard deviation of the temperature fluctuations to classify different climate types, finding slightly different exponents for each type. *Kurnaz* (2004b) applied DFA to monthly temperatures from 384 stations in the Western US, finding similar results.

Latitude Dependence

Pattantyús-Ábrahám et al. (2004) analysed daily temperature from 61 stations in Australia with DFA, finding that the scaling exponent varies from station to station. Generally it decreases with increasing distance from equator. They also found different scaling exponents for minimum and maximum temperature from the same station, but no pattern for magnitude. *Király and Jánosi* (2005) applied DFA to daily temperature records from 61 stations in Australia and 18 stations in Hungary. 48 of the Australian stations were based on the continent, while the remaining 13 were located on islands. They found a decreasing correlation exponent with increasing distance from the equator for the Australian station temperatures, in agreement with *Pattantyús-Ábrahám et al.* (2004). For the stations on islands the temperature analysis is in agreement with *Weber and Talkner* (2001) and *Monetti et al.* (2003). *Huybers and Curry* (2006) used the NCEP-NCAR instrumental re-analysis (*Kalnay et al.*, 1996) to find a global map for β using spectral analysis. In this study it was found that β is smaller over land than over ocean, but also that β is smaller toward higher latitudes, in agreement with *Király and Jánosi* (2005). They also found that the temperature for the Southern Hemisphere has a larger β than the Northern Hemisphere, probably because of larger ocean areas in the Southern Hemisphere. Temperature proxies together with observational data were analysed to get a patched periodogram for high latitudes and the tropics. Between annual and centennial time scales the tropical marine compilation has $\beta \approx 0.56$ and the high-latitude compilation $\beta \approx 0.37$, in agreement with the findings in their global map. At centennial time scales the spectra look more similar, but for time scales longer than centuries the tropics has $\beta \approx 1.29$ and the high latitudes $\beta \approx 1.64$. *Király and Jánosi* (2006) analysed several thousands of temperature records from the Global Daily Climatology Network with DFA. They did not find systematic dependence on geographic parameters similarly to *Pattantyús-Ábrahám et al.* (2004), *Király and Jánosi* (2005). It was concluded that the pattern for scaling exponent has no simple dependence on latitude, longitude or distance from oceans. *Vyushin and Kushner* (2009) did a study on monthly mean re-analysis air temperature (ERA-40, *Uppala et al.* (2005)), where the Hurst exponent was calculated at each longitude, latitude and pressure. They used both DFA and spectral methods. A decrease of H from the tropics to the extratropics was found, and the spectral methods showed a pronounced maximum in the Southern Hemisphere. The authors

attribute the latter finding to linear trends, since they did not find this with DFA.

Scaling Regimes

In studies of daily temperature, several scaling regimes have been found (*Pelletier*, 1997, *Talkner and Weber*, 2000, *Weber and Talkner*, 2001, *Caballero et al.*, 2002), although with different cross-over scale (from 3 days to 1 month). The typical time scale of general weather regimes is about 10 days, and up to this time scale the weather is highly correlated, explaining the high scaling exponents on these scales. Different cross-over scales for daily records may be found if they are from sites with different climate types. Different methods may also yield different cross-over scales. For monthly and annual instrumental records, cross-overs are usually not apparent. Longer records must be used to investigate if there are new regimes at scales above 100 years, and proxies going far back in time may be studied to indicate such changes. *Pelletier* (1997) studied a Vostok deuterium record converted into degrees Celsius. The Vostok station is located in Antarctica, and the record is the based on the isotopic fractions between ^{18}O and ^{2}H in ice cores. Three scaling regimes were found with the periodogram. For frequencies less than $f \approx 1/(40\text{ kyr})$, $\beta \approx 0$, i.e. the time series is a white noise. The regime between $f \approx 1/(40\text{ kyr})$ and $f > 1/(2\text{ kyr})$ displayed $\beta \approx 2$ (Brownian motion), and for the regime with $f \approx 1/(2\text{ kyr})$ $\beta \approx 0.5$ was found. *Pelletier and Turcotte* (1999) also applied the periodogram to the Vostok record, with the same result discussed in *Pelletier* (1997). Solar luminosity was studied in this paper, showing regimes in the periodogram similar to that of the Vostok record. The authors concluded that the physics of the radiating layer of the sun must strongly resemble the physics of the Earth's atmosphere. *Markonis and Koutsoyiannis* (2013) studied a number of temperature time series consisting of satellite, instrumental, proxy and reconstruction data. They applied a type of variogram combining the standard deviation as a function of scale for all the temperature series, spanning scales from 1 month to 50 million years. The authors did not find several scaling regimes, but an overall slope corresponding to $\beta = 0.84$, unlike *Pelletier* (1997). However, variogram methods are not always an accurate tool to investigate scaling properties, as they are in some cases biased and do not incorporate detrending. It is also disputable how well the variogram follows this slope at different time scales.

Difference Between Local, Regional and Global Temperature

Baillie and Chung (2002) analysed two different annual temperature series for the Northern Hemisphere, Southern Hemisphere and the entire globe (total of six records) with a FARIMA$(0,d,0)$ model. They found $d = 0.38$ and $d = 0.33$ ($\beta = (0.76, 0.66)$) for the global temperature, $d = 0.40$ and $d = 0.30$ ($\beta = (0.80, 0.60)$) for the Northern Hemisphere temperature and $d = 0.25$ and $d = 0.32$ ($\beta = (0.50, 0.64)$) for the Southern Hemisphere temperature. *Alvarez-Ramirez et al.* (2008) applied DFA to four monthly temperature sets for continents and oceans in the Northern and Southern Hemisphere. They estimated the scaling exponent for subsample windows of approximately 20 years with a 2-month slide to test for time-varying degrees of long-range memory, finding the same persistence pattern in time. Using the full record they found for Northern Hemisphere land temperature, Northern Hemisphere ocean temperature, Southern Hemi-

sphere land temperature and Southern Hemisphere ocean temperature the scaling exponents $\alpha \approx (0.69, 0.93, 0.78, 0.90)$, corresponding to $\beta \approx (0.38, 0.86, 0.56, 0.80)$, respectively. They also claimed that multifractality is present in the temperature data, and more evident for the land temperature. In *Lennartz and Bunde* (2009), DFA2 was applied to monthly land air, sea surface and combined temperatures of the globe, and the Northern and Southern Hemisphere. They found scaling exponents of $\alpha \approx 1.22$ ($\beta \approx 1.44$) for Northern Hemisphere sea surface and $\alpha \approx 0.79$ ($\beta \approx 0.58$) for Northern Hemisphere land temperature. The result for the sea surface temperature in the Northern Hemisphere is in agreement with *Monetti et al.* (2003), while the exponent for land temperature is higher for the Northern Hemisphere than for local stations. This may indicate that spatial averaging increases persistence. *Alvarez-Ramirez et al.* (2008) found smaller values for β for the Northern Hemisphere than *Lennartz and Bunde* (2009), and ocean temperature with $\beta < 1$, which is not in agreement with previous studies. Like *Huybers and Curry* (2006), they find higher values for the Southern Hemisphere than the Northern Hemisphere, while *Baillie and Chung* (2002) found the opposite. This might be due to different methods estimating β, or differences in the records (e.g. trends) influencing the estimation.

Trends

Bloomfield and Nychka (1992) studied the significance of a linear trend in a global annual temperature record (*Folland et al.*, 1990) using 7 different short-range and long-range memory noise models. They found that the trend was significant for all of the models. *Beran and Feng* (2002) suggested a semi-parametric method for simultaneous estimation of trends and parameters for FARIMA$(p,d,0)$. The method was applied to temperature data for the Northern Hemisphere. For land+sea data, $d = 0.38$ ($\beta = 0.76$) was found and the trend was just at the border of significance at the 5% level. For land temperature only, $d = 0.09$ ($\beta = 0.18$), and no significant long memory was found. The trend was clearly significant. *Koutsoyiannis* (2003) applied a variogram approach to a Northern Hemisphere temperature record (*Jones et al.*, 1998). For the Northern Hemisphere, $H = 0.88$ ($\beta = 0.76$) was found, and a trend study showed no strong evidence that temperature increase was of an unusual change of climate. They also studied a Paris temperature time series, yielding $H = 0.79$ ($\beta = 0.58$) and no significant trends. *Craigmile et al.* (2004) suggested using the discrete wavelet transform to extract a polynomial trend from an LRM record. They applied this approach to a 150 year record of the sea surface temperature from the Seychelles, in the Indian Ocean (*Charles et al.*, 1997), and found that the large scale variations in the record could be attributed to the stochastic variations rather than to a deterministic trend. In *Gil-Alana* (2005) the monthly Northern Hemisphere temperature record (*Jones and Briffa*, 1992) was examined by means of fractional integration techniques. It was found that the record follows a FARIMA$(0,d,0)$ with $0.3 < d < 0.4$ $(0.6 < \beta < 0.8)$, and that there is a statistically significant linear trend in the record. *Cohn and Lins* (2005) considered the Northern Hemisphere temperature by *Jones et al.* (1999). They found almost the same value for the slope of a linear trend under different noise models, but different significance levels. For white noise and short-range memory processes, the trend was significant, while for long-range memory processes it was not. *Fatichi et al.* (2009) analysed 26 temperature records in the Tuscany region with three different non-parametric trend detection procedures, using FARIMA processes to model the records. They

found significant linear trends for 9 of the station records. *Rybski and Bunde* (2009) studied trend significance in temperature records from six stations. They used a DFA-based technique to estimate linear trends, and found the probability that a given long-term correlated record contains a certain trend. They found significant trends for two of the six records. *Lennartz and Bunde* (2009) performed a trend analysis to decide if a linear trend could be a natural part of an LRM record, or if the trend was of external origin. They found that the trends were more significant in global than local records, and that the annual increase over the last 50 years was a weaker indicator of an anthropogenic trend than the lower increase over the last 100 years. *Franzke* (2010) used spectral analysis to estimate LRM parameters for temperature at eight Antarctic stations, finding $0.16 < \beta < 0.56$. A trend study was done under two null models: that the data are represented by an AR(1) (SRM) and that the data are represented by a FARIMA$(0,d,0)$ (LRM). A significant trend was found for 3 stations under the SRM hypothesis, and for 1 station under the LRM hypothesis. In *Franzke* (2012a) the significance of trends in temperature records was tested against three null models: SRM, LRM and phase scrambling. The records analysed were daily temperature records from central England (CET), Stockholm, Faraday-Vernadsky and Alert. The last two stations are in two polar regions that have experienced some of the most dramatic environmental changes in the last two decades. Different trends were investigated, and the cubic polynomial fit had the smallest RMS error for all four time series. The temperature record at Faraday-Vernadsky showed the largest warming, which could not arise by chance for any of the null models. For CET and Stockholm temperature records, the warming trends were significant under the SRM and phase scrambling null model, but not for LRM. The Alert temperature record had a warming trend which could be reproduced by all three models, i.e., the trend was not significant. A similar approach was done in *Franzke* (2012b) on daily mean temperatures from 109 stations in the Eurasion Arctic region. This resulted in significant trends in 17 temperature records against the SRM null model, in 3 temperature records against the LRM null model and 8 temperature records against the phase scrambling null model.

The trend studies show variable results for how significant trends are in temperature records. The significance depends on the null model (LRM/SRM), the trend model and the location, length, and temporal and spatial resolution of the records. When the trend is found to be insignificant, it does not mean that the record is not affected by global warming, but rather that properties like high variance and persistence may make it hard to detect a global warming signal.

Comparing Different Models

Talkner and Weber (2000), *Weber and Talkner* (2001) found slightly different scaling exponents using the periodogram and DFA. *Caballero et al.* (2002) estimated the scaling parameter d for daily mean temperature from Central England, Chicago and Los Angeles using four methods: Periodogram, aggregated variance, differenced variance and maximum likelihood estimation of an FARIMA$(1,d,1)$ process. They found that long memory was present with two scaling regimes in the temperature data. This was best captured by the FARIMA$(1,d,1)$ noise model with $d = (0.20, 0.13, 0.23)$, corresponding to $\beta = (0.40, 0.26, 0.46)$, for Central England, Chicago and Los Angeles respectively. The cross-over was around $1/(6 \text{ days})$. The other meth-

ods yielded slightly different scaling parameters. If pure noises are studied, the different methods should yield similar scaling exponents. Different scaling exponents may be due to influence of trends or multiple scaling regimes.

Percival et al. (2001) analysed the Sitka, Alaska, winter air temperature record, and fitted two models, AR(1) and FARIMA$(0,d,0)$, to the time series. They compared the autocorrelation function and periodogram with the theoretical ACFs and PSDs for the models with parameters from the record, and applied a goodness-of-fit test. They found that there was no statistical evidence to favour one model over the other.

5.2 Simulated Temperature From Model Experiments

Many different climate models have been used in the study of LRM in simulated temperature, but most of them are coupled Atmosphere-Ocean General Circulation Models (AOGCM). Different climate model experiments from the same models are often available, where the difference lies in which forcings are kept fixed and which are dynamic. The forcings used in the model experiments usually consist of total solar forcing, volcanic forcing, CO_2 or GHG forcing and aerosol forcing, and for some model experiments forcings related to land use change and orbital forcing are also included.

Controlruns/Fixed Forcings

In *Fraedrich and Blender* (2003) DFA was applied to global fields of observed and simulated surface temperatures from an AOGCM climate model experiment. The result from observational data was mostly in agreement with previous studies of temperature in oceanic and coastal regions, but the authors found $\alpha \approx 0.5$ corresponding to white noise in inner continents. A 1000-year simulation from the model experiment yielded similar exponents to what was found for the observational data in this study. They did not find decreasing exponents with increasing distance from equator like *Király and Jánosi* (2005), who comment that this might be due to lower spatial resolution over Australia in *Fraedrich and Blender* (2003). *Blender et al.* (2006) compared application of DFA to Greenland ice core $\delta^{18}O$ time series with near surface temperature from an AOGCM simulation. The analysis showed LRM scaling up to millennial time scales during the Holocene in the ice core data, and that the LRM was reproduced by a 10000 year simulation.

Dynamic CO_2 Forcing

Syroka and Toumi (2001) studied persistence in observed temperature, the NCEP re-analysis (*Kalnay et al.*, 1996) and temperature from a HadCM3 model experiment with daily resolution. Data from Central England (CET), the El Niño region and global data were used. A variogram approach was used to determine scaling regimes and exponents. They found different scaling regimes for the different records, and anti-persistence on scales larger than 1 year in the El Niño

region. The authors found that the temperature from the model experiment reproduces high persistence on time scales less than one year, but that the persistence on larger time scales is smaller than for the global temperature and CET. For the El Niño region the model experiment and NCEP produce temperatures with similar features. The authors also concluded that neither the observations nor the model simulations can be interpreted in terms of an AR(1) process by comparing power spectra of the data and synthetic noise. One should be careful with the interpretation of these results, as variogram approaches do not incorporate detrending, and cannot yield $\beta > 1$. In the El Niño region, the temperature have oscillations, which also affects the scaling behaviour. Still, this study may give indications of how well the climate model simulation of temperature reproduces the observed temperature.

In *Bunde et al.* (2001), observed maximum daily temperatures from 6 sites, among them Prague, was studied with DFA. Temperature data from Prague from different climate model experiments with AOGCM models were analysed with the same methods, and the results compared. They used time series cut off at the year 1992, and time series extending into the future. For all the observational data they found an $\alpha \approx 0.65$ corresponding to $\beta \approx 0.3$ for time scales above 10 days. The temperatures from the model experiments showed good scaling for the data from the CSIROMK2 experiment, with a scaling exponent close to that for the Prague record. The temperatures from the experiments with ECHAM4/OPYC3 and HadCM3 showed a crossover after about 3 yr, where the data had an exponent corresponding to white noise. In *Govindan et al.* (2001) a similar study was done, but with comparison between observational data and data from model experiments from two sites, Prague and Melbourne. The same models were used. The Prague results were discussed as in *Bunde et al.* (2001). For Melbourne the results were similar, except for the temperature from the HadCM3 experiment, which yielded a slightly higher exponent than the observational data. In *Vjushin et al.* (2002) temperature records at four sites from seven climate model experiments were analysed with DFA. All the models were AOGCM's. Historical forcing records were used up to 1990, and a 1% increase in CO_2 level was assumed after that. The authors found that the different model experiments varied significantly with regards to LRM, and also found variations from location to location. Scaling exponents differed from those found for observational temperature record. They concluded that the gradual addition of CO_2 makes the temperature from the model experiments lose their memory, and that the results may be improved by changing the method for adding CO_2.

Blender and Fraedrich (2003) did a similar analysis as *Fraedrich and Blender* (2003) with temperature from two different model experiments with dynamic greenhouse gas forcing included. The results of these two studies are in agreement.

Dynamic CO_2 and Aerosol Forcing

Govindan et al. (2002) made another study applying DFA to observational data and temperature from climate model experiments. They used records from six sites, and temperature from more model experiments than previously used in *Bunde et al.* (2001) and *Govindan et al.* (2001). In addition, both experiments with dynamic greenhouse gas forcing and with dynamic greenhouse

gas plus aerosol forcing were included. The main conclusion was that the temperature from the model experiments fail to reproduce the scaling behaviour found for observational data, and that the models display large differences in scaling at different sites. Of the two scenarios, the one with dynamic greenhouse gas plus aerosol forcing performed better. The authors claimed that since LRM is underestimated in the temperature from the climate model experiments, it follows that anticipated global warming is overestimated. *Bunde and Havlin* (2002) applied DFA to mean daily temperature from a number of sites with different climate, as the locations are on continents, coast lines, islands and in the ocean, finding exponents in agreement with the DFA studies of instrumental records already mentioned. They compared to simulated temperature, where three types of climate model runs were used: control runs with all forcings fixed, run with dynamic greenhouse gas forcing and run with dynamic greenhouse gas plus aerosol forcing. The authors found that the experiments with dynamic greenhouse gas and aerosol forcing produce temperatures that perform best with regards to scaling exponents, but they are not perfectly reproducing that found for observational temperature. *Govindan et al.* (2003) analysed the temperature volatility, i.e., the increments, from a few selected sites and compared with temperature volatility from climate model experiments, using DFA. For the observational data they found scaling exponents similar to that of the direct analysis of observational temperature. Temperature from model experiments were obtained with the same three types of model runs as in *Bunde and Havlin* (2002). Here the temperature volatility showed a wider range of scaling exponents and conclusions were harder to draw.

Including Dynamic Volcanic Forcing

In *Vyushin et al.* (2004) temperature from model experiments with no forcings, greenhouse gas, sulphate aerosol, ozone, solar, volcanic forcing and various combinations were studied (these forcings were dynamic, other forcings fixed). Scaling exponents for temperature at 16 land sites and 16 sites in the Atlantic Ocean were estimated. They found that dynamic volcanic forcing was the most relevant for obtaining scaling exponents close to those found for observational records. *Rybski et al.* (2008) used model experiments with constant forcing and with dynamic solar, volcanic and greenhouse gas forcing. They analysed data from grid cells all over the globe. They found that for the forced run experiment, the temperature showed a scaling exponent in agreement with observational temperature, but that the temperature from the control run generally yields somewhat lower scaling exponents.

Comparing Different Models

Vyushin et al. (2012) analysed a large number of records from temperature re-analysis and temperature from model experiments. They used methods for estimating both the lag-one autocorrelation ϕ for an AR(1) and the scaling parameter H and made geographical maps of these noise parameters. The H values for the simulations were largely consistent with the results in *Fraedrich and Blender* (2003), *Blender and Fraedrich* (2003), *Blender et al.* (2006), *Rybski et al.* (2008). Goodness-of-fit tests were also applied, but it was found that neither AR(1) nor fGn provided a better fit to the observed and simulated data. Their method was similar to the one we

applied in Figures 4.7 and 4.8, but the sampling intervals considered was limited to monthly and annual. Application of a larger range of sampling rates would have given a clearer picture and favoured the fGn model. *Zhu et al.* (2010) used experiments from COSMOS to investigate temperature simulations all over the globe. LRM properties and forecast experiments based on an AR(1) linear predictor were studied, with various results. They demonstrated the existence of long-range memory in the near-surface temperature field in high-latitude oceans, while in areas with LRM, the prediction skills of the AR(1) predictor were poor. In the central South Atlantic on the other hand, the predictable component by AR(1) was enhanced due to local strong decadal and bicentennial fluctuations, while LRM scaling were poor.

5.3 Reconstructed Temperature

Rybski et al. (2006) apply DFA2 to six reconstructed temperature records for the Northern Hemisphere. This resulted in scaling exponents corresponding to $0.6 < \beta < 1$. For the Moberg reconstructed temperature they found $\beta = 0.72$. In their study the authors concluded that their work support the claim that the most recent observed warming is inconsistent with the hypothesis of purely natural dynamics. In *Mills* (2007) the periodogram was applied to the Moberg Northern Hemisphere reconstructed temperature, as well as fitting a FARIMA$(2,d,2)$ to the record. The author analysed the temperature divided into different subperiods as well as the full record, to find $d \approx 0.5$ ($\beta \approx 1$) for most of the periods. The periodogram and FARIMA$(2,d,2)$ approach gave almost the same scaling parameter with a few exceptions. *Rea et al.* (2011) analysed six temperature reconstructions for Northern Hemisphere, Western USA, Colorado (USA), Shihua (near Beijing, China), Tasmania (Australia) and Torneträsk (Sweden) with 11 different estimators of the scaling exponent. They found a wide range of scaling exponents for each location, and concluded that although some of the methods suggest a good fit to long-range memory processes, there are phenomena present in the data, e.g., oscillations, that cannot be explained by LRM. They claim that the apparent long-range memory is merely an artefact of the method of analysis, but do not consider the possibility that the reconstructions could be a long-range memory process superposed on a trend. *Halley and Kugiumtzis* (2011) did a non-parametric testing of linear trends in 9 temperature reconstructions for the Northern Hemisphere using a type of surrogate data preserving the LTP structure of the records. They found that the rising trend had a low probability of being natural fluctuations.

For reconstructed temperature there are few studies, and it is harder to draw an overall picture. The studies of the Northern Hemisphere reconstructions mostly indicate that this temperature is a highly persistent noise with a superposed trend.

Chapter 6

Summary of Papers

The papers in this thesis focus on long-range memory in time series of surface temperature. We have mostly studied global and hemispheric temperature means, since such records are far less studied than local temperature time series in the existing literature. A regional instrumental record, the Central England temperature (CET) was included in Paper I, together with global land temperature, combined global land temperature and ocean temperature and a Northern Hemisphere (NH) temperature reconstruction. In Paper II, only global land and global ocean temperature were used. In Paper III, the focus was on Northern Hemisphere temperature. Local data at Reykjanes Ridge were also studied, using a model temperature simulation, reconstructed temperature based on proxies and reconstructed temperature based on temperature observations. Proper error bars for the estimated scaling exponents and more rigorous testing for LRM and trends including Monte Carlo simulations distinguish Paper I from previous work on LRM in temperature records. The local data in Paper III were included to illustrate that local ocean temperature also is strongly persistent, as opposed to local continental temperature which is usually random or only weakly persistent.

The scaling behaviour was investigated with a number of methods. In Paper I, WVA, DFA, periodogram, variogram and autocorrelation function were used to find if the correlation structure of the records was consistent with that of fGns, with positive result after proper detrending. However, other noise models were not considered. The significance of trends under an AR(1) model was included in Paper II, and therefore an approach for finding the noise model that best describes the temperature records was applied. The correlation time τ_c was estimated as a function of the sampling rate Δt for the full data set, the detrended data set, and synthetic data under the two null hypothesis. This showed that for the land temperature, the time series was more consistent with fGn than AR(1) after detrending. The ocean temperature was most consistent with fBm without detrending, and with fGn after linear detrending. Due to the high persistence in the ocean temperature, we could not decide whether fGn or fBm was the best model, but AR(1) was rejected. In Paper III, the periodogram, DFA and WVA was applied to temperature from model experiments and reconstructed temperature for the Northern Hemisphere and Reykjanes Ridge. Based on the results from these methods and the results in Paper II, it was not found necessary to test an AR(1) hypothesis. To avoid the effects of linear trends associated with anthropogenic

global warming, the time series were cut off at 1750 AD. Instead of assuming that the records should be divided into a trend and a noise, an approach considering a deterministic and stochastic response to external forcing was performed.

Our papers confirm that global and hemispheric temperature means are more persistent than local temperature, and that temperature over oceans is more persistent than temperature over land. In Paper I, we found for CET $H = 0.64 \pm 0.07$, corresponding to $\beta = 0.28 \pm 0.14$, for global land temperature $H = 0.75 \pm 0.07$ ($\beta = 0.50 \pm 0.14$), for combined land and ocean temperature $H \approx 1$ ($\beta \approx 1$) and for the NH reconstruction $H = 0.9 \pm 0.1$ ($\beta = 0.8 \pm 0.2$). In Paper III we found a somewhat lower scaling exponent for the NH reconstruction, $0.6 < \beta < 0.7$. For the NH temperature simulated by climate model experiments, $0.6 < \beta < 1$ was found, and the temperature from Reykjanes Ridge showed slightly lower persistence with $0.4 < \beta < 0.6$.

Trends are important in our papers in two ways: they tend to influence the estimation of memory exponents, and LRM tends to influence the statistical significance of trends. In Paper I, three trend models were considered: linear, cubic, and 7th order polynomial. For CET, the linear detrending resulted in the best scaling behaviour, while for the global land temperature and combined global land and ocean temperature, the cubic trend model gave the best result, although the results for the latter record were slightly harder to interpret due to $\beta \approx 1$. For the NH reconstruction, polynomial detrending did not give good scaling, so a wavelet filtering approach was done to simulate an oscillation. The record was well described by an fGn with such a trend superposed. In Paper II, a rigorous study of significance of trends in global land temperature and global ocean temperature was performed. A trend model consisting of a linear function and an oscillation was chosen, but the procedure could have been used for any trend model. The method of hypothesis testing was emphasized, as the testing of significance of trends in LRM records previously have been done with different approaches, leading to different conclusions. In our approach, a correlated noise was chosen as the null model, with the alternative model that a trend was present in the temperature records. Three noise models were used: AR(1), fGn and fBm (for ocean temperature only). The noise parameters for each noise model were estimated for the two records, and ensembles of synthetic noises with the same parameters constructed. The trend model parameters were estimated for each ensemble, and used to obtain 95% confidence contours of the distribution for the trend parameters. When the trend parameters for the ocean temperature were compared to the 95% confidence, the trend was significant for AR(1) and fGn, but not for the fBm null model. For the land temperature, the linear part of the trend was clearly significant. The null hypothesis could then be rejected, and a new null model including the linear trend was formed. It was then found that the oscillation also was significant in the land temperature record.

Chapter 7

Concluding Remarks

The study of different surface temperature time series, including instrumental records, reconstructions and climate model simulations, shows that LRM is present on time scales from months to centuries. Most of the time series can be described as persistent noise. Global ocean temperature may be described as a highly persistent noise or a nonstationary motion, but the distinction is unclear because of biases and errors in the methods when $\beta \approx 1$. Significant rising trends can be found in temperature time series over the last 100-200 years, but may be hard to detect in local records due to high variance and in ocean records due to high persistence. The global land temperature over the last 160 years works as a great example of a temperature record showing a clear rising trend as well as an oscillation with a period of ~ 70 years. The study of Northern Hemisphere temperature from climate model experiments shows that external forcing alone cannot explain LRM in temperature, since LRM is found in both temperature from control runs and in the residual from a deterministic response to forcing. The scaling exponents are in this case close to that of the temperature from the experiments with full dynamic forcing. Although some of the previous studies of simulated temperature indicate that some types of dynamic forcing is important for reproducing LRM in agreement with that found in observational data, this is not always the case. These studies focus on local data, so the smaller scaling exponents could perhaps be explained by other features masking the LRM, e.g., higher variance, oscillations, or generally lower persistence in local than in global data. The lack of persistence in control runs is not found for the Northern Hemisphere mean used in our studies. This suggests that LRM arises from internal dynamics of the climate system, and since sea surface temperature is more persistent than land air temperature, ocean dynamics must be a crucial component for LRM in temperature.

Bibliography

Abry, P. and P. Veitch. Wavelet analysis of long-range-dependent traffic. *Information Theory, IEEE Transactions on*, 44(1):2 –15, 1998.

Alvarez-Ramirez, J., J. Alvarez, L. Dagdug, E. Rodriguez, and J. C. Echeverria. Long-term memory dynamics of continental and oceanic monthly temperatures in the recent 125 years. *Physica A*, 387(14):3629 – 3640, 2008. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/j.physa.2008.02.051.

Baillie, R. T. and S.-K. Chung. Modeling and forecasting from trend-stationary long memory models with applications to climatology. *Int. J. Forecasting*, 18(2):215 – 226, 2002. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(01)00154-6.

Beran, J. *Statistics for Long-Memory Processes*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1994.

Beran, J. and Y. Feng. SEMIFAR models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Comput. Stat. Data An.*, 40(2):393 – 419, 2002. doi: http://dx.doi.org/10.1016/S0167-9473(02)00007-5.

Blender, R. and K. Fraedrich. Long time memory in global warming simulations. *Geophys. Res. Lett.*, 30, 2003. doi: 10.1029/2003GL017666.

Blender, R., K. Fraedrich, and B. Hunt. Millennial climate variability: GCM-simulation and Greenland ice cores. *Geophys. Res. Lett.*, 33(4), 2006. doi: 10.1029/2005GL024919.

Bloomfield, P. and D. Nychka. Climate spectra and detecting climate change. *Climatic Change*, 21(3):275–287, 1992.

Bodri, L. Fractal analysis of climatic data: Mean annual temperature records in Hungary. *Theoretical and Applied Climatology*, 49(1):53–57, 1994. ISSN 0177-798X. doi: 10.1007/BF00866288.

Bodri, L. Short-term climate variability and its stochastic modeling. *Theor. Appl. Climatol.*, 51 (1-2):51–58, 1995. ISSN 0177-798X. doi: 10.1007/BF00865539.

Box, G. E. and G. M. Jenkins. *Time series analysis; forecasting and control*. San Francisco: Holden-Day, 1970.

Brohan, P., J. J. Kennedy, I. Harris, S. F. Tett, and P. D. Jones. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, 111 (D12), 2006.

Bunde, A. and S. Havlin. Power-law persistence in the atmosphere and in the oceans. *Physica A*, 314, 2002.

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

Caballero, R., S. Jewson, and A. Brix. Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation. *Clim. Res.*, 21:127–140, 2002.

Charles, C., D. Hunter, and R. G. Fairbanks. Interaction between the ENSO and the Asian monsoon in a coral record of tropical climate. *Science*, 277(5328):925–928, 1997.

Cohn, T. A. and H. F. Lins. Nature's style: Naturally trendy. *Geophys. Res. Lett.*, 32(23), 2005. doi: 10.1029/2005GL024476.

Craigmile, P. F., P. Guttorp, and D. B. Percival. Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics*, 15(4):313–335, 2004.

Delignieres, D., S. Ramdani, L. Lemoine, K. Torre, M. Fortes, et al. Fractal analyses for 'short' time series: A re-assessment of classical methods. *J. Math. Psychol.*, 50(6):525 – 544, 2006. ISSN 0022-2496. doi: http://dx.doi.org/10.1016/j.jmp.2006.07.004.

Eichner, J. F., E. Koscielny-Bunde, A. Bunde, S. Havlin, and H.-J. Schellnhuber. Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. *Phys. Rev. E*, 68, 2003. doi: 10.

Fatichi, S., S. M. Barbosa, E. Caporali, and M. E. Silva. Deterministic versus stochastic trends: Detection and challenges. *J. Geophys. Res.*, 114(D18), 2009. doi: 10.1029/2009JD011960.

Flandrin, P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inf. Theory*, 38(2):910–917, 1992.

Folland, C. K., T. Karl, and K. Vinnikov. Observed climate variations and change. *Climate change: the IPCC scientific assessment*, 195:238, 1990.

Fraedrich, K. and R. Blender. Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models. *Phys. Rev. Lett.*, 90:108501, 2003. doi: 10.1103/PhysRevLett.90.108501.

Franzke, C. Long-range dependence and climate noise characteristics of Antarctic temperature data. *J. Climate*, 23:6074–6081, 2010. doi: doi:http://dx.doi.org/10.1175/2010JCLI3654.1.

Franzke, C. Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J. Climate*, 25(12):4172–4183, 2012a.

Franzke, C. On the statistical significance of surface air temperature trends in the Eurasian Arctic region. *Geophys. Res. Lett.*, 39(23), 2012b.

Franzke, C. L., T. Graves, N. W. Watkins, R. B. Gramacy, and C. Hughes. Robustness of estimators of long-range dependence and self-similarity under non-gaussianity. *Philos. T. Roy. Soc. A*, 370(1962):1250–1267, 2012.

Gil-Alana, L. A. Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques. *J. Climate*, 18(24):5357–5369, 2005.

Goldberger, A. L. and B. J. West. Fractals in physiology and medicine. *Yale J. Biol. Med.*, 60: 421–435, 1987.

Govindan, R. B., D. Vjushin, S. Brenner, A. Bunde, S. Havlin, et al. Long-range correlation and trends in global cliamte models: Comparison with real data. *Physica A*, 294:239–248, 2001.

Govindan, R. B., D. Vjushin, A. Bunde, S. Brenner, S. Havlin, et al. Global climate Models Violate Scaling of the Observed Atmospheric Variability. *Phys. Rev. Lett.*, 89(2), 2002. doi: 10.1103/PhysRevLett.89.028501.

Govindan, R., A. Bunde, and S. Havlin. Volatility in atmospheric temperature variability. *Physica A*, 318(3–4):529 – 536, 2003. doi: 10.1016/S0378-4371(02)01552-2.

Grossmann, A. and J. Morlet. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. *SIAM J. Math. Anal.*, 15(4):723–736, 1984.

Halley, J. and D. Kugiumtzis. Nonparametric testing of variability and trend in some climatic records. *Climatic Change*, 109(3-4):549–568, 2011. doi: 10.1007/s10584-011-0053-5.

Hansen, J. and S. Lebedeff. Global trends of measured surface air temperature. *J. Geophys. Res.*, 92(D11):13345–13372, 1987.

Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann. Earth's energy imbalance and implications. *Atmos. Chem. Phys.*, 11(24):13421–13449, 2011.

Heneghan, C. and G. McDarby. Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Phys. Rev. E*, 62:6103–6110, 2000. doi: 10.1103/PhysRevE.62.6103.

Hu, K., P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E*, 64:011114, 2001.

Hurst, H. E. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808, 1951.

Hurst, H. E., R. P. Black, and Y. Simaika. *Long-term storage: an experimental study*. Constable, 1965.

Huybers, P. and W. Curry. Links between annual, Milankovitch and continuum temperature variability. *Nature*, 441:329–332, 2006. doi: 10.1038/nature04745.

Jones, P. and K. Briffa. Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details. *The Holocene*, 2(2):165–179, 1992.

Jones, P., K. Briffa, T. Barnett, and S. Tett. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. *The Holocene*, 8(4):455–471, 1998.

Jones, P., D. Lister, T. Osborn, C. Harpham, M. Salmon, et al. Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res.*, 117(D5), 2012.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, 37(2):173–199, 1999.

Jones, P. D. and A. Moberg. Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, 16(2):206–223, 2003.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, et al. The NCEP/NCAR 40-year reanalysis project. *B. Am. Meteorol. Soc.*, 77(3):437–471, 1996.

Kantelhardt, J. W., E. Koscielny-Bunde, H. H. Rego, S. Havlin, and A. Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A*, 295(3-4):441 – 454, 2001.

Király, A. and I. M. Jánosi. Detrended fluctuation analysis of daily temperature records: Geographic dependence over Australia. *Meteorol. Atmos. Phys.*, 88:119–128, 2005. ISSN 0177-7971. doi: 10.1007/s00703-004-0078-7.

Király, I. B. A. and I. M. Jánosi. Correlation properties of daily temperature anomalies over land. *Tellus A*, 58(5):593–600, 2006. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2006.00195.x.

Klein Tank, A., J. Wijngaard, G. Können, R. Böhm, G. Demarée, et al. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.*, 22(12):1441–1453, 2002.

Koscielny-Bunde, E., A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, et al. Indication of a Universal Persistence Law Governing Atmospheric Variability. *Phys. Rev. Lett.*, 81:729–732, 1998a. doi: 10.1103/PhysRevLett.81.729.

Koscielny-Bunde, E., H. Eduardo Roman, A. Bunde, S. Havlin, and H.-J. Schellnhuber. Long-range power-law correlations in local daily temperature fluctuations. *Philos. Mag. B*, 77(5): 1331–1340, 1998b. doi: 10.1080/13642819808205026.

Koutsoyiannis, D. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol. Sci. J.*, 48(1):3–24, 2003.

Kurnaz, M. L. Application of Detrended Fluctuation Analysis to Monthly Average of the Maximum Daily Temperature to Resolve Different Climates. *Fractals*, 12(04):365–373, 2004a. doi: 10.1142/S0218348X04002665.

Kurnaz, M. L. Detrended fluctuation analysis as a statistical tool to monitor the climate. *J. Stat. Mech. Theor. Exp.*, 2004(07):P07009, 2004b.

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

Malamud, B. D. and D. L. Turcotte. Self-affine time series: measures of weak and strong persistence. *J. Statist. Plann. Inference*, 80:173–196, 1999.

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

Mandelbrot, B. B. and J. R. Wallis. Some long-run properties of geophysical records. *Water Resour. Res.*, 5(2):321–340, 1969. ISSN 1944-7973. doi: 10.1029/WR005i002p00321.

Markonis, Y. and D. Koutsoyiannis. Climatic Variability Over Time Scales Spanning Nine Orders of Magnitude: Connecting Milankovitch Cycles with Hurst-Kolmogorov dynamics. *Surv. Geophys.*, 34(2):181–207, 2013. ISSN 0169-3298. doi: 10.1007/s10712-012-9208-9.

Matheron, G. Principles of geostatistics. *Econ. Geol.*, 58(8):1246–1266, 1963.

McLeod, A. I., H. Yu, and Z. L. Krougly. Algorithms for Linear Time Series Analysis: With R Package. *J. Stat. Softw.*, 23, 2007.

McLeod, A. and K. W. Hipel. Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst Phenomenon. *Water Resour. Res.*, 14(3):491–508, 1978.

Mielniczuk, J. and P. Wojdyłło. Estimation of Hurst exponent revisited. *Comput. Stat. Data An.*, 51(9):4510 – 4525, 2007. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/j.csda.2006.07.033.

Mills, T. C. Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends? *J. Roy. Stat. Soc. A Sta.*, 170(1):83–94, 2007. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2006.00443.x.

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

Parker, D. E., T. Legg, and C. K. Folland. A new daily central England temperature series, 1772–1991. *Int. J. Climatol.*, 12(4):317–342, 1992.

Pattantyús-Ábrahám, M., A. Király, and I. M. Jánosi. Nonuniversal atmospheric persistence: Different scaling of daily minimum and maximum temperatures. *Phys. Rev. E*, 69:021110, 2004. doi: 10.1103/PhysRevE.69.021110.

Pelletier, J. D. Analysis and Modeling of the Natural Variability of Climate. *J. Climate*, 10:1331 – 1342, 1997. doi: doi:http://dx.doi.org/10.1175/1520-0442(1997)010<1331:AAMOTN>2.0.CO;2.

Pelletier, J. D. and D. L. Turcotte. Self-affine time series: II. Applications and Models. In Dmowska, R. and B. Saltzman, editors, *Long-Range Persistence in Geophysical Time Series*, volume 40 of *Advances in Geophysics*, pages 91 – 166. Elsevier, 1999. doi: http://dx.doi.org/10.1016/S0065-2687(08)60294-0.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, et al. Mosaic organization of DNA nucleotides. *Phys. Rev. E.*, 49:1685–1689, 1994. doi: 10.1103/PhysRevE.49.1685.

Percival, D. B., J. E. Overland, and H. O. Mofjeld. Interpretation of North Pacific variability as a short-and long-memory process*. *J. Climate*, 14(24):4545–4559, 2001.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, 108(D14), 2003.

Rea, W., M. Reale, and J. Brown. Long memory in temperature reconstructions. *Climatic Change*, 107(3-4):247–265, 2011. doi: 10.1007/s10584-011-0068-y.

Rybski, D. and A. Bunde. On the detection of trends in long-term correlated records. *Physica A*, 388(8):1687 – 1695, 2009. doi: http://dx.doi.org/10.1016/j.physa.2008.12.026.

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

Rybski, D., A. Bunde, and H. von Storch. Long-term memory in 1000-year simulated temperature records. *J. Geophys. Res.*, 113, 2008. doi: 10.1029/2007JD008568.

Rypdal, K. Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions. *J. Geophys. Res.*, 117(D6), 2012.

Rypdal, M. and K. Rypdal. Long-memory effects in linear-response models of Earth's temperature and implications for future global warming. Submitted to J. Climate, 2013.

Stanley, H., S. Buldyrev, A. Goldberger, J. Hausdorff, S. Havlin, et al. Fractal landscapes in biological systems: Long-range correlations in DNA and interbeat heart intervals. *Physica A*, 191(1–4):1 – 12, 1992. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/0378-4371(92)90497-E.

Syroka, J. and R. Toumi. Scaling and persistence in observed and modeled surface temperature. *Geophys. Res. Lett.*, 28(17):3255–3258, 2001. ISSN 1944-8007. doi: 10.1029/2000GL012273.

Talkner, P. and R. O. Weber. Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Phys. Rev. E*, 62:150–160, 2000. doi: 10.1103/PhysRevE.62.150.

Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, et al. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012, 2005. ISSN 1477-870X. doi: 10.1256/qj.04.176.

Vandewalle, N. and M. Ausloos. Coherent and random sequences in financial fluctuations. *Physica A*, 246(3–4):454 – 459, 1997. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/S0378-4371(97)00366-X.

Vjushin, D., R. B. Govindan, S. Brenner, A. Bunde, S. Havlin, et al. Lack of scaling in global climate models. *J. Phys.-Condens. Mat.*, 14:2275–2282, 2002.

Voss, R. F. Characterization and measurement of random fractals. *Phys. Scr.*, 1986(T13):27, 1986.

Vyushin, D., I. Zhidkov, S. Havlin, A. Bunde, and S. Brenner. Volcanic forcing improves Atmosphere-Ocean Coupled General Circulation Model scaling performance. *Geophys. Res. Lett.*, 31, 2004. doi: 10.1029/2004GL019499.

Vyushin, D. I., P. J. Kushner, and F. Zwiers. Modeling and understanding persistence of climate variability. *J. Geophys. Res.*, 117(D21), 2012. ISSN 2156-2202. doi: 10.1029/2012JD018240.

Vyushin, D. I. and P. J. Kushner. Power-law and long-memory characteristics of the atmospheric general circulation. *J. Climate*, 22(11), 2009.

Weber, R. O. and P. Talkner. Spectra and correlations of climate data from days to decades. *J. Geophys. Res.*, 106:20,131–20,144, 2001. doi: 10.1029/2001JD000548.

Weron, R. Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, 312(1–2):285 – 299, 2002. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/S0378-4371(02)00961-5.

Zhu, X., K. Fraedrich, Z. Liu, and R. Blender. A demonstration of long-term memory and climate predictability. *J. Climate*, 23(18):5021–5029, 2010.

Zorita, E., T. Stocker, and H. von Storch. How unusual is the recent series of warm years? *Geophys. Res. Lett.*, 35(24), 2008.

Paper I

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

Accepted for publication in *J. Geophys. Res. Atmos.*, Volume 118, 16 July 2013.

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 ± 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 ϕ. 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 β 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 β 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 τ_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 τ_c and of a white noise on scales longer than τ_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 β 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 μ and σ^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 nth 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 nth 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

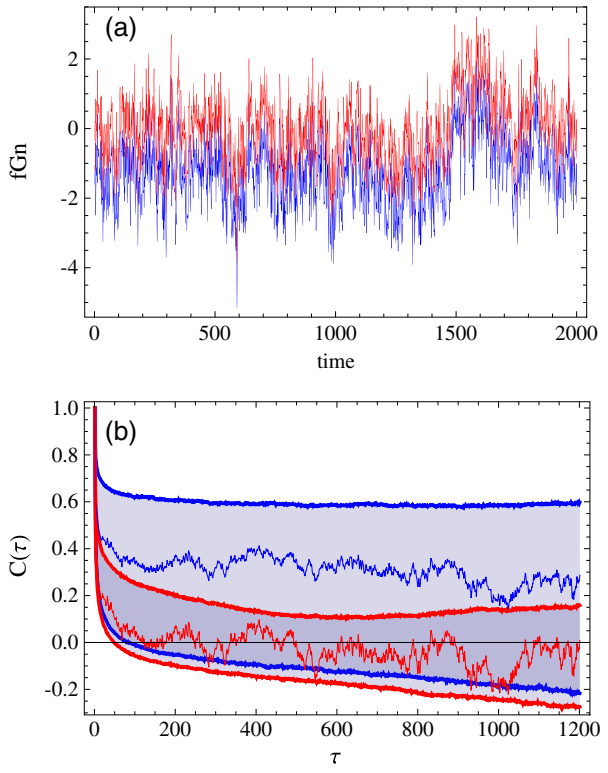


Figure 1. (a) Blue: a numerical realization of an fGn with $\beta = 0.8$ ($H = 0.9$). Red: the same signal, but with zero sample mean. (b) The blue, irregular curve is the unbiased ACF estimate from the blue signal in Figure 1a. The red, irregular curve is the biased ACF estimate from the red signal 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 τ 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 τ and enumerated by the index $\nu = 1, \ldots, N_\tau$. In each segment an nth 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 DFAn. 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 τ due to the bias [*Lennartz and Bunde*, 2009b]. This negative bias is more pronounced for H approaching unity. At any given τ 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

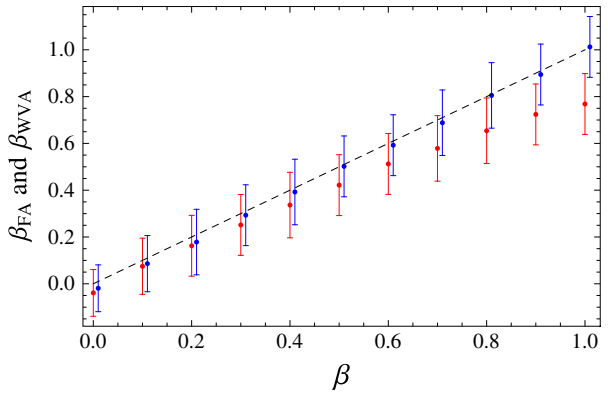


Figure 2. β_{FA} (red) and β_{WVA} (blue) plotted against β. Every point has been computed from an ensemble of numerically generated fGn records of only 2000 data points. The error bars are 95% confidence intervals and are approximately ± 0.14 for both estimation methods.

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 β_{WVA} is negligible, the bias for β_{FA} when β approaches 1 is close to –0.2. The $\pm 2\sigma$ error for both estimates over the entire β-interval is approximately ± 0.14. Since $H = (\beta + 1)/2$, the corresponding figures for H is ± 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 β 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 τ_+ is reduced for stronger trend. The curvature for $\tau \leq \tau_- \approx 3$ is inherent in the wavelet method. Hence,

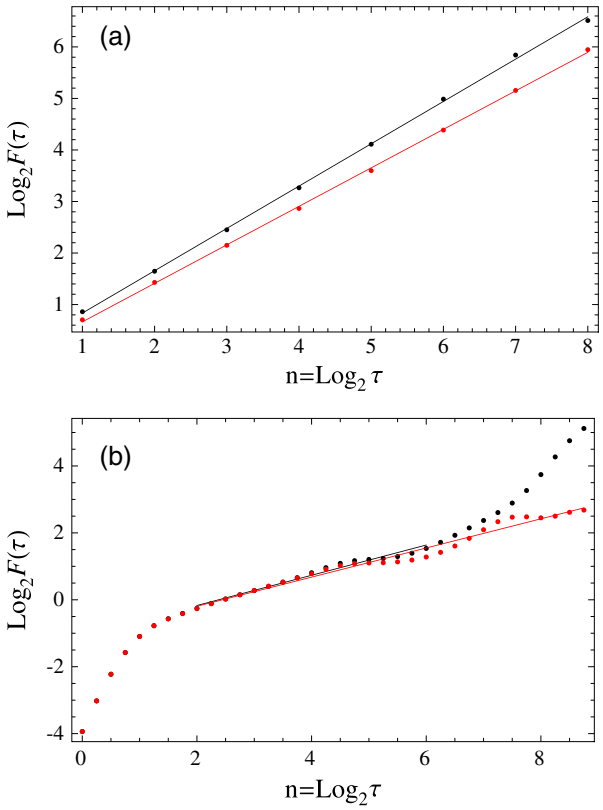


Figure 3. (a) FA of synthetic fGn record with $\beta = 0.5$ ($H = 0.75$), $\sigma = 1$ and length 2000 data points (red) and FA of the same record with the linear trend with slope 0.001 (black). Over the entire record, the trend implies an increase of 2σ. The slope of the red curve is $H = 0.75$ and 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 τ 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.

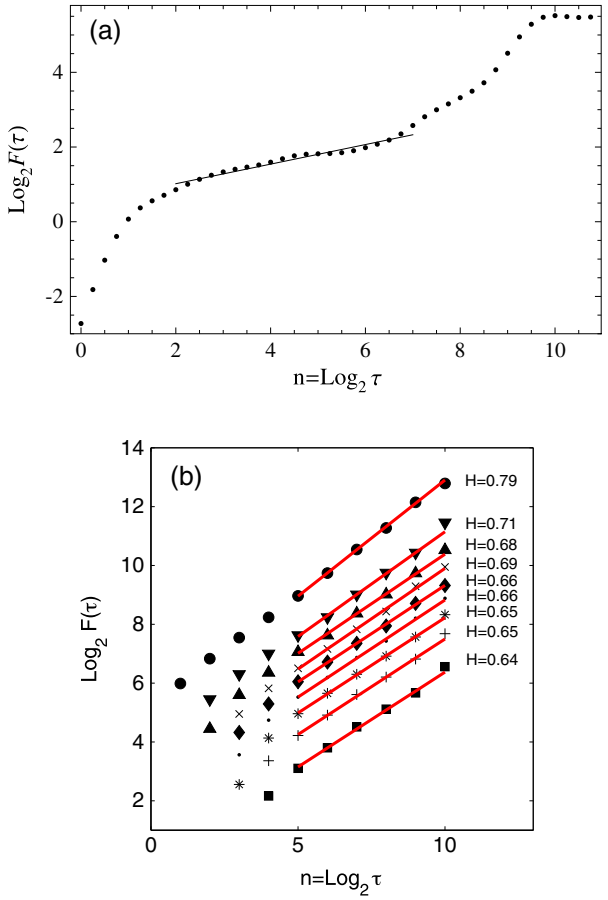


Figure 4. (a) WVA of the CET record. The slope of the black line is $\beta = 0.26$, corresponding to $H = 0.63$. (b) DFA0–8 of CET. The upper curve is the fluctuation function for DFA0, the ones below are DFAn, $n = 1, \ldots 8$, with DFA$n+1$ coming as the curve right below DFAn.

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 β 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 τ. 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 qth 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 τ, 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 τ_+ 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

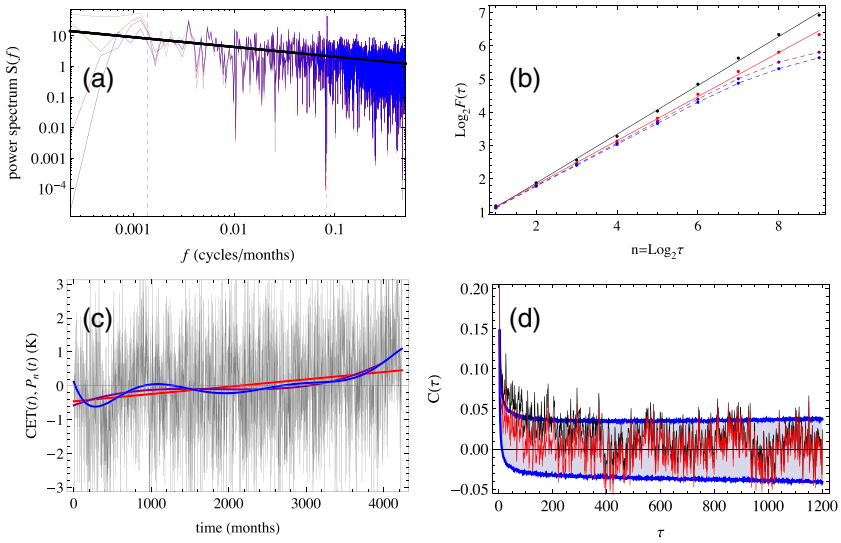


Figure 5. (a) PSDs of deseasonalized monthly CET record 1659–2011 A.D. with variable degree of detrending. Gray: undetrended. Red: P_1 detrended. Purple: P_3 detrended. Blue: P_7 detrended. Thick line has slope $-\beta = -0.30$, corresponding to $H = 0.65$. Vertical dashed lines mark the 60 year period (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 τ_+ 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 τ_+. 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 ± 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 β, 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

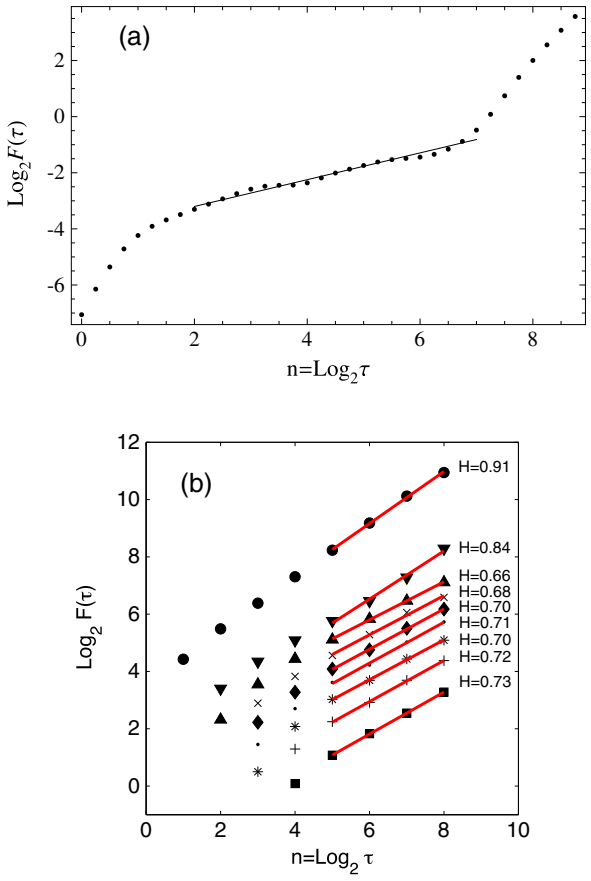


Figure 6. (a) WVA of the GMLT record. The slope of the black line is $\beta = 0.48$, corresponding to $H = 0.74$. (b) DFA0–8 of GMLT. The upper curve is the fluctuation function for DFA0, the ones below are DFAn, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFAn.

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 γ. 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 (σ 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 τ, and the reduction is larger for larger τ. 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 τ is $F(\tau)/\tau$. Hence, if the data set exhibits scaling with Hurst exponent H, the fluctuations of the moving average scales as τ^{H-1}. It may not come as a surprise that the fluctuation level of the CET data set decreases faster with increasing τ 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 τ 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 ± 0.07. This should yield $H = 0.80 \pm 0.07$. H for WVA has negligible bias and error ± 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) ± 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

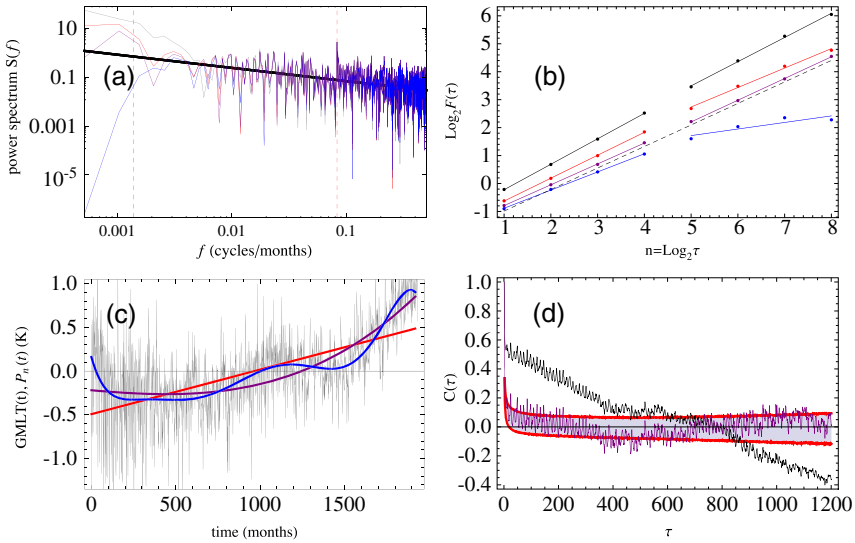


Figure 7. (a) PSD of monthly GMLT record 1850–2010 A.D. Gray: undetrended. Red: P_1 detrended. Purple: P_3 detrended. Blue: P_7 detrended. Thick line has slope $-\beta = -0.54$, corresponding to $H = 0.77$. Vertical dashed lines mark the 60 year period (blue), and the 1 year period (red). (b) FA of the 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 ± 0.07 and for DFA3 is ± 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 ± 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 τ. 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

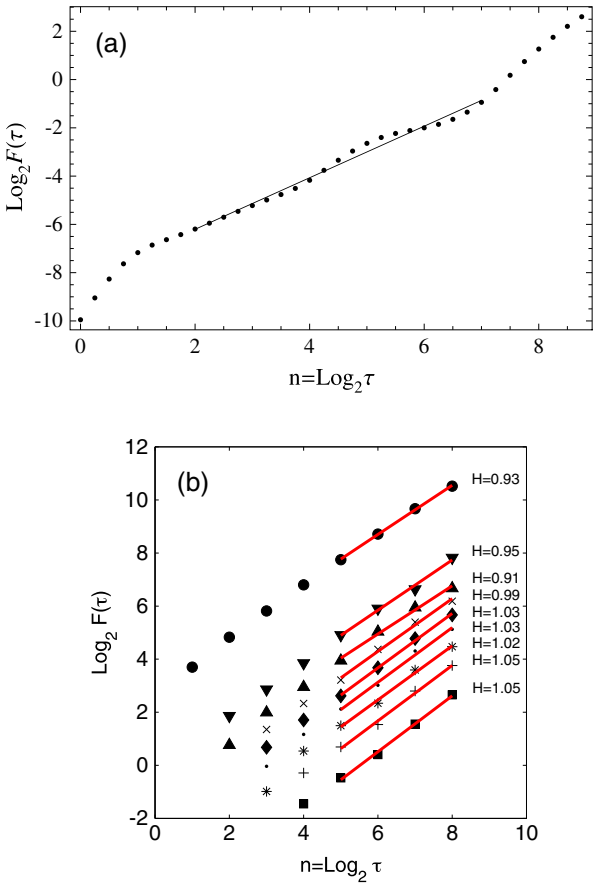


Figure 8. (a) WVA of the GMST record. The slope of the black line is $\beta = 1.07$, corresponding to $H = 1.03$. (b) DFA0–8 of GMST. The upper curve is the fluctuation function for DFA0, the ones below are DFAn, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFAn.

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 τ) 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 τ. 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

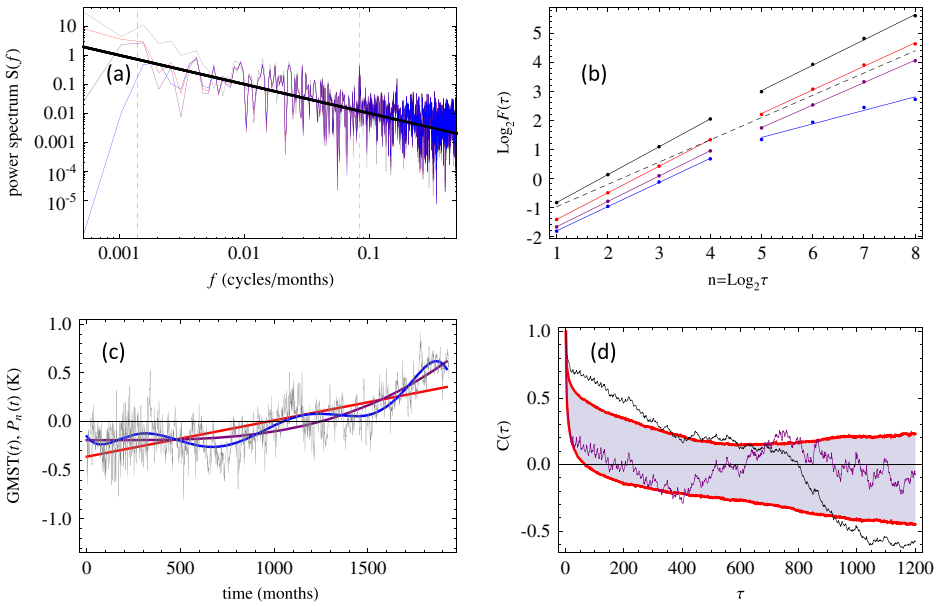


Figure 9. (a) PSD of monthly GMST record 1850–2010 A.D. Gray: undetrended. Red: P_1 detrended. Purple: P_3 detrended. Blue: P_7 detrended. Thick line has slope $-\beta = -1.0$, corresponding to $H = 1.0$. Vertical dashed lines mark the 60 year period (blue), and the 1 year period (red). (b) FA of the 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 τ 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 τ 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

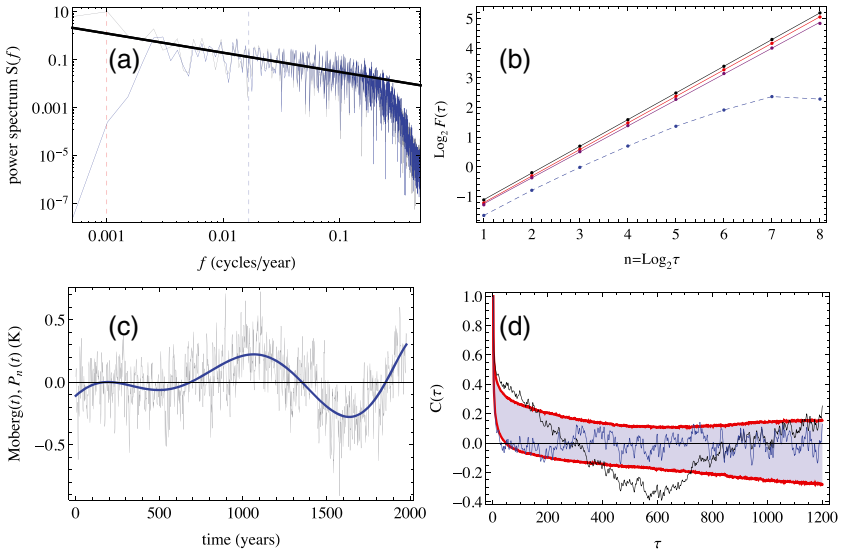


Figure 10. (a) PSD of undetrended Moberg record 0–1979 A.D. (gray) and of P_7-detrended record (blue). Thick line has slope $-\beta = -0.80$, corresponding to $H = 0.90$. Vertical dashed lines mark the 60 year period (blue), and the 1000 year period (red). (b) Variogram of the "profile" $y(t)$ of the 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 $H = 0.90$.

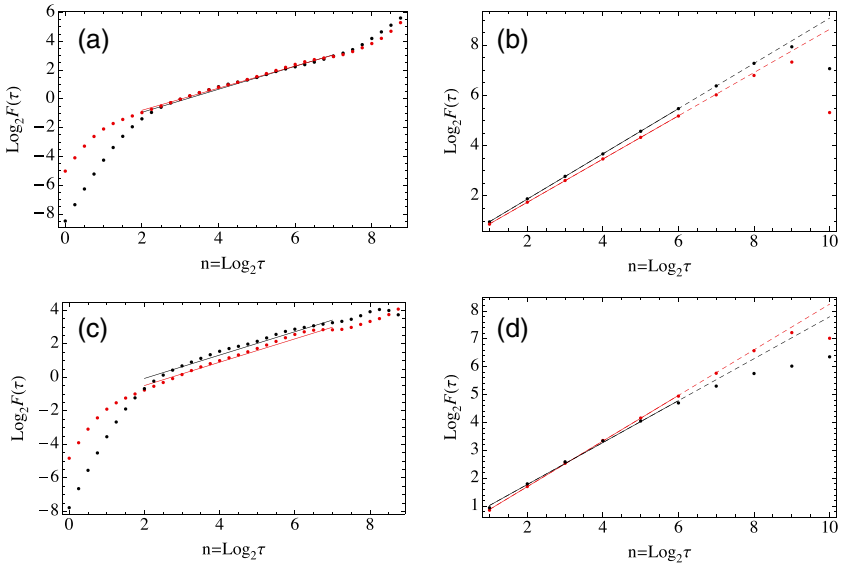


Figure 11. (a) WVA of undetrended Moberg record (black) and of synthetic record consisting of wavelet-filtered signal plus fGn with $H = 0.87$ (red). The slope of the black line is $\beta = 0.80$ ($H = 0.90$) and of the red line $\beta = 0.76$ ($H = 0.88$). (b) FA of the same signals as in Figure 11a. 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 τ.

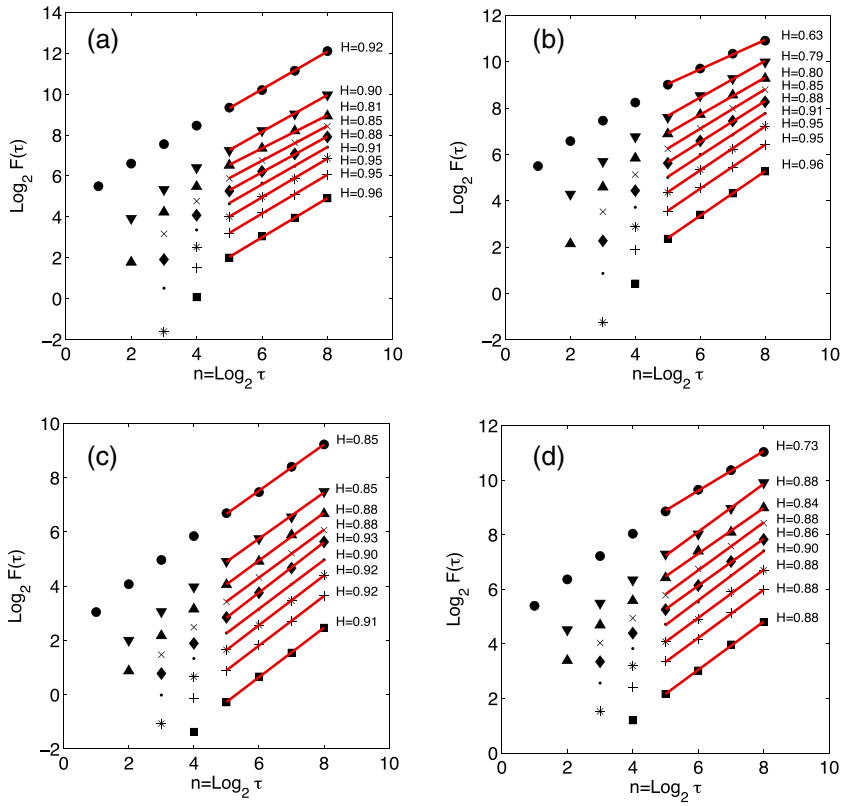


Figure 12. (a) DFA0–8 of undetrended Moberg record. The upper curve is the fluctuation function for DFA0, the ones below are DFAn, $n = 1, \ldots 8$, with DFA$n + 1$ coming as the curve right below DFAn. (b) DFA0–8 of Moberg record detrended by subtraction of wavelet-filtered signal. (c) DFA0–8 of synthetic 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 τ. 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 τ 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 β for FA is now –0.30 when β approaches unity, and the error for both FA and DFA is approximately ± 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 ± 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%

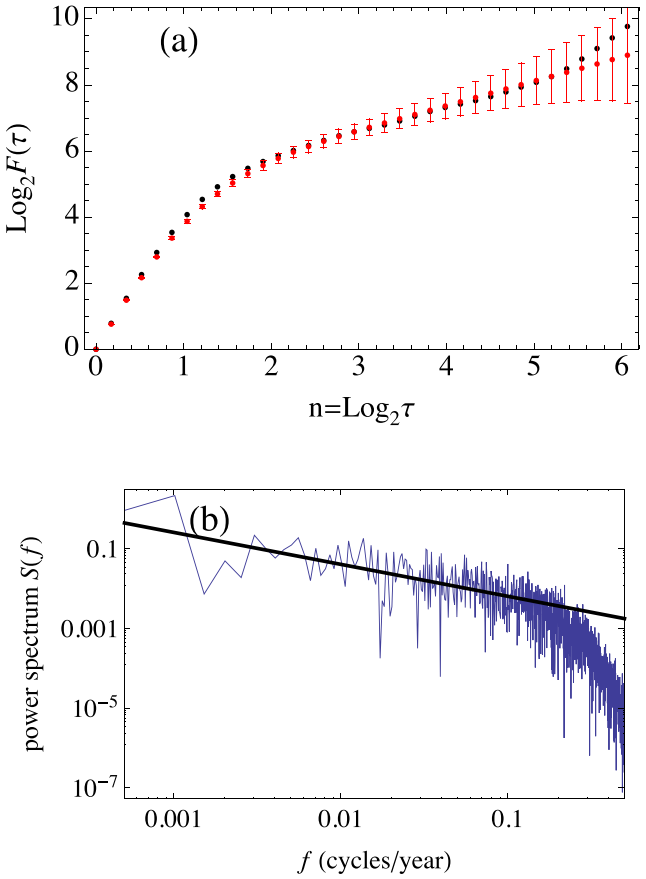


Figure 13. (a) Black dots: WVA fluctuation function for the Moberg record. Red dots: Mean WVA fluctuation function for an ensemble of filtered fGns with $\beta = 0.75$ ($H = 0.875$). The error bars indicate the 95% confidence intervals estimated from the ensemble. (b) Periodogram for one realization 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 β, which is the β estimated from the WVA. But as our analysis shows, this estimate has an uncertainty of ± 0.2, so we cannot be certain that we have used the right β 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 β 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

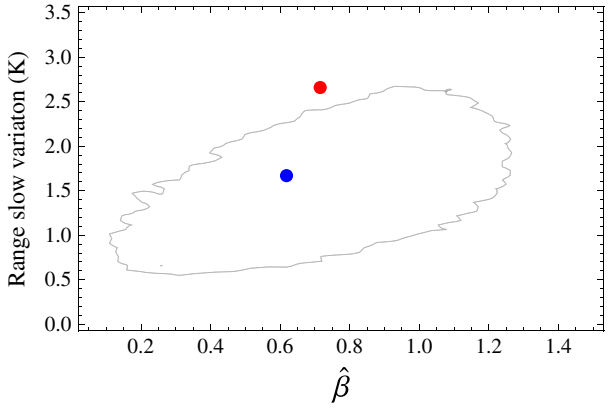


Figure 14. The closed contour indicates the line of constant $p(\hat{r}, \hat{\beta})$ inside which the integrated probability is 0.95. The red dot is the estimated $(\hat{r}_{\rm obs}, \hat{\beta}_{\rm obs})$ for the Moberg record. The blue dot is the same estimate for the modified Moberg record where 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 (± 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 ϕ 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.

Paper II

Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years

Submitted to *Earth System Dynamics Discussions*.

Manuscript prepared for J. Name
with version 5.0 of the LaTeX class copernicus.cls.
Date: 20 February 2014

Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years

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

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Norway
[2]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway

Correspondence to: Lene Østvand
(lene.ostvand@uit.no)

Abstract. Various interpretations of the notion of a trend in the context of global warming are discussed, contrasting the difference between viewing a trend as the deterministic response to an external forcing and viewing it as a slow variation which can be separated from the background spectral continuum of long-range persistent climate noise. The emphasis in this paper is on the latter notion, and a general scheme is presented for testing a multi-parameter trend model against a null hypothesis which models the observed climate record as an autocorrelated noise. The scheme is employed to the instrumental global sea-surface temperature record and the global land temperature record. A trend model comprising a linear plus an oscillatory trend with period of approximately 70 yr, and the statistical significance of the trends, are tested against three different null models: first-order autoregressive process, fractional Gaussian noise, and fractional Brownian motion. The parameters of the null models are estimated from the instrumental record, but are also checked to be consistent with a Northern Hemisphere temperature reconstruction prior to 1750 for which an anthropogenic trend is negligible. The linear trend in the period 1850-2010 AD is significant in all cases, but the oscillatory trend is insignificant for ocean data and barely significant for land data. However, by using the significance of the linear trend to constrain the null hypothesis, the oscillatory trend in the land record appears to be statistically significant. The results suggest that the global land record may be better suited for detection of the global warming signal than the ocean record.

1 Introduction

At the surface of things, the conceptually simplest approach to detection of anthropogenic global warming should be the estimation of trends in global surface temperature throughout the instrumental observation era starting in the mid-nineteenth century. These kinds of estimates, however, are subject to deep controversy and confusion originating from disagreement about how the notion of a trend should be understood. In this paper we adopt the view that there are several, equally valid, trend definitions. Which one that will prove most useful depends on the purpose of the analysis and the availability and quality of observation data.

At the core of the global change debate is how to distinguish anthropogenically forced warming from natural variability. A complicating factor is that natural variability has forced as well as internal components. Power spectra of climatic time series also suggest to separate internal dynamics into quasi-coherent oscillatory modes and a continuous and essentially scale-invariant spectral background. Over a vast range of time scales this background takes the form of a persistent, fractional noise or motion (*Lovejoy and Schertzer*, 2013; *Markonis and Koutsoyannis*, 2013). Hence, the issue is threefold: (i) to distinguish the climate response to anthropogenic forcing from the response to natural forcing, (ii) to distinguish internal dynamics from forced responses, and (iii) to distinguish quasi-coherent, oscillatory modes from the persistent noise background. This conceptual structure is illustrated by the Venn diagram in Figure 1a. Figure 1b illustrates three possible trend notions based on this picture. Fundamental for all is the separation of the observed climate record into a trend component (also termed the *signal*) and a *climate noise* component. The essential difference between these notions is how to make this separation.

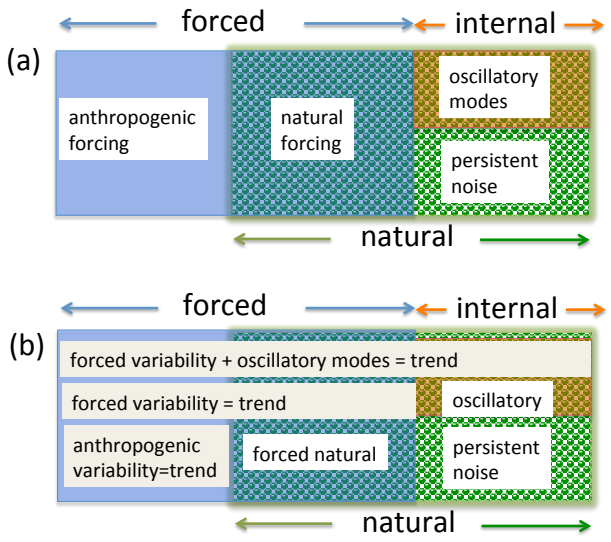


Fig. 1. Venn diagrams illustrating the interplay between forced, internal, and natural variability and various definitions of trend. (a): Natural variability can be both forced and internal. Forced variability can be both anthropogenic and natural. Internal variability is natural, but can consist of quasiperiodic oscillatory modes as well as a continuum of persistent noise. (b): The three different trend notions discussed in the text.

The widest definition of the trend is to associate it with all forced variability and oscillatory modes as illustrated by the upper row in Figure 1b. With this notion the methodological challenge will be to develop a systematic approach to extract the trend from the observed record, and then to subtract this component to establish the persistent noise component. The physical relevance of this separation will depend on to what extent we can justify to interpret the extracted trend as a forced response with internally generated oscillatory modes superposed. If detailed information on the time evolution of the climate forcing is not used or is unavailable such a justification is quite difficult. In this case we will first construct a parametrized model for the trend based on the appearance of the climate record at hand and our physical insight about the forcing and the nature of the dynamics. The next step will be to estimate the parameters of the trend model by conventional regression analysis utilizing the observed climate record. The justification of interpreting this trend as something forced and/or coherent different from background noise will be done through a test of the null hypothesis which states that the climate record can be modeled as a long-range memory (LRM) stochastic process. Examples of such processes are persistent fractional Gaussian noises (fGns) or fractional Brownian motions (fBms). LRM processes exhibit stronger random fluctuations on long time scales than short-memory processes and hence a null model based on LRM-noise will make it more difficult to reject the null hypothesis for a given estimated trend. For comparison we will also test the null hypothesis against a conventional short-memory notion of climate noise, the first-order autoregressive process (AR(1)). In general, rejection of the null hypothesis will be taken as an acceptance of the hypothesis that the estimated trend is significant, and will strengthen our confidence that these trends represent identifiable dynamical features of the climate system.

A trend can be rendered significant under the AR(1) null hypothesis, but insignificant under an LRM-hypothesis, and then it could of course be argued that the value of this kind of analysis of statistical significance is of little interest, unless one can establish evidence that favors one null model over another. One can, however, test the null models against the observation data, and here analysis seems to favor the fGn/fBm models over short-memory models. There are dozens of papers that demonstrate scaling properties consistent with fGn or fBm properties in instrumental temperature data (see *Rypdal et al.* (2013) for a short review and some references). But, since the instrumental records may be strongly influenced by the increasing trend in anthropogenic forcing, it is difficult to disentangle LRM introduced by the forcing from that arising from internal, unforced variability. Detrending methods such as the detrended fluctuation analysis (*Kantelhardt et al.*, 2001) are supposed to do this, but the short duration of the instrumental records does not seem to allow us to make an undisputable distinction between AR(1) and fGn/fBm. We analyze this issue in section 3.3, where we also comment the methods and conclusions in a recent study by (*Vyushin et al.*, 2012).

There are also other approaches that favor the LRM models for description of random internal variability in global data on time scales from months to centuries. One is based on analysis of temperature reconstructions for the last millennium prior to the anthropocene (*Rybski et al.*, 2006; *Rypdal et al.*, 2013). These temperature data are not influenced by an anthropogenic trend, but exhibit self-similar scaling properties with spectral exponent $\beta \approx 1$ (to be explained in section 2) on time scales at least up to a century. Short-memory processes like the AR(1) will typically exhibit scaling with $\beta \sim 2$ up to the autocorrelation time, and a flat ($\beta \sim 0$) spectrum on time scales longer than this, but this is not observed in these data. Another line of investigation has been to use available time-series information about climate forcing in a parametrized, linear, dynamic-stochastic model for the climate response (*Rypdal and Rypdal*, 2013). The trend then corresponds to the deterministic solution to this model, i.e., the solution with the known (deterministic) component of the forcing. In this model the persistent noise component of the temperature record is the response to a white noise stochastic forcing. In (*Rypdal and Rypdal*, 2013) analysis of the residual obtained by subtracting the deterministic forced solution from the observed instrumental global temperature record

shows scaling properties consistent with an fGn model and inconsistent with an AR(1) model.

The method is described in *Rypdal and Rypdal* (2013), where only exponential and scale-free long-range persistent responses are modeled, without allowing for quasi-coherent oscillations. The approach in that paper adopts the trend definition described in the second row of Figure 1b. Here the trend is the forced variability, while all unforced variability is relegated to the realm of climate noise. It is possible, however, to incorporate forced and natural oscillatory dynamics into such a response model.

The lower row in Figure 1b depicts the trend notion of foremost societal relevance; the forced response to anthropogenic forcing. Once we have estimated the parameters of the forced response model, we can also compute the deterministic response to the anthropogenic forcing separately. One of the greatest advantages of the forced-response methodology is that it allows estimation of this anthropogenic trend/response and prediction of future trends under given forcing scenarios, subject to rigorous estimates of uncertainty. On the other hand, that method is based on the assumption that the forcing data employed are correct. The construction of forcing time series relies heavily on uncertain observations and modeling, hence there is an obvious case for complementary approaches to trend estimation that do not rely on this kind of information. This is the approach that will be explored in the present paper.

2 Trend Detection Methodology

2.1 The null models

The noise modeling in this paper makes use of the concept of long-range memory (LRM), or (equivalently) long-term persistence (LTP) (*Beran*, 1994). In global temperature records this has been studied in e.g., *Pelletier and Turcotte* (1999); *Lennartz and Bunde* (2009); *Rybski et al.* (2006); *Rypdal and Rypdal* (2010); *Efstathiou et al.* (2011); *Rypdal et al.* (2013); *Rypdal and Rypdal* (2013). Emanating from these studies is the recognition that ocean temperature is more persistent than land temperature and that the 20'th century rising trend is stronger for land than for ocean. LRM is characterized by a time-asymptotic ($t \to \infty$) autocorrelation function (ACF) of power-law form $C(t) \sim t^{\beta-1}$ for which the integral $\int_0^\infty C(t)\mathrm{d}t$ diverges. Here β is a power-law exponent indicating the degree of persistence. The corresponding asymptotic ($f \to 0$) power spectral density (PSD) has the form $S(f) \sim f^{-\beta}$, hence β is also called the spectral index of the LRM process. For $0 < \beta < 1$ the process is stationary and is termed a persistent fGn. For $1 < \beta < 3$ the process is non-stationary and termed an fBm. As a short-memory alternative we shall also consider the AR(1) process which has an exponentially decaying ACF and is completely characterized by the lag-one autocorrelation ϕ (*von Storch and Zwiers*, 1999).

2.2 Previous work using LRM null models

Bloomfield and Nychka (1992) studied the signficance of a linear trend in 128 years of global temperature assuming different stochastic models, including fractionally integrated white noise. They found that the trend in the record could not be explained as natural variability by any of the models.

Significance of linear trends under various null models, some exhibiting LRM, was also studied by *Cohn and Lins* (2005). One of their main points was that trends classified as statistically significant under a short-memory null hypothesis might end up as insignificant under an LRM hypothesis. The paper is a theoretical study of trend significance and is motivated by the strong persistence which is known to exist in hydroclimatic records. As an example they study the Northern Hemisphere (NH) temperature record and find that their test renders the trend insignificant under the LRM null hypothesis. They conclude that the trend *might* be due to natural dynamics. Analyses with similar and other methodologies on other records indicate that the global trend signal is significant in spite of LRM (*Gil-Alana*, 2005; *Rybski et al.*, 2006; *Lennartz and Bunde*, 2009; *Halley and Kugiumtzis*, 2011; *Rypdal et al.*, 2013). We show in the present paper that the global land temperature record turns out to exhibit a stronger trend and weaker LRM than the NH temperature which is sufficient to establish trend significance. In contrast, the weaker trend and stronger LRM of global ocean temperature yield a less significant trend for this signal.

Some recent papers on LRM and trends are *Fatichi et al.* (2009); *Rybski et al.* (2009); *Franzke* (2009, 2010); *Franzke and Woollings* (2011); *Franzke* (2012a,b); *Franzke et al.* (2012). *Fatichi et al.* (2009) and *Rybski et al.* (2009) study station temperatures under different LRM null hypotheses, and find significant linear trends in some, but not all, of the records. *Franzke* (2012b) applies a methodology similar to that of *Cohn and Lins* (2005) to single-station temperature records in the Arctic Eurasian region. He emphasises that almost all stations show a positive trend, and that the melting of Arctic sea ice leaves no doubt about the reality of an anthropogenic warming signal in the Arctic. By evaluating all station data together, for instance by analysing the regional averaged temperature, one would most likely arrive at a significant trend. His point is that the natural variability for single stations is so large and long-range correlated that it may mask the warming signal at the majority of individual stations at the present stage of global warming. This is an important message to convey to those policymakers who have got the impression that local climate projections universally are sufficiently reliable to implement adaptive measures.

2.3 Hypothesis testing methodology

In the present paper our main objective is to establish beyond doubt the significance of the global warming signal, and if possible also the multidecadal oscillation. From the studies discussed above, we know that there are many temperature records from which this significance cannot be established under an LRM null hypothesis, so we should search for a signal that is optimal for trend detection. Such an optimal signal seems to be the instrumental global land temperature record HadCRUT3 (*Jones et al.*, 2012). We will contrast this with analysis of the global ocean record (*Kennedy et al.*, 2011). These records are land-air and sea-surface temperature anomalies relative to the period 1961-90, with monthly resolution from 1850 to date. The analysis is made using a trend model which contains a linear plus a sinusoidal trend, although the methodology developed works for any parametrized trend model. We test this model against the null model that the full temperature record is a realization of an AR(1) process, an fGn, or an fBm (the fBm model is of interest only for the strongly persistent ocean data).

The significance tests are based on generation of an ensemble of synthetic realizations of the null models; AR(1) processes ($\phi < 1$), fGns ($0 < \beta < 1$), and fBms ($1 < \beta < 3$). Each realization is fully characterized by a pair of parameters; $\theta \equiv (\sigma, \phi)$ for AR(1) and $\theta \equiv (\sigma, \beta)$ for fGn and fBm, where σ is the standard deviation of the stationary AR(1) and fGn processes and the standard deviation of the differenced fBm. For an LRM null model the estimated value of $\hat{\beta}$ depends on which null model (fGn or fBm) one adopts. As we will show below, for ocean data, it is not so clear whether an fGn or an fBm is the most proper model (*Lennartz and Bunde*, 2009; *Rypdal et al.*, 2013), so we will test the significance of the trends under both hypotheses.

Technically, we make use of the R package by *McLeod et al.* (2007) to generate synthetic fGns and to perform a maximum-likelihood estimation of β. Since generation of fBms is not included in this package, synthetic fBms with memory exponent $1 < \beta < 3$ are produced by generating an fGn with exponent $\beta - 2$ and then forming the cumulative sum of that process. This is justified because the one-step differenced fBm with $1 < \beta < 3$ is an fGn with memory exponent $\beta - 2$ (*Beran*, 1994). Maximum-likelihood estimation of β for synthetic fBms and observed data records modeled as an fBm is done by forming the one-time-step increment (differentiation) process, estimate the memory exponent β_{incr} for that process and find $\beta = \beta_{\text{incr}} + 2$. There are some problems with this method when $\beta \approx 1$. Suppose we have a data record (like the global ocean record) and we don't know whether $\beta < 1$ or $\beta > 1$. For all estimation methods there are large errors and biases for short data records of fGns/fBms for $\beta \approx 1$ (*Rypdal et al.*, 2013). This means that there is an ambiguity as to whether a record is a realization of an fGn or an fBm when we obtain estimates of β in the vicinity of 1. For the MLE method this ambiguity becomes apparent from Figure 2. Here we have plotted the MLE estimate $\hat{\beta}$ with error bars for an ensemble of realizations of fGns (for $0 < \beta < 1$) and of fBms ($1 < \beta < 2$) with 2000 data points. The red symbols are obtained by adopting an fGn model when β is estimated. Hence, for $\beta > 1$ we find the estimate $\hat{\beta}$ from a realization of an fBm with a model that assumes that it is an fGn. It would be expected that the analysis would give $\hat{\beta} \approx 1$ for an fBm, but we observe that it gives $\hat{\beta}$ considerably less than 1 in the range $1 < \beta < 1.4$, so if we observe a $\hat{\beta}$ in the vicinity of 1 by this analysis we cannot know whether it is an fGn or an fBm. The ambiguity remains by estimating with a model that assumes that the record is an fBm, because this yields a corresponding positive bias as shown by the green symbols when the record is an fGn. This ambiguity seems difficult to resolve for ocean data as short as the monthly instrumental record.

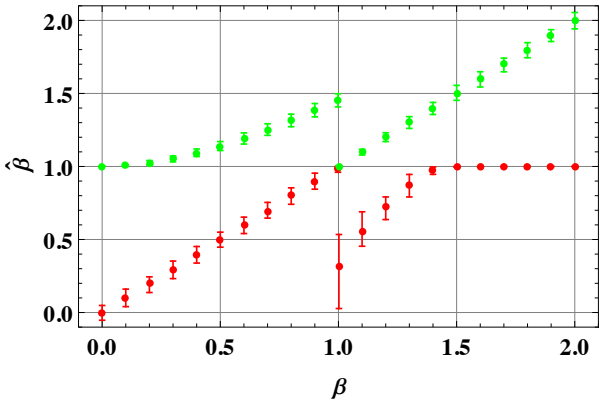


Fig. 2. The red symbols and 95% confidence intervals represent the maximum-likelihood estimate $\hat{\beta}$ for realizations of fGns/fBms with memory parameter β by adopting an fGn model. Hence, for $\beta > 1$ we find the estimate $\hat{\beta}$ from a realization of an fBm with a model that assumes that it is an fGn. The green symbols represent the corresponding estimate by adopting an fBm model, i.e., for $\beta < 1$ we we find the estimate $\hat{\beta}$ from a realization of an fGn with a model that assumes that it is an fBm. "Adopting an fBm model" means that the synthetic record is differentiated, then analyzed as an fGn by the methods of *McLeod et al.* (2007) to obtain $\hat{\beta}_{\text{incr}}$, and then finally $\beta = \hat{\beta}_{\text{incr}} + 2$.

The standard method for establishing a trend in time-series data is to adopt a parametrized model $T(A;t)$ for the trend, e.g., a linear model $A_1 + A_2 t$ with parameters $A = (A_1, A_2)$, and estimate the model parameters by a least-square fit of the model to the data. Another method, which brings along additional meaning to the trend concept, is the MLE method. This method adopts a model for the stochastic process; $x(t) = T(A;t) + \sigma w(t)$, where $w(t)$ is a correlated or uncorrelated random process and establishes the set of model parameters A for which the likelihood of the stochastic model to produce

the observed data attains its maximum. The method applied to uncorrelated and Gaussian noise models is described in many standard statistics texts (*von Storch and Zwiers*, 1999), and its application to fGns is described in *McLeod et al.* (2007). If $w(t)$ is a Gaussian, independent and identically distributed (i.i.d.) random process, the MLE is equivalent to the least-square fit. If $w(t)$ is a strongly correlated (e.g., LRM) process, and the trend model provides a poor description of the large-scale structures in the data, MLE may assign more weight to the random process (greater σ) than the least-square method. On the other hand, if the trend model is chosen such that it can be fitted to yield a good description of the large-scale structure, the parameters estimated by the two methods are quite similar, even if $w(t)$ used in the MLE method is an LRM process. In this case we can use least-square fit to establish the trend parameters without worrying about whether the residual noise obtained after subtracting the estimated trend can be modeled as a Gaussian, i.i.d. random process.

In the following, we make some definitions and outline the methodology we adopt to assess the significance of the estimated trend. Concepts defined are named with boldface fonts. Our methodology is based on standard hypothesis testing, where the trend hypothesis (termed the "alternative hypothesis") is accepted (although not verified, which is stronger) by rejection of a "null hypothesis." Failure of rejection of the null hypothesis implies failure of acceptance of the alternative hypothesis, and hence the trend will be characterized as insignificant under this null hypothesis. Hence, it is clear that the outcome of the significance test will depend on the choice of alternative hypothesis (trend model) as well as on the null hypothesis (noise model).

Let us take the pragmatic point of view that a trend is a simple and slowly varying (compared with a predefined time scale τ) function $T(A;t)$ of t, parametrized by the trend coefficients $A = (A_1, \ldots, A_n)$. It is also required that for the optimal choice of parameters, $A = \hat{A}_{\text{obs}}$ the trend $T(\hat{A}_{\text{obs}};t)$ makes a good fit to the large-scale structure of the data record. In practice, this means that the trend should be close to a low-pass filtered version of the signal, for instance a moving average over time-scale τ. The trend is significant with respect to a particular null model if the fitted $T(\hat{A}_{\text{obs}};t)$ is very unlikely to be realized in an ensemble of fits $T(\hat{A};t)$ to realizations of the null model.

The alternative hypothesis can be formulated as follows: The observed record $x(t)$ is a realization of the stochastic process

$$T(A;t) + \sigma w(t), \tag{1}$$

where the trend $T(A;t)$ is a specified function of t depending on the trend coefficients $A = (A_1, \ldots, A_n)$, and $w(t)$ is a Gaussian stationary random process of unit variance. These coefficients are estimated from a least-square fit to $x(t)$ and have the values $\hat{A}_{\text{obs}}$. We assume that the trend model can be fitted so well to the data that MLE-estimates of A with different noise models (white noise vs. strongly persistent fGn) give approximately the same $\hat{A}_{\text{obs}}$.

The null hypothesis states that the record $x(t)$ is a realization of a stochastic process

$$\varepsilon(\theta;t), \tag{2}$$

with certain properties to be specified (e.g., the process is AR(1), fGn, or fBm). Like for the alternative hypothesis, the parameters θ should be restricted to be close to the values $\hat{\theta}_{\text{obs}}$ found from estimating it from fitting the null model (2) to the data record by means of MLE.

The Monte Carlo null ensemble is the collection of realizations $x_i(\theta)\,, i = 1, 2, \ldots$, of the null model process (2).

The best choice of null model would be to utilize all our possible knowledge about the true parameter set θ. This implies considering θ as a random variable, and hence a Bayesian approach (*Gelman et al.*, 2004). We generate the null ensemble by drawing θ from the conditional distribution $P(\theta|\hat{\theta}_{\text{obs}})$, i.e., the probability that the "real" parameters of the observed process are θ given that the estimated parameters from the observed data are $\hat{\theta}_{\text{obs}}$. One way of establishing this distribution is to generate an ensemble of realizations of the noise process with θ varied in a range around $\theta \approx \hat{\theta}_{\text{obs}}$ and establish the conditional distribution $P(\hat{\theta}|\theta)$. From Bayes' theorem one then has $P(\theta|\hat{\theta}) = P(\hat{\theta}|\theta)P(\theta)/P(\hat{\theta})$. By setting $\hat{\theta} = \hat{\theta}_{\text{obs}}$, and assuming a flat prior distribution $P(\theta)$ in the range in the vicinity of θ_{obs} corresponding to the width of the distribution we the find $P(\theta|\hat{\theta}_{\text{obs}}) = P(\hat{\theta}_{\text{obs}}|\theta)$.

As an alternative to the Bayesian ideas described above one could employ a frequentist approach. This means that we assume that the null model has a fixed true parameter value θ. This parameter value is unknown, and the strategy is to create the Monte Carlo null ensemble $x_i(\hat{\theta}_{\text{obs}})\,, i = 1, 2, \ldots$, using the θ-values estimated from the observed data. We must then take the uncertainty in the θ-estimates into account, since $\hat{\theta}_{\text{obs}}$ may deviate from the true θ. This estimation error can be quantified using the bootstrap method, which assumes that the error in the parameter estimates in the null model with parameters θ can be well approximated by the corresponding errors for the null model with parameters $\hat{\theta}_{\text{obs}}$. When estimation errors are quantified one can easily adjust for these in the hypothesis tests.

Pseudotrend estimates $\hat{A}^{(i)}$ are the coefficients obtained by least-square fit of the trend model $T(A;t)$ to the realizations $x_i(\theta;t)$ of the null ensemble.

Pseudotrend distribution is the n-dimensional PDF $P(\hat{A})$ over the null ensemble.

Null-hypothesis confidence region is the region Ω in n-dimensional A-space for which $P(A) > P_{\text{thr}}$, where P_{thr} is chosen such that $\int_\Omega P(A)\,\mathrm{d}A = 0.95$.

Significance of the trend model is established if the null hypothesis is rejected, e.g., the full n-dimensional trend is 95% significant if $\hat{A}_{\text{obs}} \notin \Omega$.

If the null hypothesis is rejected by this procedure, we are rejecting only those aspects of the null model that are relevant to the full trend model, i.e., the trend model (alternative hypothesis) produces trend coefficients $\hat{A}_{\rm obs}$ that give a good fit to the large-scale structure of the data, while it is very improbable that the null model can produce $\hat{A}$ in the vicinity of $\hat{A}_{\rm obs}$.

2.4 The trend model explored in this work

We will apply the method described in the previous subsection to global temperature record using the following trend model:

$$T(A;t) = \delta + A_1 t + A_2 \sin(2\pi f t + \varphi). \tag{3}$$

This is a simplified version of the models used in several works by N. Scafetta (e.g., *Scafetta*, 2011, 2012) and the oscillation is supposed to model the 60-yr cycle observed in the instrumental record (*Schlesinger and Ramankutty*, 1994). The frequency f is not to be considered as a free model parameter to be estimated from the observed record and from realizations of the null ensemble. When estimating pseudotrends it has little meaning to let f be a free parameter, since the synthetic noise records contain no preferred frequencies. We rather treat f as a fixed quantity which is an inherent part of the alternative hypothesis. In practice we select f from a least-square fit of the trend model to the observed record varying all five parameters including f, but this is not essential. We could just as well have hypothesized a reasonable value of f by inspection of the record or from other evidence of this oscillation presented in the literature. The important thing to keep in mind is that the value of f is part of the hypothesis. Of the estimated pseudotrend coefficients $(\hat{A}_1, \hat{A}_2, \hat{\delta}, \hat{\varphi})$ only $(\hat{A}_1, \hat{A}_2)$ quantify the strength of the trend, so the relevant pseudotrend distribution to establish is $P(\hat{A}_1, \hat{A}_2)$ irrespective of the values of irrelevant parameters $(\hat{\delta}, \hat{\varphi})$. Table 1 shows the estimated $\hat{\theta}_{\rm obs}$ according to the null model in (2) using AR(1), fGn and fBm as the stochastic process $\varepsilon(\theta;t)$. Also in this table are the estimated trend parameters $(\hat{A}_1, \hat{A}_2)$ from applying the trend model in (3) and the period $T = 1/f$ of the oscillatory trend. Since, as mentioned above, this period has been selected from a fitting procedure it has slightly different values for the ocean and land records.

2.5 Results

The results of the analysis are shown in Figure 3. We observe that the trend parameters $(\hat{A}_1, \hat{A}_2)_{\rm obs}$ are outside the null-hypothesis 95% confidence region for all three noise models and for ocean as well as land records. But we also observe that the significance is more evident for land than for ocean, and is reduced as more strongly persistent noise models are employed. For the fBm model applied to ocean data the trend is barely outside the 95% confidence region.

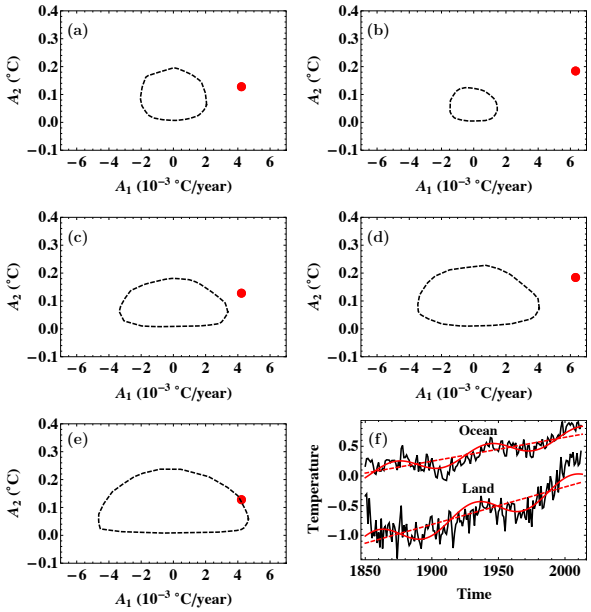


Fig. 3. In panels (a-e) the red dots represent the estimated trend coefficients $(\hat{A}_1, \hat{A}_2)_{\rm obs}$ and the dashed, closed curve the 95% confidence contour of the distribution $P(\hat{A}_1, \hat{A}_2)$. (a): ocean data and AR(1) null model. (b): land data and AR(1) null model. (c): ocean data and fGn null model. (d): land data and fGn null model. (e): ocean data and fBm null model. (f): Black curves: The global ocean and land temperature records. Red curves: the linear and sinusoidal trends.

It is the full trend model equation (3) that is accepted by this test, but something can also be said about the separate significance of the individual trends represented by the individual trend coefficients from the pseudotrend distribution $P(\hat{A}_1, \hat{A}_2)$. For the AR(1) and fGn null models it is apparent from Figure 3a-d that the linear trend is highly significant since $\hat{A}_{1,\rm obs}$ is located far to the right of the confidence region. On the other hand, except for the AR(1) model applied to land data in Figure 3b, $A_{2,\rm obs}$ is not totally above the confidence region. This means that the linear pseudotrends observed in the null ensemble has negligible chance of getting near the observed trend, while there is some chance to find oscillatory trends in the null ensemble which are as large as $\hat{A}_{2,\rm obs}$. The significance of those separate trends against these null models is determined by forming the separate one-dimensional PDFs, $P(\hat{A}_1) \equiv \int P(\hat{A}_1, \hat{A}_2) \mathrm{d}\hat{A}_2$ and $P(\hat{A}_2) \equiv \int P(\hat{A}_1, \hat{A}_2) \mathrm{d}\hat{A}_1$ and form the confidence intervals in the standard way. In Figure 4 we have formed the corresponding one-dimensional cumulative distribution functions (CDFs) from the two-dimensional PDFs for ocean data shown in Figure 3a, c, and e. We observe that the linear trend is significant for the AR(1) and fGn null models, but barely significant for

Table 1. Estimated noise parameters $\hat{\theta}_{\text{obs}}$ from the null hypotheses in (2) and trend parameters $\hat{A}_{\text{obs}}$ estimated from the trend model (3). The units for the trend estimation are months for $\hat{\tau}_{\text{obs}}$, 10^{-3} °C/yr for $\hat{A}_{1,2,\text{obs}}$, and yr for the oscillation period T.

	AR(1)	fGn		fBm		Trend		
	$\hat{\tau}_{\text{obs}}$	$\hat{\beta}_{\text{obs}}$	$\hat{\sigma}_{\text{obs}}$	$\hat{\beta}_{\text{obs}}$	$\hat{\sigma}_{\text{obs}}$	$\hat{A}_{1,\text{obs}}$	$\hat{A}_{2,\text{obs}}$	T
Ocean	21.3	0.994	0.25	1.45	0.086	4.21	0.128	69.7
Land	3.43	0.654	0.49			6.34	0.186	73.4

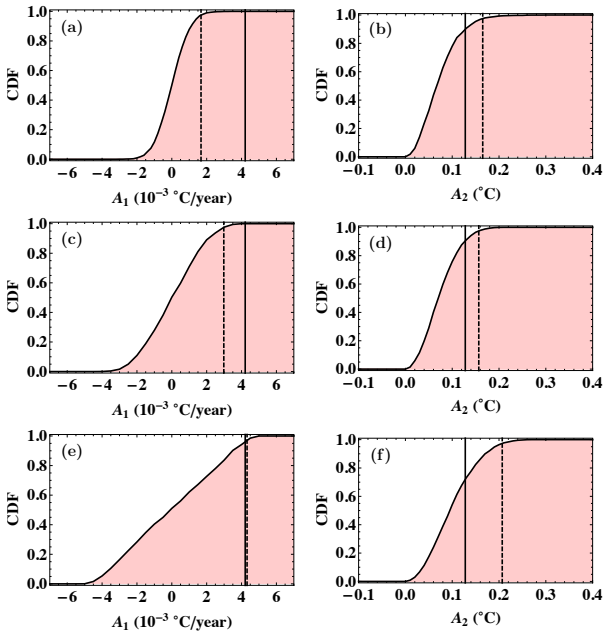


Fig. 4. Curved lines are CDFs for trend coeffecients $\hat{A}_1$ and $\hat{A}_2$ established from the null model ensemble for ocean data. Vertical dashed line marks the upper 95% confidence limit. Vertical solid line marks $\hat{A}_{1,2,\text{obs}}$. (a) and (b): AR(1) null model. (c) and (d): fGn null model. (e) and (f): fBm null model.

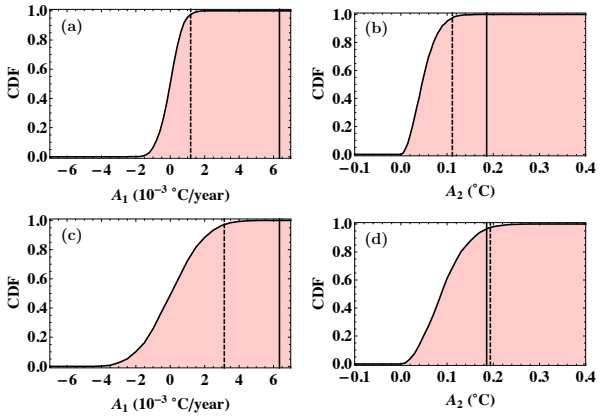


Fig. 5. Curved lines are CDFs for trend coeffecients $\hat{A}_1$ and $\hat{A}_2$ established from the null model ensemble for land data. Vertical dashed line marks the upper 95% confidence limit. Vertical solid lines mark $\hat{A}_{1,2,\text{obs}}$. (a) and (b): AR(1) null model. (c) and (d): fGn null model.

the fBm model. The oscillatory trend is insignificant for all models.

The corresponding CDFs for land data are shown in Figure 5. The linear trend is even more significant than for ocean data, while the oscillatory trend is significant for the AR(1) model, but barely significant for the fGn model.

One important lesson to learn from this analysis is that the stronger persistence in the ocean temperature record makes it harder to detect significant trends as compared to the land record. This effect outweighs the increased trend significance from the lower noise levels in the ocean record compared to the land record. Another is that the land record analysis establishes beyond doubt that there is a significant global linear trend throughout the last century, and that the reality of an oscillatory trend is probable, but not beyond the 95% confidence limit.

3 Constraining and Evaluating the Null Hypothesis

By estimating the parameters for the null model from the full observed record (without detrending), and allowing this model to be an LRM noise, we have selected the fractional noise model that is most likely to explain the variance of the full record. Hence it can be considered as the null model for the climate noise that is least likely to be rejected by the observed trend. If this null model is rejected, i.e., if the trend is found to be significant under this null, it is very unlikely that it will be found insignificant under other reasonable null hypotheses. Since we have found that the linear trend in the global land record *is* significant under this null, we should have very high confidence in this result. The non-significance of the oscillatory trend, however, deserves a reassessment in the light of the established significance of the linear trend. In a Bayesian spirit, it would be appropriate to investigate the oscillatory trend further by including the linear trend as

Table 2. Estimated noise parameters $\hat{\theta}_{\text{obs}}$ from the new null hypotheses in (4). The units are same as in Table 1.

	AR(1)	fGn	
	$\hat{\tau}_{\text{obs}}$	$\hat{\beta}_{\text{obs}}$	$\hat{\sigma}_{\text{obs}}$
Land	2.04	0.584	0.391

an established fact and construct a null model constrained to accept the existence of the linear trend;

3.1 A constrained null model yields significant oscillation

$$\hat{\delta}_{\text{obs}} + \hat{A}_{1,\text{obs}} t + \varepsilon(\theta; t). \tag{4}$$

We now first estimate a new $\hat{\theta}_{\text{obs}}$ by fitting the new null model (4) to the observed land record. The new estimated noise parameters are shown in Table 2. Then we produce a new null ensemble of records from the null model by drawing θ from the conditional distribution $P(\theta|\hat{\theta}_{\text{obs}})$. Finally we fit the trend model (3) to each realization in the ensemble and form $P(\hat{A}_1, \hat{A}_2)$. The result is shown for land data and $\varepsilon(\theta; t)$ modeled as an fGn in Figure 6a. The inclusion of the linear trend in the null model will imply that we shall fit $\varepsilon(\theta; t)$ to the record $\tilde{x}(t) \equiv x(t) - (\hat{\delta}_{\text{obs}} + \hat{A}_{1,\text{obs}} t)$ rather than to $x(t)$. Since we already have established that $x(t)$ contains a significant linear trend, the variability of $\tilde{x}(t)$ may be considerably less than the variability of $x(t)$ and hence the new estimated noise parameters $\hat{\theta}_{\text{obs}}$ may correspond to smaller $\hat{\sigma}_{\text{obs}}$ and $\hat{\beta}_{\text{obs}}$ than we obtained for the original null model. This reduction in noise parameters leads to narrowing of $P(\hat{A}_1, \hat{A}_2)$, and a narrower CDF for the oscillation trend parameter $\hat{A}_2$, as shown in Figure 6b. The result is that this constrained test establishes that the oscillatory trend is also significant.

3.2 Evaluation of the null model

The long-range memory associated with fractional noises and motions gives rise to larger fluctuations on long time scales that allows description of such variability as part of the noise background rather as trends. The implication is that variability which has to be described as significant trends under white noise or short-memory noise hypotheses may have to be classified as insignificant trends under an LRM null hypothesis. The issue of the most proper choice of null hypothesis was touched upon in section 2, but let us re-examine the issue in the light of the results we have obtained so far.

One way to deal with this issue is to apply an estimator that characterizes the correlation structure of the observed record and compare the outcome with those arising from applying the same estimator to different models for the climate-noise background. There are several estimators, for instance wavelet variances and detrended fluctuation analysis, that are well suited for extracting the scaling properties of a time series and estimating a β-exponent. For LRM processes such as fBm and fGn (which are respectively self-similar processes and the differences of self-similar processes) the fluctuation level of a time series varies as a power law versus time scale τ, and one can therefore analyze data using double-logarithmic plots of the so-called fluctuation functions. For processes with a characteristic time scale τ_c, such as the AR(1) processes, the fluctuation functions will not be power laws, and this can be seen from the estimated fluctuation functions. For an AR(1) process, which has an autocorrelation function on the form e^{-t/τ_c}, the time series behaves like a Brownian motion ($\beta = 2$) for time scales $t \ll \tau_c$ and a white noise process ($\beta = 0$) for $t \gg \tau_c$. If a time series is sufficiently long, the crossover between these two scaling regimes is clearly visible in the estimated fluctuation functions, and since we do not observe such crossovers in global temperature records, we can use fluctuation functions to illustrate that LRM processes are better suited than AR(1) processes as models for the global temperature. This idea is pursued in (*Rypdal and Rypdal*, 2013), where detrended fluctuation analysis is employed to show that a residual signal (constructed by subtracting the deterministic response to the external forcing) is inconsistent with an AR(1) process, but consistent with an LRM process.

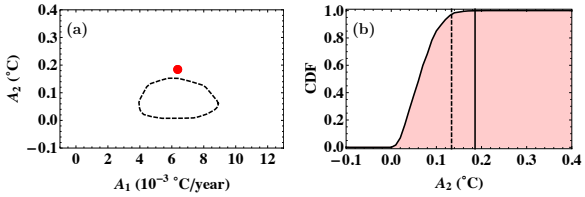


Fig. 6. (a): The 95% confidence contour of the distribution $P(\hat{A}_1, \hat{A}_2)$ for land data obtained by the new null model (4) with $\varepsilon(\theta; t)$ an fGn process. (b): The CDF derived from $P(\hat{A}_2)$ for this null model, with upper 95% confidence limit marked as dotted vertical line.

The test described above utilizes a method designed to estimate the scaling exponent β in LRM processes. As an alternative, we can use a test based on an estimator for the correlation time τ_c in an AR(1) process. For this test we should think of our time series as a discrete-time sampling of a continuous-time stochastic process. The continuous-time analog of an AR(1) process is the Ornstein-Uhlenbeck (OU) process. If a time series T_k is obtained from an OU process by sampling it at times $t_k = k\Delta t$, then the one-lag autocorrelation of T_k is $\phi^{(\Delta t)} = e^{-\Delta t/\tau_c}$. We can obtain a standard sample estimate $\hat{\phi}^{(\Delta t)}$ of the lag-one autocorrelation, and from this we obtain an estimate of the correlation time:

$$\hat{\tau}_c = \frac{\Delta t}{-\log \hat{\phi}^{(\Delta t)}}. \tag{5}$$

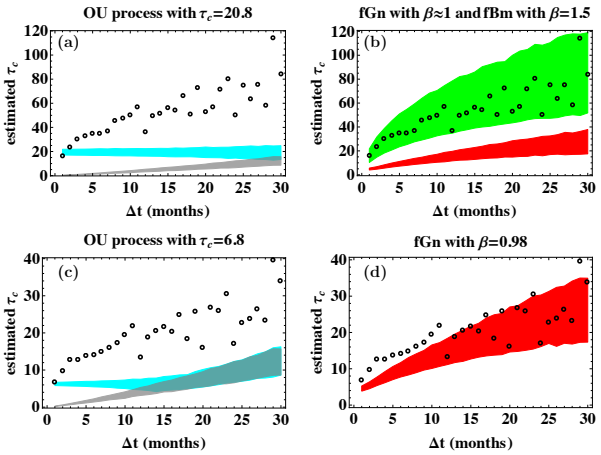


Fig. 7. Panels (a) and (b) show the estimated decorrelation time τ_c as a function of sampling time Δt for the ocean temperature (black circles) and for ensembles of synthetic realizations of three different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) and fBms (green) in panel (b). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the colored areas are the 95% confidence regions for these estimates. The gray area in panel (a) is the confidence region for τ_c for a white noise process. Panels (c) and (d) show the decorrelation time of the linearly detrended ocean temperature and for the synthetic realizations of the processes generated from the new null model; equation (4).

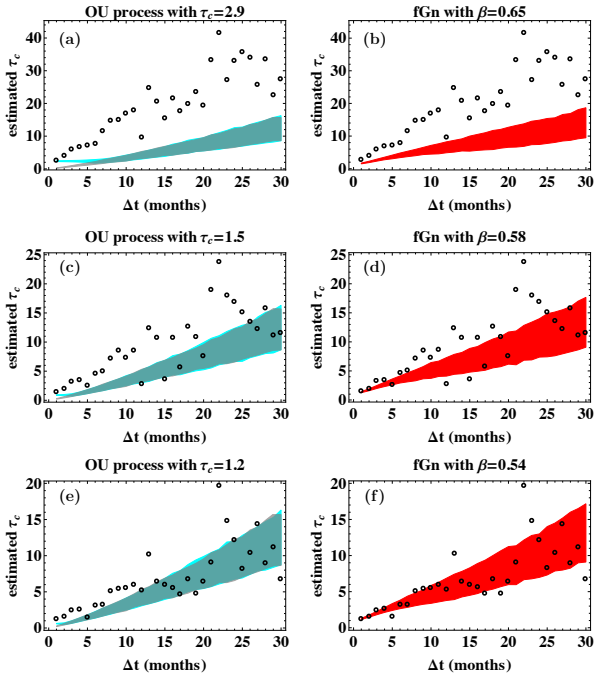


Fig. 8. Panels (a) and (b) show the estimated decorrelation time τ_c as a function of sampling time Δt for the land temperature (black circles) and for ensembles of synthetic realizations of three different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) in panel (b). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the colored areas are the 95% confidence regions for these estimates. The gray area in panel (a) is the confidence region for τ_c from a white noise process. Panels (c) and (d) show the decorrelation time of the linearly detrended land temperature and for the synthetic realizations of the processes generated from the new null model; equation (4). Panels (e) and (f) show the decorrelation time of the land temperature after removing the full trend; equation (3), and for the synthetic realizations of the processes generated from the detrended record by the MLE method.

Monte Carlo simulations show that this estimate is independent of Δt, as long as $\Delta t < \tau_c$. However, if the process is an fGn rather than an OU process, then the autocorrelation function of the time series T_k is approximated well by $(\beta+1)\beta(k\Delta t)^{\beta-1}$, and hence the lag-one autocorrelation is

$$\phi^{(\Delta t)} \approx (\beta+1)\beta\Delta t^{\beta-1}\,.$$

If τ_c is defined via $\tau_c = \Delta t/(-\log\phi^{(\Delta t)})$, then

$$\tau_c = \frac{\Delta t}{-\log(\beta+1)\beta-(\beta-1)\log\Delta t}\,.$$

This shows that OU processes and fGns can be distinguished by how an estimator of the correlation length depends on the sampling rate for the time series: For an OU process the estimate of τ_c is independent of Δt as long as $\Delta t < \tau_c$, and for fGns the estimates of τ_c grow with Δt. In Figures 7 and 8 we have plotted the estimates of τ_c according to equation (5) for ocean and land temperatures respectively, with and without linear detrending. For the land temperature, full detrending (removing the trend (3)) is also included. The estimates are shown as the circular plot markers in the figures. There is a clear increase in the τ_c estimate as Δt varies from 1 to 30 months. We have compared the results with Monte Carlo simulation of a white noise process, OU processes, fGns and fBms. Here the synthetic temperature series are constructed using parameters obtained by MLE. For the ocean temperature without detrending the test shows that the data is most consistent with a nonstationary fBm, and after linear detrending it is more consistent with an fGn than with an OU process. For the land temperature we observe that neither of the processes fit the data unless we perform a detrending, and for the detrended data there are only small differences between a white noise process, an OU process and the fGn with $\beta = 0.54$. The reason for this is that the ML estimate of τ_c is so small (close to the monthly time resolution of the temperature record) that the model OU process is effectively reduced to a white noise on all resolved time scales. The white noise process is a special case of an fGn, so the fGn class of pro-

cesses is clearly preferred in this case as well, although the test presented here is not suitable for estimating the β exponent. There are other tests that are better suited for accurate estimation of β, and if we apply these we will see that a persistent process ($\beta > 0$) is a better model for detrended land temperatures than white noise ($\beta = 0$) (*Rypdal et al.*, 2013).

The model selection test we have described here illustrates the important point that if one decides to model global temperature fluctuations as OU processes, then the choice of optimal model depends strongly on the time resolution of the time series. The same is not true for fGns and fBms, and this reflects the fact that global temperature data to a good approximation are scale invariant.

The method presented here can be seen as a generalization of the method presented by *Vyushin et al.* (2012), who attempt to distinguish between scale-free processes and AR(1) processes by considering estimates of $\phi^{(\Delta t)}$ for two different time resolutions Δt (monthly and annual). However, our results show that this test fails if the estimated τ_c is less than a year, which turns out to be the case for the land record. *Vyushin et al.* (2012) analyze a large number of local and regional time series and find that some are consistent with fGns, other with AR(1)s, but most are inconsistent with both. It is reasonable to expect that many of these records are in the category for which the test fails.

4 Conclusions

In this paper we have attempted to classify the various possible ways to understand the notion of a trend in the climate context, and then we have focused on the detection of a combination of a rising and oscillatory trend in global ocean and land instrumental data when no information about the climate forcing is used. It is well known that the statistical significance of the trends depends on the degree of autocorrelation (memory) assumed for the random noise component of the climate record (*Cohn and Lins*, 2005; *Rybski et al.*, 2006, 2009). It is also known that the linear trends are easier to detect and appear to be more significant in global than in local data (*Lennartz and Bunde*, 2009), although local records exhibit weaker long-term persistence than global records. Despite this fact, much effort is spent on establishing trends and their significance in data from local stations (e.g., *Franzke*, 2012b) with variable results. The failure of detecting consistent trends in local data records reflects the tendency of internal spatiotemporal variability to mask the trend that signals global warming, and we believe therefore that investigation of such trends should be performed on globally averaged data. For global data records our study demonstrates very clearly that the long-range memory observed in sea-surface temperature data leads to lower significance of detected trends compared to land data. This does not mean, of course, that the global warming signal and internal oscillations are not present in all of those records. It is just not possible to establish the statistical significance of these trends from these records alone, since the large short-range weather noise in local temperatures and the slower fluctuations in ocean temperature both reduce the possibilities of trend detection. Hence, one needs to search for the optimal climate record to analyze for detection of the global warming signal, and our results suggest that the global land temperature signal may be the best candidate for such trend studies.

While a linear trend is only marginally significant under the long-range memory null hypothesis in ocean data, it is clearly significant in land data. Hence, there should be no doubt about the significance of a global warming signal over the last 160 years even under null hypotheses presuming strong long-range persistence of the climate noise.

Assessment of the statistical significance of a linear trend is of course not the only way to detect the global warming signal in temperature records. An alternative hypothesis in the form of a second- or third-order polynomial trend would give a more precise, but more technically complex assessment. Other approaches are not based on trend estimates at all. Some methods compare spatiotemporal observations to patterns of natural variability obtained from global climate models. These patterns represent the null model, and the detection is typically performed through "fingerprint methods" rather than using just single observable such as the global temperature (*Hasselmann*, 1993; *Hegerl*, 1996). The validity of the method depends, of course, on the assumption that the climate model correctly describes the relevant aspects of the pattern of natural variability, e.g., the long-range correlation structure in space and time. This is not an obvious assumption, since there are significant differences between climate models in this respect (*Govindan et al.*, 2001; *Blender and Fraedrich*, 2003).

Other methods are based on null models like those considered in the present paper, but rather than estimating trends one estimates the probability of observing the recent clustering of record-breaking temperatures at the end of the instrumental record (*Zorita et al.*, 2008). The method is conceptually and technically simpler than the trend assessment, but it depends crucially on the assumption that the null model is strictly true on the shortest inter-annual time scales, since it assumes that the probability of variation from one year to the next is determined by this model. In contrast, the trend assessment emphasizes the properties of the null model on time scales up to a century, so it rather assumes the null model is strictly true on multi-decadal to century scales. The two approaches are complementary, but we believe the trend approach is better designed to detect the smooth, monotonic global warming signal, since it will be insensitive to particular interannual to decadal variability such as ENSO, or variability due to forcing from clusters of volcanic eruptions or solar-cycle variations. The elimination of these variabilities may be important for detection of the anthropogenic trend, as was shown by multiple regression techniques by *Foster and Rahmstorf* (2011) and *Lean and Rind* (2009). Moreover,

in the approach of *Zorita et al.* (2008) inclusion of the 70 yr oscillation in the null model would lead to enhanced probability of clustering of record-breaking temperatures at the end of the twentieth century, and hence a reduction of the significance of the warming signal. These are examples illustrating that one may arrive at misleading results without careful selection of the alternative as well as null models based on the data at hand and existing knowledge. In a Bayesian framework this is obvious.

Our initial analysis leaves some doubt about the significance of the 70 yr oscillatory mode in the global signal, as shown in Figure 4d and 4f and Figure 5d. By means of a Bayesian iteration, however, utilizing the established significance of a linear trend to formulate a constrained null hypothesis, we are able to establish statistical significance of the oscillatory trend in the land data record. We believe this is an important result, because it means that we cannot dismiss this oscillation as a spontaneous random fluctuation in the climate noise background. By the analysis presented here we cannot decide whether this oscillation is an internal mode in the climate system or an oscillation forced by some external influence. Such insights can be obtained from a generalization of the response model of *Rypdal and Rypdal* (2013) by employing information about the climate forcing, and will be the subject of a forthcoming paper. There are various published hypotheses about the nature of this oscillation. The least controversial is that this is a global manifestation of the Atlantic Multidecadal Oscillation (AMO) which is essentially an internal climate mode (*Schlesinger and Ramankutty*, 1994). Some authors go further and suggest that this oscillation is synchronized and phase locked with some astronomical influence (*Scafetta*, 2011, 2012). Although some of these suggestions seem very speculative, there are some quite well-documented connections between periodic tidal effects on the Sun from the motion of the giant planets and radioisotope paleorecord proxies for solar activity on century and millennium time scales (*Abreu et al.*, 2012). So far there exists no solid evidence that these, and multidecadal, variations in solar activity have a strong influence on terrestrial climate, but the issue will probably be in the frontline of research on natural climate variability in the time to come. The work presented here cannot shed light on the physical cause of this oscillation, but it presents evidence that it is a phenomenon that stands out from the long-memory background of random temperature fluctuations. Its importance for our assessment of anthropogenic global warming is obvious from the observation that the oscillation seems to peak at the turn of the millennium and hence provides a possible explanation of the current hiatus in global temperature.

Acknowledgements. The authors are grateful to Ola Løvsletten for illuminating discussions and comments.

References

Abreu, J. A., Beer, J., Ferriz-Mas, A. , McCracken, K. G., and Steinhilber F. :, Is there a planetary influence on solar activity?, Astronomy and Astrophys., A88, 548-557, doi:10.1051/0004-6361/201219997, 2012.

Beran, J.: Statistics for Long-memory Processes, Monographs on statistics and applied probability, Chapman& Hall/CRC, Boca Raton, 1994.

Blender, R. and Fraedrich, K.: Long time memory in global warming simulations, Geophys. Res. Lett., 30, 1769, doi: 10.1029/2003GL017666, 2003.

Bloomfield, P. and Nychka, K.: Climate spectra and detecting climate change, Clim. Change, 21, 275-287, doi: 10.1007/BF00139727, 1992.

Cohn, T. A. and Lins, H. F.: Nature's style: Naturally trendy, Geophys. Res. Lett., 32, L23402, doi:10.1029/2005GL024476, 2005.

Efstathiou, M. N., Tzanis,C., Cracknell, A.P., and Varotsos, C. A.: New features of land and sea surface temperature anomalies, Int. J. Remote Sensing, 32, 3231-3238, doi:10.1080/01431161.2010.541504, 2011.

Fatichi, S., Barbosa, S. M., Caporali, E., and Silva, M. E.: Deterministic versus stochastic trends: Detection and challenges, J. Geophys. Res., 114, doi: 10.1029/2009JD011960, 2009.

Flandrin, P.: Wavelet analysis and synthesis of fractional Brownian motion, IEEE Trans. Inform. Theory, 38, 910-917, doi:10.1109/18.119751, 1992.

Foster, G. and Rahmstorf, S.: Global temperature evolution 1979–2010, Environ. Res. Lett., 6, 044022, doi:10.1088/1748-9326/6/4/044022, 2011.

Franzke, C.: Multi-scale analysis of teleconnection indices: Climate noise and nonlinear trends, Nonlinear Processes Geophys., 16, 65-76, 2009.

Franzke, C.: Long-range dependence and climate noise characteristics of Antarctic temperature data, J. Clim., 23, 6074-6081, doi:10.1175/2010JCLI3654.1, 2010.

Franzke, C. and T. Woollings, T.: On the persistence and predictability properties of North Atlantic climate variability, J. Clim., 24, 466-472, 2011.

Franzke, C.: Nonlinear trends, long-range dependence and climate noise properties of surface air temperature, J. Clim., 25, 4172-4183, 2012a.

Franzke, C.: On the statistical significance of surface air temperature trends in the Eurasian Arctic region, *Geophys. Res. Lett.*, *39*, L23705, doi:10.1029/2012GL054244, 2012b.

Franzke, C., Graves, T. , Watkins, N.W., Gramacy, R. B., and Huges C.: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, Philps. Trans. R. Soc. A, 370, 1250-1267, doi:10.1098/rsta.2011.0349, 2012.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, Texts in Statistical Science Series, Chapman& Hall/CRC, Boca Raton, 2004.

Gil-Alana, L. A.: Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques, J. Clim., 18, 5357-5369,doi: http://dx.doi.org/10.1175/JCLI3543.1, 2005.

Govindan, R. B., Vjushin, D., Brenner, S., Bunde, A., Havlin, S., and Schellnhuber H.-J.: Long-range correlations and trends in global climate models: Comparison with real data, Physica A, 294, 239-248, 2001.

Halley, J., and Kugiumtzis, D.: Nonparametric testing of variability and trend in som climatic records, Clim. Change, 109, 549-568, doi:10.1007/s10584-011-0053-5, 2011.

Hasselmann, K.: Optimal fingerprints for the detection of time dependent climate change, J. Climate, 6, 1957-1971, 1993.

Hansen, J. and Sato, M. and Ruedy, R., et al.: Climate simulations for 1880-2003 with GISS modelE, Climate Dynam., 29, doi: 10.1007/s00382-007-0255-8, 2007.

Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K.: Earth's energy imbalance and implications. Atmos. Chem. Phys., 11, doi:10.5194/acp-11-13421-2011, 2011.

Hegerl, G. C., von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U., Jones, P. D.: Detecting greenhouse gas induced climate change with an optimal fingerprint method, J. Climate, 9, 2281-2306, 1996.

Hu, K., Ivanov, P. C., Chen, Z, Carpena, P., and Stanley H. E.: Effect of trends on detrended fluctuation analysis, Phys. Rev. E, 64, 011114, doi: 10.1103/PhysRevE.64.011114, 2001.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res.,117, D05127, doi:10.1029/2011JD017139, 2012.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. A., Havlin, S., and Bunde, A.: Detecting long-range correlations with detrended fluctuation analysis, Physica A, 295, 441-454, 2001.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220, 2011.

Lean, J. L. and Rind, D. H.: How will Earth's surface temperature change in future decades? Geophys. Res. Lett., 36, L15708, doi:10.1029/2009GL038932, 2009.

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

Lovejoy, S., and Schertzer, D.: The Weather and Climate: Emergent Laws and Multifractal Cascades, Cambridge University Press.

Malamud, B. L., and Turcotte, D.: Self-Affine Time Series: I. Generation and Analyses, Adv. Geophys., 40, 1-90, doi:10.1016/S0065-2687(08)60293-9, 1999.

Markonis, Y. and Koutsoyiannis, D.: Climatic Variability Over Time Scales Spanning Nine Orders of Magnitude: Connecting Milankovitch Cycles with Hurst-Kolmogorov Dynamics (2013), Surv. Geophys., 34, 181-207, doi:10.1007/s10712-012-9208-9, 2013.

McLeod, A. I., Yu, H., and Krougly, Z. L.: Algorithms for linear time-series analysis, J. of Statistical Software, 23, 1-26, 2007.

Pelletier, J. D., and Turcotte, D.: Self-Affine Time Series: II. Applications and Models, Adv. Geophys., 40, 91-166, doi:10.1016/S0065-2687(08)60294-0, 1999.

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

Rybski, D., and Bunde, A.: On the detection of trends in long-term correlated records, Physica A, 388,1687-1695, doi: 10.1016/j.physa.2008.12.026, 2009.

Rypdal, M., and Rypdal, K.: Testing hypotheses about sun-climate complexity linking, Phys. Rev. Lett., 104, 128501, doi:10.1103/PhysRevLett.104.128501, 2010.

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

Rypdal, M., and Rypdal, K.: , 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, 2012.

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

Rypdal, M., and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, submitted to J. Climate, http://arxiv.org/pdf/ 1305.5080v1.pdf.

Scafetta, N.: A shared frequency set between the historical mid-latitude aurora records and the global surface temperature, Journal of Atmospheric and Solar-Terrestrial Physics, 74, 145-163, doi:10.1016/j.jastp.2011.10.013, 2011.

Scafetta, N.: Testing an astronomically based decadal-scale empirical harmonic climate model versus the IPCC (2007) general circulation models, Journal of Atmospheric and Solar-Terrestrial Physics, 80, 124-137, doi:10.1016/j.jastp.2011.12.005, 2012.

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

Vallis, G. K.: Climate and the Oceans, Princeton Primers in Climate, Princeton University Press, Princeton, 2012.

von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, 1999.

Zorita, E., Stocker, T. F., and von Storch, H.: How unusual is the recent series of warm years, Geophys. Res. Lett., 35, L24706, doi:10.10029/2008GL036228, 2008.

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

Paper III

Long-Range Memory in Millennium-Long ESM and AOGCM Experiments

Submitted to *Earth System Dynamics Discussions*.

Manuscript prepared for J. Name
with version 5.0 of the LaTeX class copernicus.cls.
Date: 27 February 2014

Long-Range Memory in Millennium-Long ESM and AOGCM Experiments

L. Østvand[1], T. Nilsen[2], K. Rypdal[2], D. Divine[3], and M. Rypdal[2]

[1]Department of Physics and Technology, UiT The Arctic University of Norway, Norway
[2]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway
[3]Norwegian Polar Institute, Tromsø, Norway

Correspondence to: Lene Østvand
(lene.ostvand@uit.no)

Abstract. Northern Hemisphere (NH) temperature records from a reconstruction and a number of millennium-long climate model experiments are investigated for long-range memory (LRM). The models are two Earth system models and two atmospheric-ocean general circulation models. The periodogram, detrended fluctuation analysis and wavelet variance analysis are applied to examine scaling properties and to estimate a scaling exponent of the temperature records. A simple linear model for the climate response to external forcing is also applied to the reconstruction and the forced climate model runs, and then compared to unforced control runs to extract the LRM generated by internal dynamics of the climate system. With one exception the climate models show strong persistent scaling with power spectral densities of the form $S(f) \sim f^{-\beta}$ with $0.8 < \beta < 1$ on time scales from years to several centuries. This is somewhat stronger persistence than found in the reconstruction ($\beta \approx 0.7$). The exception is the HadCM3 model, which exhibits $\beta \approx 0.6$. We find no indication that LRM found in these model runs are induced by external forcing, which suggests that LRM on sub-decadal to century time scales in NH mean temperatures is a property of the internal dynamics of the climate system. Temperature records for a local site, Reykjanes Ridge, are also studied, showing that strong persistence is found also for local ocean temperature.

1 Introduction

The presence of long-range memory (LRM) in climatic records is well documented in the geophysics literature. LRM is characterized by an algebraically decaying autocorrelation function $\lim_{t\to\infty} C(t) \propto t^{-\gamma}$ such that $\int_0^\infty C(t)\mathrm{d}t = \infty$, i.e., $0 < \gamma \leq 1$. Equivalently, the power spectral density (PSD) of LRM time series follows a power law $\lim_{f\to 0} S(f) \propto f^{-\beta}$, where $\beta = 1-\gamma$ and $0 < \beta < 1$. A typical model for an LRM stochastic process is the persistent fractional Gaussian noise (fGn). This is a stationary process with $0 < \beta < 1$. The cumulative integral (or sum) of such a process has the PSD of the form $S(f) \sim f^{-\beta}$, but with $\beta \to \beta+2$. Such a process with $1 < \beta < 3$ is a non-stationary LRM process called a fractional Brownian motion (fBm).

Because of the noisy nature of PSD estimators like the periodogram, other methods for estimating β are preferred (Beran et al., 2013). In this paper we use the periodogram as the first crude characterization of the data and for detection of spectral peaks indicating lack of power-law scaling, but employ Detrended Fluctuation Analysis (DFA) and Wavelet Variance Analysis (WVA) for parameter estimation.

Most of the LRM studies of climatic time series investigate local time records (e.g., Pelletier, 1997; Weber and Talkner, 2001; Eichner et al., 2003), but LRM has also been found in global observed temperature records (Lennartz and Bunde, 2009) and reconstructed temperature records for the Northern Hemisphere (Rybski et al., 2006; Mills, 2007). Some surface temperature records from AOGCM climate models have been analyzed with the main result that LRM is not reproduced in agreement with that of observational temperature (Syroka and Toumi, 2001; Bunde et al., 2001; Govindan et al., 2001, 2002; Bunde and Havlin, 2002). Some of the model experiments produce temperature with multiple scaling regimes, and some of them yield smaller scaling exponents than the observational temperature. However, in (Syroka and Toumi, 2001; Bunde et al., 2001; Govindan et al., 2001) the model experiments all had greenhouse gas forcing as the only dynamic forcing, while remaining external forcings, such as total solar irradiance and volcanic effects,

were kept constant. Govindan et al. (2002) and Bunde and Havlin (2002) used experiments where (i) all forcings were fixed, (ii) with fixed forcings except greenhouse gas forcing, and (iii) with fixed forcings except greenhouse gas plus aerosol forcing. Their main conclusion was that the temperature from the model experiments fail to reproduce the scaling behaviour found in observational data, and that the models display large differences in scaling from different sites. Of these scenarios, the one with dynamic greenhouse gas plus aerosol forcing performed better with respect to producing the scaling observed in instrumental temperature records. Global fields of observed and simulated surface temperatures from an AOGCM climate model experiment were studied in Fraedrich and Blender (2003). The experiment was run with fixed forcings. The result from observational data was mostly in agreement with previous studies of temperature in oceanic and coastal regions, but the authors found white noise scaling ($\beta \approx 0$) at continental interiors. Analysis of a 1000-year temperature simulation from the model experiment produced similar scaling exponents to what was found for the observational data in this study. Blender and Fraedrich (2003) made a similar analysis of temperature from two different model experiments with dynamic greenhouse gas forcing, giving results in agreement with Fraedrich and Blender (2003).

Temperature from model experiments with constant forcings, and time-varying greenhouse gas, sulfate aerosol, ozone, solar, volcanic forcing and various combinations was studied in Vyushin et al. (2004). Scaling exponents for temperature at 16 land sites and 16 sites in the Atlantic ocean were estimated. They found that inclusion of volcanic forcing considerably improved the scaling behavior. Rybski et al. (2008) used model experiments with all constant forcing and with dynamic solar, volcanic and greenhouse gas forcing. They analyzed data from grid cells all over the globe, but did not investigate global or hemispheric means. They found that for the forced run experiment the temperature showed a scaling exponent in agreement with observational temperature, while the temperature from the control run showed generally lower persistence.

Studies of LRM in temperature records from climate model experiments mostly use temperature from local sites, and some also use temperature spatially averaged over larger regions. Global mean temperature was studied by Syroka and Toumi (2001), but hemispheric means have not been studied with regards to LRM. For observational and reconstructed temperature, global and hemispheric means are also far less studied than local data.

In the present study we analyze scaling properties of surface temperature for the Northern Hemisphere from paleoclimate simulations and compare to those of temperature reconstruction by Moberg et al. (2005) which spans the last two millennia. Hemispheric temperature records from four different Earth system climate models are analyzed, and both forced runs and control runs are investigated. In order to avoid effects of anthropogenic forcing only data up to the year 1750 is used. This will give an idea of what role other natural external forcing like solar, CO_2, volcanic and aerosol forcing play in producing LRM, and indicate if LRM can arise from internal dynamics alone.

Separation of the LRM arising from internal dynamics from the LRM induced by external forcing can also be achieved from reconstructed and simulated temperature data if the forcing data are known. The method makes use of a simple linear model for the global temperature response (Rypdal and Rypdal, 2013). The response to the external forcing can then be computed and subtracted from the observed or modeled temperature record to yield a residual which represents the internal variability of the climate system. Analysis of this residual and temperature from forced runs and control runs are compared for those models where temperatures from both forced runs and control runs are available.

This paper is organized as follows: Section 2 describes the DFA and WVA methods and the response model. Information about the models and the data used can be found in Section 3, and the results from the analysis are presented in section 4. Discussion and conclusion follow in section 5.

2 Methods

2.1 Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis (DFA) (Peng et al., 1994; Kantelhardt et al., 2001) was explicitly designed to remove polynomial trends in the data. The method can be summarized in four steps. First, we construct the cumulative sum (the "profile") of the temperature time series $x(t)$; $Y(i) = \sum_{t=1}^{i} x(t) - \langle x \rangle$, where $\langle x \rangle$ denotes the mean. In the second step the profile is divided into $N_\tau = N/\tau$ non-overlapping segments of equal length τ. This is done both starting at the beginning and at the end of the profile, so $2N_\tau$ segments are obtained altogether. In the third step, an n'th order polynomial is fitted to, and then subtracted from, each segment. Thus, at this stage we have formed the detrended profile $Y_\tau(i) = Y(i) - p_\nu(i)$, where $p_\nu(i)$ is the polynomial fitted to the ν'th segment. In the final step, the variance of each segment, $F^2(\nu,\tau) = \frac{1}{\tau}\sum_{i=1}^{\tau} Y_\tau^2[(\nu-1)\tau + i]$, is computed. The fluctuation function is given by the square root of the average over all the segments,

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

The scaling parameter β is found through the relation

$$F(\tau) \propto \tau^{(\beta+1)/2}.$$

What we have described is the n'th order detrended fluctuation analysis, denoted DFAn. It has the property of eliminating the effect of an $n-1$'th order polynomial trend. In this

paper we employ second order DFA, denoted DFA2, which eliminates linear trends.

2.2 Wavelet Variance Analysis

The continuous wavelet transform is the convolution between a time series $x(t)$ and the rescaled wavelet $\Psi(t/\tau)$;

$$W(t,\tau;x(t),\Psi(t)) = \int_{-\infty}^{\infty} x(t')\frac{1}{\sqrt{\tau}}\Psi\left(\frac{t'-t}{\tau}\right)\mathrm{d}t'.$$

The mother wavelet $\Psi(t)$ and all rescaled versions of it must fulfill the criteria $\int_{-\infty}^{\infty}\Psi(t')\,\mathrm{d}t' = 0$. For LRM time series, the variance $F(\tau) = (1/N)\sum_{t=1}^{N} W^2(t,\tau)$ scales as a power-law (Flandrin, 1992; Malamud and Turcotte, 1999),

$$F(\tau) \sim \tau^{\beta}.$$

The method is therefore known as the wavelet variance analysis (WVA). In this study we have used the Mexican hat wavelet, which is capable of eliminating linear trends, and denote the method WVA2. The properties of the WVA2 analysis are similar to the DFA2 in that it usually yields similar values of β. It is, however, much more sensitive to the presence of additional oscillations in the data, which show up as wavy structures in the function $F(\tau)$. We use it in this paper mainly as a tool (in addition to the periodogram) to detect such oscillations.

2.3 The response model residual analysis

For the preindustrial period the most important contributions to the external radiative forcing $F(t)$ are orbital, solar variability, and aerosols from volcanic eruptions. Orbital forcing can be computed with high accuracy, and total solar irradiation has been reconstructed for the last ten millennia. Existing reliable reconstructions of volcanic forcing cover the last millenium. The forcing data used here are further described in Section 3. The evolution of the global mean surface temperature anomaly T on decadal to millennial time scales can tentatively be modeled as a linear response to $F(t)$ in addition to a response to stochastic forcing from unresolved spatiotemporal "turbulence" (e.g., forcing of the sea-surface temperature from atmospheric weather systems). A simple stochastic-dynamic model (SDM) with an LRM response function is (Rypdal and Rypdal, 2013):

$$T(t) = \mu[\underbrace{\int_0^t (t-s)^{\beta/2-1}F(s)\mathrm{d}s}_{\text{deterministic solution}} + \sigma\underbrace{\int_0^t (t-s)^{\beta/2-1}\mathrm{d}B(s)}_{1/f^{\beta}\text{ noise}}].$$

Here $B(s)$ is the Wiener stochastic process whose increments $\mathrm{d}B(s)$ is a Gaussian white noise process and $\sigma\mathrm{d}B(s)$ represents the stochastic component of the forcing. $T(t)$ is the temperature relative to the temperature T_0 at time $t = 0$ (the beginning of the record) and $F = \tilde{F} + F_0$ is perturbed forcing $\tilde{F}$ relative to that of a radiative equilibrium at surface temperature T_0 plus the actual radiative imbalance F_0 at $t = 0$. By definition $\tilde{F}(0) = 0$. F_0 is a model parameter which is estimated from the data along with the other model parameters β, μ, and σ. 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 fractional Gaussian noise (a stationary process) if $\beta < 1$ and a fractional Brownian motion (nonstationary) if $1 < \beta < 3$.

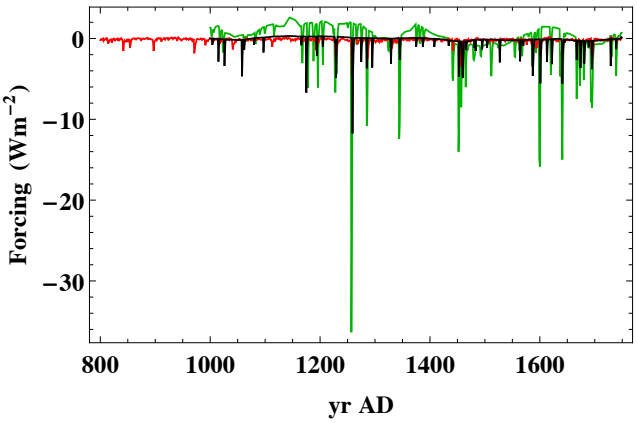


Fig. 1. The different forcings used as input to the response model, i.e. Crowley forcing used with the Moberg reconstruction (black), forcing used in the COSMOS experiment (red) and forcing used in the LOVECLIM experiment (green).

Time-series information about global climate forcing and its various components exists for the instrumental period as well as for the last millennium. This information can be used in conjunction with the observed temperature records to perform maximum-likelihood estimates (MLE) of the parameters of the model. The details of the MLE method applied to this response model are explained in Rypdal and Rypdal (2013). In a short-range memory response model, the power-law kernel $(t-s)^{\beta/2-1}$ in the response model is replaced with an exponential $e^{-(t-s)/\tau}$, where τ is the time constant. In this case the parameter μ^{-1} can be interpreted as the effective heat capacity of the climate system. In the LRM response model μ^{-1} does not have a simple physical interpretation, although it is (in combination with β) a measure of the thermal inertia of the system. The memory parameter β estimated from this model should be interpreted as the LRM parameter for the internal temperature response, and hence the problem of separating the LRM contribution from the forcing and the internal LRM has been eliminated. The β estimated in Rypdal and Rypdal (2013) is $\beta \approx 0.75$, which is not much lower than the value estimated for the full temperature record from detrending techniques like DFA and WVA. This shows that

the detrending techniques effectively eliminate the contribution to β from the anthropogenic trend.

In the present paper the Crowley forcing (Crowley, 2000) is used for Moberg reconstructed temperature (Moberg et al., 2005) and for the temperature from the ECHO-G forced run experiment. The COSMOS experiment was run with a different forcing, and this forcing is used as input to our response model. For the temperature from the LOVECLIM experiment, solar and volcanic forcings were used together with forcings from CO_2 and tropospheric aerosols corresponding to the Crowley forcing. The full forcing data in these three cases are shown in Figure 1.

3 Data

3.1 The reconstruction of Moberg et al. (2005)

The reconstructed temperature presented in Moberg et al. (2005) is a Northern Hemisphere reconstruction covering the time period 1-1979 AD. The reconstruction is created from 11 low-resolution proxy time series (e.g. ice cores and sediments, 1-180 year resolution) and 7 tree-ring records (annual resolution). The 18 local reconstructed temperature time series were first divided into an Eastern and a Western part. Linear interpolation was then applied to all time series in order to create annual mean values. The beginning and end of the time series were padded with surrogate data so that they all covered the time period 300 BC - 2300 AD to minimize edge effects of the wavelet transform. The wavelet transform (WT) with the Mexican hat wavelet basis function was then applied using the set of 22 scales to generate 22 time series. For each scale 1-9 (Fourier timescales $<$80 years), the WT from the tree-ring proxy series were averaged. For the scales 10-22 (Fourier timescales $>$80 years), the WT from the low-resolution proxy series were averaged. Scale 1-22 were then merged, creating two full WT time series, one for the Eastern and one for the Western Northern Hemisphere. The two subsets were then averaged, and the inverse WT was calculated, creating a dimensionless NH temperature reconstruction. Finally, the mean and variance of the reconstructed temperature time series were calibrated to correspond to the instrumental data available for the time period 1856-1978.

3.2 Marine sediment SST reconstruction; Reykjanes Ridge

The local sea surface temperature (SST) reconstruction applied in the following study is presented in detail in Miettinen et al. (2012). Past August SST has been reconstructed by analyzing marine planktonic diatoms from a composite marine sediment core, recovered at the Reykjanes Ridge in the western subpolar North Atlantic, (57°27.09'N, 27°54.53'W, at 2630 m water depth). The composite core consist of a 54.3 cm long box core, and a 3.725 m long gravity core. The general assumption is that the down-core diatomic microfossil assemblages are related with past environmental conditions at the core site. Marine diatoms are unicellular, photosynthetic algae with siliceous frustules. For this particular analysis, the down-core diatomic assemblages were converted to August SST estimates by the weighted-average partial least squares technique (ter Braak and Juggins, 1993). The SST reconstruction has an average temporal resolution of 2 years for year 1770-2000 (box core), and 8-10 years for year 1000-1770 (gravity core).

3.3 SST reconstruction from observations; Reykjanes Ridge

A reconstruction based on instrumental observations was developed in Smith and Reynolds (2005). For the ocean, sea surface temperature (SST) was used, while surface marine air temperatures where left out due to biases in the daytime temperatures. The SST analysis and a separate land surface air temperature analysis were merged to form a monthly merged analysis from 1880 to 1997. The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) SST observations release 2 was the primary SST data, but the combined satellite and in situ SST analysis of Reynolds et al. (2002) was also included. The reconstruction was separated into low- and high-frequency components, which were added for the total reconstruction. The low frequency was reconstructed using spatial and temporal filtering, with a time filter of 15 yr. The low-frequency component was subtracted from the data before reconstrucion of the high-frequency component using spatial covariance modes. The method for reconstructing the data is described in detail in Smith and Reynolds (2004). This reconstruction contains improvements over many earlier studies: It is globally complete, incorporates updates in ICOADS, the analysis variance have less dependence on sampling compared to some earlier analysis, and uncertainty estimates indicate when and where the analysis is most reliable.

3.4 LOVECLIM model and experiment

The Earth system model LOVECLIM version 1.2 contains a quasi-geostrophic model for the atmosphere (ECBilt2), coupled to an ocean GCM (CLIO3) (Goosse et al., 2010). The two models have 3 and 20 vertical levels, respectively. A thermodynamic sea-ice model is incorporated into the OGCM, and the vegetation model VECODE is used to simulate the dynamics of trees, grasses and deserts. It includes the evolution of the terrestrial carbon cycle, while a separate model LOCH simulates the ocean carbon cycle. Both the solubility and the biological pumps are included in this model. Incorporated in LOVECLIM is also the ice-sheet model AGISM, which consists of 3 modules; ice sheet flow, visco-elastic bedrock and mass balance at the ice-atmospehere and ice-ocean interfaces.

Table 1. Information on temperature from model experiments

Climate model	LOVECLIM ESM v. 1.2	COSMOS ESM	ECHO-G	HadCM3
Complexity	Interm.	GCM	GCM	GCM
Time period covered	500-1750 AD	800-1750 AD	1000-1750 AD	850-1750 AD
Temp.res	annual	monthly	monthly	monthly
Spat.res	5.63x5.63 degrees	3.75x3.75 degrees	3.75x3.75 degrees	1.25x1.25 degrees

We apply surface temperature data from one experiment with this model; "LOVECLIM Climate Model Simulation Constrained by Mann et al. 2009 Reconstruction" (Goosse et al., 2012). In this experiment, simulations are constrained by the mean surface temperature reconstruction of Mann et al. (2009). External forcing includes TSI (total solar irradiance), volcanic eruptions, land cover changes, orbital forcing, greenhouse gases and aerosols. When we implement the response model to these data, only time series for the solar, volcanic and greenhouse gas forcing are applied. The solar forcing time series is based on the reconstruction by Muscheler et al. (2007). The volcanic activity time series originate from Crowley et al. (2003) , while the greenhouse gas forcing used is obtained from (Crowley, 2000).

3.5 COSMOS ESM model and experiments

The COSMOS ESM model consists of GCMs for the atmosphere and the ocean (Jungclaus et al., 2010). The atmospheric model ECHAM5 (Roeckner et al., 2003) has 19 vertical levels, while the ocean model MPIOM (Marsland et al., 2003) has 40. A thermodynamic sea-ice model is incorporated into the OGCM. Additional modules include the ocean biogeochemistry model HAMOCC5 (Wetzel et al., 2006), and the terrestrial biosphere model JSBACH (Raddatz et al., 2007).

The surface temperature data applied in our analysis are extracted from one experiment in a set of experiments referred to as "Ensemble Simulation of the Last Millenium using the Comprehensive COSMOS Earth System Model" (Jungclaus et al., 2010). External forcing used in the forced simulations include TSI, volcanoes, orbital forcing, greenhouse gases and land use change. An unforced control run is also used here in the comparative LRM study.

For the response model, time series for solar, volcanic and greenhouse gas forcing are applied. The forcing time series used are created specifically for this model and experiment (Jungclaus et al., 2010). The solar forcing time series is based on a combination of reconstructions; from the Maunder Minimum (1647-1715 AD) until today the total solar irradiance (TSI) is based on historical sunspot records (Krivova and Solanki, 2007; Balmaceda et al., 2007), and between 800 AD and the Maunder Minimum the TSI is reconstructed from estimates of the solar open magnetic flux based on ^{14}C concentrations in tree rings (Solanki et al., 2004; Krivova and Solanki, 2008; Usoskin et al., 2011). An 11-year solar cycle has been superposed on this part of the reconstruction.

The relative radiative forcing from volcanic eruptions is calculated from aerosol optical depth (AOD) and effective radius R_{eff}. Satellite data from the 1991 Mt. Pinatubo eruption is the basis for these estimates. The greenhouse gas forcing includes CO_2, where concentrations are computed within the model, based on historical records of fossil fuel emissions by Marland et al. (2003).

3.6 ECHO-G model and experiments

The coupled model ECHO-G (Legutke and Voss, 1999) version 4 consist of GCMs for the ocean/sea ice and the atmosphere. The atmospheric model ECHAM4 (Roeckner et al., 1996) includes 19 vertical levels, while the ocean model HOPE-G (Legutke and Maier-Reime, 1999) includes 20 levels. External forcing includes volcanoes, solar irradiance and greenhouse gases, all derived from Crowley (2000) and is used in the response model study. Surface temperature from two experiments is used for analysis; one forced run and one control run with forcing values fixed to year 1990 (Zorita et al., 2003; González-Rouco et al., 2003; von Storch et al., 2004).

3.7 HadCM3 model and experiment

The Hadley Centre coupled model 3 is an AOGCM (Gordon et al., 2000), with 19 levels in the atmospheric component HADAM3 (Pope et al., 2000) and 20 levels on the ocean HADOM3 component. External forcing is constant for this experiment (Collins et al., 2000).

The complexity, time period coverered, and temporal and spatial resolution for each model experiment are given in Table 1.

4 Results

When applying DFA2 and WVA2 directly to the temperature reconstruction records, all data up to the year 1750 are used. Because the Crowley forcing record starts at 1000 AD, only data from this year and forward were used in the re-

Table 2. Estimated β from applying DFA2 and WVA2 directly to the full temperature record (all temperatures), from DFA2 and WVA2 applied to the residuals from the deterministic response, and from the response model using MLE (temperature reconstruction and temperature from the forced climate model run experiments). The scaling ranges for DFA2 and WVA2 are also shown in years for Moberg and LOVECLIM, and in months otherwise. The same scaling range has been used for the full record and the residual, except for WVA2 applied to the forced ECHO-G experiment, where the upper scale used for the residual is in parenthesis.

	Full record		Residual		Response model	Scaling range					
	DFA2	WVA2	DFA2	WVA2	MLE	DFA2			WVA2		
Moberg	0.69	0.60	0.59	0.42	0.75	16	-	512	16	-	215
LOVECLIM	0.98	0.96	1.0	0.97	0.95	8	-	256	4	-	181
COSMOS forced	0.82	0.79	0.77	0.72	0.61	8	-	4096	4	-	4096
COSMOS ctrl	0.82	0.87				8	-	2048	4	-	512
ECHO-G forced	0.91	0.90	0.72	0.84	0.75	8	-	2048	4	-	1448 (304)
ECHO-G ctrl	0.85	0.87				8	-	2048	4	-	724
HadCM3 ($\tau < 10$ yr)	1.9	2.0				8	-	128	4	-	45
HadCM3 ($\tau > 10$ yr)	0.63	0.65				128	-	8192	45	-	4096

sponse model residual analysis of the Moberg reconstructed temperature and the temperature from the LOVECLIM experiment. Therefore the scales shown in the plots for DFA2 and WVA2 may differ somewhat between the full record and the residual from the deterministic response. Table 2 shows the resulting β from applying DFA2 and WVA2 directly to the full temperature and to the residuals from the deterministic response. The β estimated using the response model is also given, where the parameters of the response model are estimated with the MLE method (Rypdal and Rypdal, 2013). The response model residual analysis is applied to the temperature reconstructions and the temperature from forced climate model experiments, while the direct analysis also include temperature from control runs. The scaling ranges used to find β with DFA2 and WVA2 are also indicated in this table. Note that the scaling range is given in years for the Moberg and LOVECLIM temperatures, and in months otherwise. Figures 2-12 show the analysis of the temperature records. For the full records, the figures show (a) the temperature data, and (b) PSD, (c) DFA2, and (d) WVA2 applied to the data set. For the response model results, the figures show (a) the temperature data and deterministic response, (b) the residual, and (c) PSD and (d) DFA2 applied to the residual. The residual is the deterministic response interpolated to have the original time resolution subtracted from the full-resolution temperature data. For DFA2 and WVA2 95% confidence areas are shown, computed from Monte Carlo ensembles of synthetic fGns. For the response model analysis with DFA2, they are given the β estimated by MLE, otherwise they are given β corresponding to the one found from the DFA2 and WVA2 analysis respectively. The values of β used are indicated in each figure.

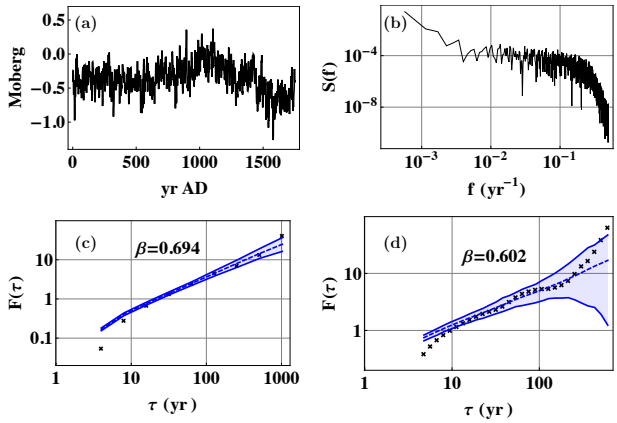


Fig. 2. (a) The Moberg reconstructed temperature. (b) PSD, (c) DFA2 and (d) WVA2 applied to the reocrd. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

4.1 Results from paleoreconstruction of Moberg et al. (2005)

Figure 2 shows the Moberg reconstructed temperature record, its power spectral density (PSD) and the results of DFA2 and WVA2 applied to the full record. As discussed above the scales up to $\sim$ 10 years are not representative for the temperature, and this is seen as a cross-over in the slope of both DFA2 and WVA2 fluctuation functions. The deviation from a straight line at the largest scales (lowest frequencies), which is most prominent in the WVA2 fluctuation function, is caused by a nonlinear trend associated with the

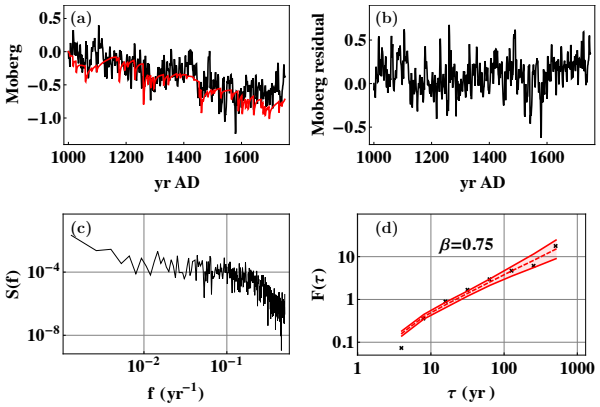


Fig. 3. (a) The Moberg reconstructed temperature (black) and the determinstic response (red). (b) The residual from the deterministic response. (c) PSD and (d) DFA2 applied to the residual. The red area is the 95% confidence for synthetic fGn with β estimated with MLE using the response model, indicated in the figure. Estimating β from the residual with DFA2 yields $\beta = 0.589$.

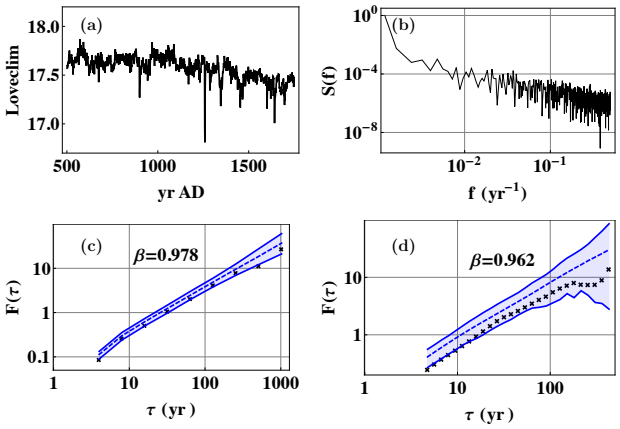


Fig. 4. (a) The temperature from the LOVECLIM experiment. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

two well known climatic features of the last 2000 years: Medieval climate anomaly and the Little ice age. The two methods yield $\beta \approx 0.69$ and $\beta \approx 0.60$, respectively.

In principle the LRM properties due to internal dynamics can be separated from those induced by the external forcing by applying the response model method of Rypdal and Rypdal (2013) described in section 2. This method allows estimation of the model parameters β, σ, and μ from the Crowley forcing data and the Moberg reconstruction record. Then we can compute the deterministic response and the residual obtained by subtracting this deterministic response from the Moberg record. The residual represents the response to the stochastic forcing and hence the internal variability of the climate system. The scaling properties of this residual can be assessed with the DFA method which also provides a consistency test on the maximum-likelihood estimate of β.

A caveat here is the low-pass filtered nature of the Moberg record. The MLE method tends to emphasize the shorter scales on which the reconstruction record is smooth, and this will spuriously yield $\beta \approx 1$. A way to avoid this could be to coarse grain both temperature and forcing time series by averaging over successive time windows of a certain length t_A, such that the temperature series is no longer smooth. This will give a more reasonable maximum-likelihood estimate of β, but the coarse-grained data cannot capture the causal connection between the almost instantaneous volcanic forcing spikes and the temperature response to them. The resulting blurring of the causal connection on time scales shorter than a decade has the effect that the MLE method interprets the variability on these short scales as stochastic, and hence overestimates the stochastic forcing strength σ, and also yields incorrect estimates of μ and β. The lesson to learn from this is that we cannot expect to obtain a correct separation of deterministic and stochastic forcing and correct parameter estimates from the low-resolution reconstruction data. Another approach to circumvent this problem was suggested in Rypdal and Rypdal (2013), where the response model parameters computed from the annual-resolution instrumental data were applied to the millennium-long annual-resolution forcing record to produce a deterministic-response record with annual resolution. The Moberg record and this deterministic response is shown in Figure 3(a). The residual obtained by subtracting the deterministic response from the reconstructed record is shown in Figure 3(b), and provides a good representation of the internal variability on time-scales longer than a decade. On shorter time-scales the residual is strongly influenced by the forced response due to the smooth character of the temperature reconstruction, but we do not need to use those scales to estimate model parameters if we do not insist on using MLE. The PSD of the residual is shown to exhibit good scaling in Figure 3(c), and the DFA2 fluctuation function of this residual on the longer time-scales should provide good estimates of β for the internal variability, as shown in Figure 3(d).

4.2 Results from LOVECLIM experiment

The NH temperature record for the period 1000-1750 AD for the LOVECLIM experiment, its power spectral density (PSD), and the DFA2 and WVA2 fluctuation functions are shown in Figure 4. DFA2 and WVA2 show good scaling with $\beta \approx 0.97$ at least on time scales up to a few hundred yr. The response model gives similar value of β, which suggests that the persistence observed in the modeled record is

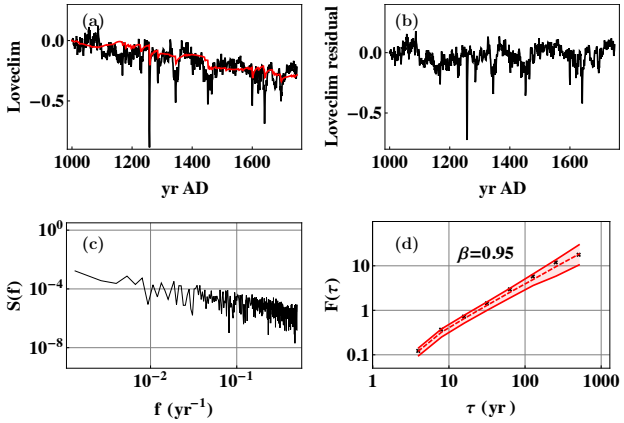


Fig. 5. (a) Temperature from the LOVECLIM experiment with 1-year resolution (black) and deterministic response (red). (b) The residual from the deterministic response. (c) PSD and (d) DFA2 applied to the residual. The red area is the 95% confidence for synthetic fGn with β estimated with MLE using the response model, indicated in the figure. Estimating β from the residual with DFA2 yields $\beta = 1.01$.

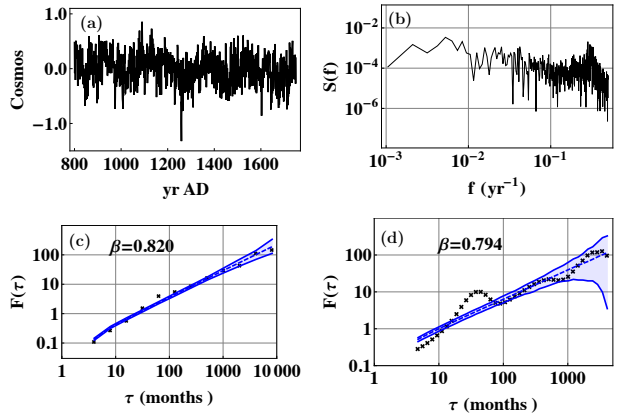


Fig. 6. (a) The temperature from the COSMOS experiment. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

due to LRM in the internal dynamics and not a reflection of LRM in the forcing. In this model simulation both forcing input and simulation output are given with annual resolution. This allows us to handle volcanic forcing and the response to volcanic eruptions in a realistic manner. The results from the response model estimates with annual resolution are shown in Figure 5.

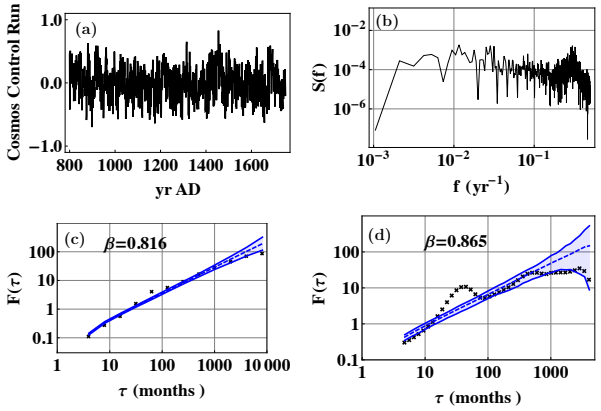


Fig. 7. (a) The temperature from the COSMOS controlrun. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

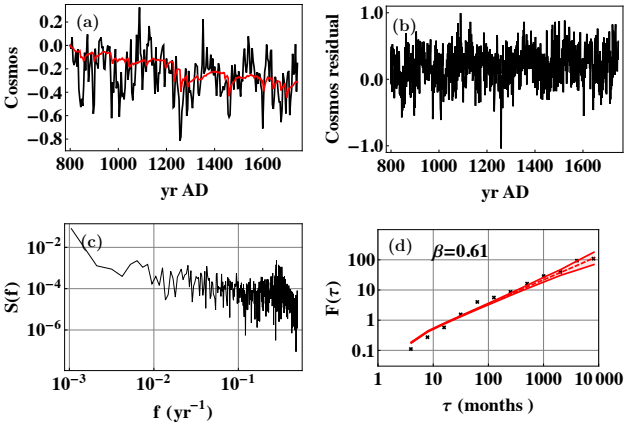


Fig. 8. (a) 4-year average of the temperature from the COSMOS experiment (black) and deterministic response (red). (b) The residual from the deterministic response. (c) PSD and (d) DFA2 applied to the residual. The red area is the 95% confidence for synthetic fGn with β estimated with MLE using the response model, indicated in the figure. Estimating β from the residual with DFA2 yields $\beta = 0.772$.

4.3 Results from COSMOS experiment

The temperature from the COSMOS forced run experiments exhibits some oscillations. In particular a prominent peak corresponding to a multiannual mode is seen in the PSD in Figure 6(b), and in the WVA plot in Figure 6(d). Otherwise scaling is fairly good with $\beta \approx 0.8$. For the control run (Figure 7) β is almost the same with DFA2, but slightly higher with WVA2. The PSD, DFA2, and WVA2 show some signs of loss of memory on scales longer than a century. The mul-

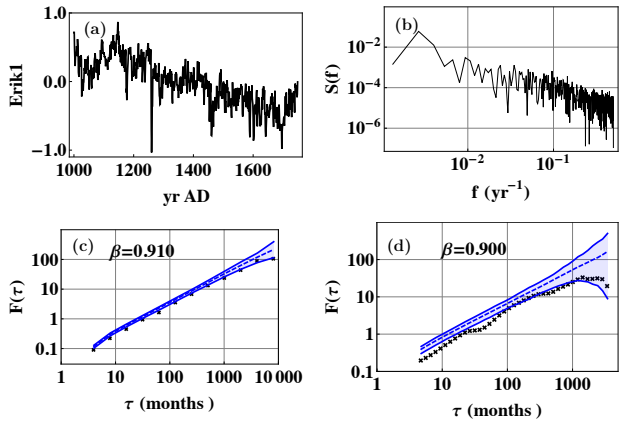


Fig. 9. (a) The temperature from the ECHO-G experiment Erik1. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

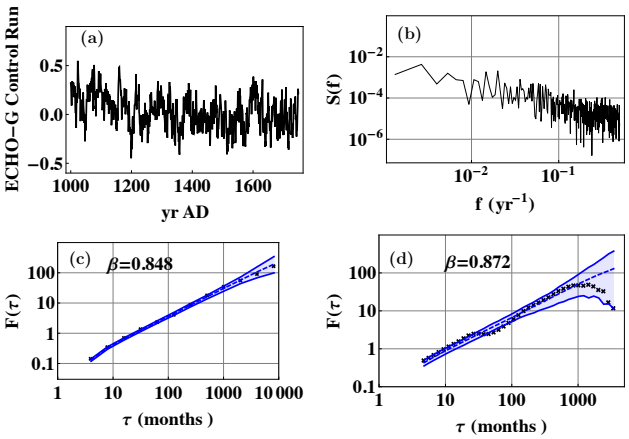


Fig. 10. (a) The temperature from the ECHO-G controlrun. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. The blue areas are the 95% confidence for synthetic fGn with β estimated with DFA2 and WVA2, indicated in the figure.

tiannual oscillation influences the maximum-likelihood estimation of model parameters in the response model, so in Figure 8 these estimates have been performed on a 4-yr coarse-grained time series, while the residual has been computed on monthly scale. As discussed above such coarse-graining creates a misrepresentation of the response to volcanic eruptions, which we believe is the main reason for the lower estimate $\beta \approx 0.6$ from the response model.

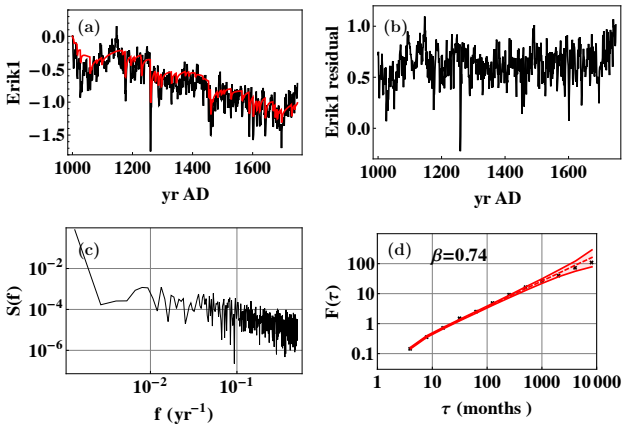


Fig. 11. (a) 1-year average of the temperature from the experiment Erik1 (black) and deterministic response (red). (b) The residual from the deterministic response. (c) PSD and (d) DFA2 applied to the residual. The red area is the 95% confidence for synthetic fGn with β estimated with MLE using the response model, indicated in the figure. Estimating β from the residual with DFA2 yields $\beta = 0.720$.

4.4 Results from ECHO-G experiments

The temperature from the forced experiment "Erik1" shows good scaling with $\beta \approx 0.9$ in Figure 9. The temperature from the control run (Figure 10) also scales well with a similar value for β. The response model yields a slightly smaller β (Figure 11). Here a 1-yr coarse graining has been applied before the parameters have been estimated, since the forcing data have 1-yr resolution.

4.5 Results from the HadCM3 experiment

The HadCM3 experiment consists of only a control run, and the scaling properties of the NH temperature series from this experiment differs from the other experiments in some important respects. Figure 12 shows a marked cross-over between two scaling regimes for $\tau \sim 100$ months in the DFA2 plot and $\tau \sim 45$ months in the WVA2 plot. The two regimes correspond to fBm-scaling with $\beta \approx 2$ (Brownian motion) for the smaller scales, and fGn-scaling with $\beta \approx 0.6$ for the larger scales. Rather than being dominated by an ENSO-like quasi-oscillatory mode up to scales of a few years, as observed in the COSMOS experiments, we observe here a non-stationary random-walk-like process for those scales. On the longer time scales this model also exhibits persistent scaling, but with somewhat lower persistence than observed in data from control runs and response model estimates in the other climate model experiments we have investigated.

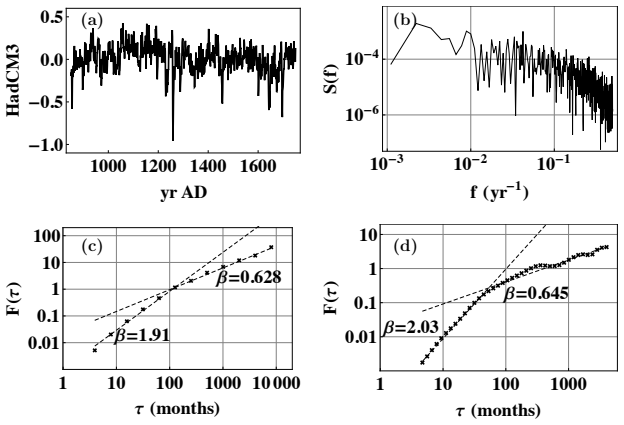


Fig. 12. (a) The temperature from the HadCM3 experiment. (b) PSD, (c) DFA2 and (d) WVA2 applied to the record. Two scaling regimes are found with β estimated with DFA2 and WVA2, indicated in the figure.

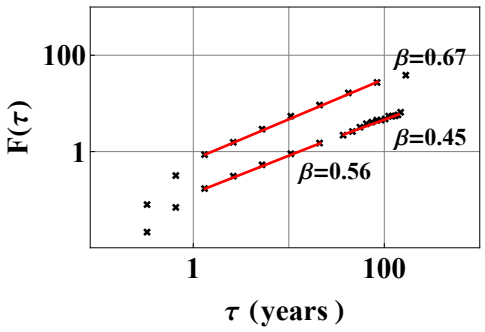


Fig. 13. DFA2 applied to sea surface or air surface temperature at Reykjanes Ridge. The upper curve is the result for the air surface temperature from Erik1 experiment. The lower left curve is based on the monthly reconstructed data for sea surface temperature, and the lower right curve on the marine sediment reconstruction of ocean temperature. The red lines indicate the scales used to estimate β.

4.6 Scaling in local data; Reykjanes Ridge

Analysis of instrumental local station data from continental interiors typically yields very low persistence on time scales up to a few decades. On the other hand, our experience is that coastal and oceanic observations in the temperate regions of the Northern Hemisphere present persistent β-values closer to those found for the hemispheric average. We also believe that this feature extends beyond the decadal scales, i.e., that good scaling with strong persistence prevails on scales up to several centuries in the Northern oceans. As an illustration we present in Figure 13 analysis of the Reykjanes ridge reconstruction from marine sediments described in Section 3.2 and the ECHO-G Erik1 experiment for the period 1000-1750 AD, and of the monthly SST reconstruction for the period 1880-1997 as described in Section 3.3. The figure shows DFA2 plots for the three data sets. ECHO-G shows good scaling in the range 1-100 yr with $\beta \approx 0.67$, as compared to $\beta = 0.91$ for the NH-average. The marine sediment reconstruction yields $\beta \approx 0.45$, and the instrumentally-based reconstruction $\beta \approx 0.56$. The greatest uncertainty in the β-estimate is in the marine sediment reconstruction, for which a very limited range of scales is available for analysis. The record has uneven time spacing, but the time step is mostly almost the same, slightly below 10 years. The data points inconsistent with this are ignored, and DFA2 applied to the remaining record. A maximum-likelihood estimation method for time series with uneven time spacing yields β close to what was found with DFA2. In spite of these uncertainties the analysis demonstrates consistently persistent scaling over time scales from years to centuries in these local data from model experiment and reconstruction data.

5 Conclusions

The temperatures from all model experiments except HadCM3 a yield higher values of β than the Moberg reconstruction when scales longer than a decade are considered. The reconstruction is said to represent temperature in the Northern Hemisphere, but most of the proxies used are in land and coastal areas. In this sense they may be considered more like representations of land or coastal temperature. Studies of observational data show higher persistence in sea surface temperature than land air temperature, and the value for β found for the Moberg record is more in agreement with the one found for the Northern Hemisphere land temperature than ocean temperature (Eichner et al., 2003; Lennartz and Bunde, 2009). Our estimate is in agreement with Rybski et al. (2006). The temperature from the model experiments is averaged over grid cells from both land and ocean areas, and the influence of the ocean might be what yields the higher value of β than found for the Moberg reconstruction.

The temperatures from the COSMOS experiments (both forced and control run) clearly show an influence of a quasi-periodic variability with a 2-3 year period, which can be associated with ENSO. The ECHO-G experiments show less influence of this oscillation, in HadCM3 it appears more like a crossover between two scaling regimes, and in LOVECLIM it is not noticeable. For the reconstructed temperature the ENSO time scales are not resolved, but in instrumental records the tropical oceans show a spectrum similar to that found from HadCM3.

The temperatures from the forced ECHO-G experiment and the LOVECLIM experiment show a more distinct Little Ice Age anomaly, in agreement with the temperature reconstruction, than the temperature from the COSMOS forced run experiment. This anomaly may also influence the estimation of β.

All the NH-averaged temperatures from forced experiments show clear persistent scaling with $0.8 < \beta < 1$ on most of the available scales, i.e., from a decade to several centuries. The control runs and the response model estimates from the forced runs, which reveal the memory properties of the internal climate dynamics, do not show systematically less persistence than obtained directly from the simulated forced temperature records. This observation does not support the notion that the observed long-range memory to great extent is generated by the forcing. Such a suggestion was made by Rybski et al. (2008), based on a global map for the parameter $\alpha = (\beta + 1)/2$ computed from both forced runs and control runs of the ECHO-G model. We believe that this discrepancy is caused by the reduction of spatiotemporal noise implied in performing an NH-average. The differences in estimated α between forced and unforced runs for local data may not be in the persistence of the underlying global signal, but rather in differences related to the amplitudes of spatiotemporal modes for the two types of runs.

Acknowledgements. For maximium-likelihood parameter estimation in the response model we have employed an R-routine developed by Ola Løvsletten.

References

Balmaceda, L., Krivova, N. A., and Solanki, S. K.: Reconstruction of solar irradiance using the group sunspot number, 2007.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R.: Long-Memory Processes: Probabilistic Properties and Statistical Methods, Springer, 2013.

Blender, R. and Fraedrich, K.: Long time memory in global warming simulations, Geophys. Res. Lett., 30, 2003.

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

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

Collins, M., Tett, S. F. B., and Cooper, C.: The internal climate variability of HadCM3, a version of the Hadley Centre coupled climate model without flux adjustments, Clim. Dynam., 17, 61–81, 2000.

Crowley, T. J.: Causes of climate change over the past 1000 years, Science, 289, 270–277, 2000.

Crowley, T. J., Baum, S. K., Kim, K.-Y., Hegerl, G. C., and Hyde, W. T.: Modeling ocean heat content changes during the last millennium, Geophys. Res. Lett., 30, 2003.

Eichner, J. F., Koscielny-Bunde, E., Bunde, A., Havlin, S., and Schellnhuber, H.-J.: Power-law persistence and trends in the atmosphere: A detailed study of long temperature records, Phys. Rev. E, 68, 2003.

Flandrin, P.: Wavelet analysis and synthesis of fractional Brownian motion, Information Theory, IEEE Transactions on, 38, 910 – 917, 1992.

Fraedrich, K. and Blender, R.: Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models, Phys. Rev. Lett., 90, 108 501, 2003.

González-Rouco, F., von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, Geophys. Res. Lett., 30, 2003.

Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Huybrechts, P., Jongma, J., Mouchet, A., Selten, F., Barriat, P.-Y., Campin, J.-M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M.-F., Morales Maqueda, M. A., Opsteegh, T., Mathieu, P.-P., Munhoven, G., Pettersson, E. J., Renssen, H., Roche, D. M., Schaeffer, M., Tartinville, B., Timmermann, A., and Weber, S. L.: Description of the Earth system model of intermediate complexity LOVECLIM version 1.2, Geoscientific Model Development Discussions, 3, 309–390, 2010.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M., Renssen, H., Sallaz-Damaz, Y., and Shindell, D.: The role of forcing and internal dynamics in explaining the Medieval Climate Anomaly, Clim. Dynam., 39, 2847–2866, 2012.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, Clim. Dynam., 16, 147–168, 2000.

Govindan, R. B., Vjushin, D., Brenner, S., Bunde, A., Havlin, S., and Schellnhuber, H.-J.: Long-range correlation and trends in global cliamte models: Comparison with real data, Physica A, 294, 239–248, 2001.

Govindan, R. B., Vjushin, D., Bunde, A., Brenner, S., Havlin, S., and Schellnhuber, H.-J.: Global climate Models Violate Scaling of the Observed Atmospheric Variability, Phys. Rev. Lett., 89, 2002.

Jungclaus, J. H., Lorenz, S. J., Timmreck, C., Reick, C. H., Brovkin, V., Six, K., Segschneider, J., Giorgetta, M. A., Crowley, T. J., Pongratz, J., Krivova, N. A., Vieira, L. E., Solanki, S. K., Klocke, D., Botzet, M., Esch, M., Gayler, V., Haak, H., Raddatz, T. J., Roeckner, E., Schnur, R., Widmann, H., Claussen, M., Stevens, B., and Marotzke, J.: Climate and carbon-cycle variability over the last millennium, Climate of the Past Discussions, 6, 1009–1044, 2010.

Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H., Havlin, S., and Bunde, A.: Detecting long-range correlations with detrended fluctuation analysis, Physica A, 295, 441 – 454, 2001.

Krivova, N. A. and Solanki, S. K.: Reconstruction of total solar irradiance since 1700 from the surface magnetic flux, Astron. Astrophys., 467, 335–346, 2007.

Krivova, N. A. and Solanki, S. K.: Models of solar irradiance variations: Current status, J. Astrophys. Astron., 29, 151–158, 2008.

Legutke, S. and Maier-Reime, E.: Climatology of the HOPE-G Global Ocean - Sea Ice General Circulation Model, DKRZ Tech. Rep.. DKRZ-TR–21, 1999.

Legutke, S. and Voss, R.: The Hamburg Atmosphere-Ocean Coupled Circulation Model ECHO-G, DKRZ Tech. Rep.. DKRZ-TR–18, 1999.

Lennartz, S. and Bunde, A.: Trend evaluation in records with long-term memory: Application to global warming, Geophys. Res. Lett., 36, 2009.

Malamud, B. D. and Turcotte, D. L.: Self-affine time series: measures of weak and strong persistence, J. Statistical Planning and Inference, 80, 173–196, 1999.

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, Science, 326, 1256–1260, 2009.

Marland, G., Boden, T. A., Andres, R. J., Brenkert, A., and Johnston, C.: Global, regional, and national fossil fuel CO2 emissions, Trends: A compendium of data on global change, pp. 34–43, 2003.

Marsland, S., Haak, H., Jungclaus, J., Latif, M., and Röske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, Ocean Modelling, 5, 91 – 127, 2003.

Miettinen, A., Divine, D., Ko, N., Godtliebsen, F., and Hall, I. R.: Multicentennial Variability of the Sea Surface Temperature Gradient across the Subpolar North Atlantic over the Last 2.8 kyr, J. Climate, 25, 4205–4219, 2012.

Mills, T. C.: Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends?, J. Roy. Stat. Soc. A Sta., 170, 83–94, 2007.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data, Nature, 433, 613–617, 2005.

Muscheler, R., Joos, F., Beer, J., Müller, S. A., Vonmoos, M., and Snowball, I.: Solar activity during the last 1000yr inferred from radionuclide records, Quaternary Science Reviews, 26, 82–97, 2007.

Pelletier, J. D.: Analysis and Modeling of the Natural Variability of Climate, J. Climate, 10, 1331 – 1342, 1997.

Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A.: Mosaic organization of DNA nucleotides, Phys. Rev. E., 49, 1685–1689, 1994.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A.: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, Clim. Dynam., 16, 123–146, 2000.

Raddatz, T., Reick, C., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.-G., Wetzel, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, Clim. Dynam., 29, 565–574, 2007.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, J. Climate, 15, 1609–1625, 2002.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U., and Schulzweida, U.: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate, MPI Tech. Rep. Rep. 218, 1996.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U., and Schulzweida, U.: The atmospheric general circulation model ECHAM-5, part 1: Model description, MPI Tech. Rep. Rep. 349, 2003.

Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, Geophys. Res. Lett., 33, 2006.

Rybski, D., Bunde, A., and von Storch, H.: Long-term memory in 1000-year simulated temperature records, J. Geophys. Res., 113, 2008.

Rypdal, M. and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, Submitted to J. Climate, 2013.

Smith, T. M. and Reynolds, R. W.: Improved Extended Reconstruction of SST (1854–1997)., J. Climate, 17, 2004.

Smith, T. M. and Reynolds, R. W.: A Global Merged Land–Air–Sea Surface Temperature Reconstruction Based on Historical Observations (1880–1997)., J. Climate, 18, 2005.

Solanki, S. K., Usoskin, I. G., Kromer, B., Schuessler, M., and Beer, J.: Unusual activity of the Sun during recent decades compared to the previous 11 000 years, Nature, 431, 1084–1087, 2004.

Syroka, J. and Toumi, R.: Scaling and persistence in observed and modeled surface temperature, Geophys. Res. Lett., 28, 3255–3258, 2001.

ter Braak, C. J. and Juggins, S.: Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages, Hydrobiologia, 269, 485–502, 1993.

Usoskin, I. G., Solanki, S. K., and Kovaltsov, G. A.: Grand minima of solar activity during the last millennia, Proceedings of the International Astronomical Union, 7, 372–382, 2011.

von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F., and Tett, S. F. B.: Reconstructing Past Climate from Noisy Data, Science, 306, 679–682, 2004.

Vyushin, D., Zhidkov, I., Havlin, S., Bunde, A., and Brenner, S.: Volcanic forcing improves Atmosphere-Ocean Coupled General Circulation Model scaling performance, Geophys. Res. Lett., 31, 2004.

Weber, R. O. and Talkner, P.: Spectra and correlations of climate data from days to decades, J. Geophys. Res., 106, 20,131–20,144, 2001.

Wetzel, P., Maier-Reimer, E., Botzet, M., Jungclaus, J., Keenlyside, N., and Latif, M.: Effects of ocean biology on the penetrative radiation in a coupled climate model, J. Climate, 19, 3973–3987, 2006.

Zorita, E., González-Rouco, F., and Legutke, S.: Testing the Mann et al.(1998) Approach to Paleoclimate Reconstructions in the Context of a 1000-Yr Control Simulation with the ECHO-G Coupled Climate Model, J. Climate, 16, 1378–1390, 2003.