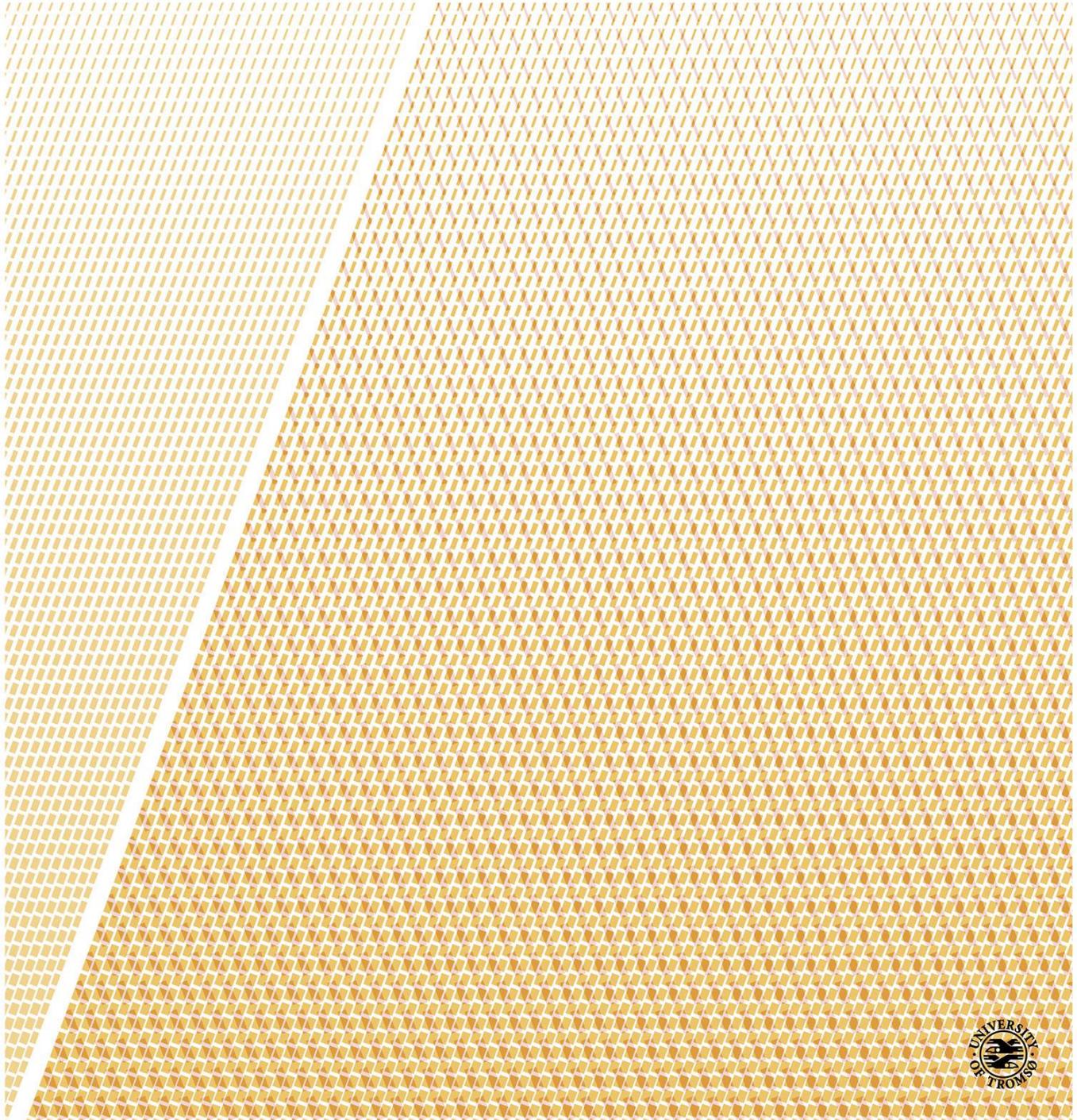


Long Range Memory in Time Series of Earth Surface Temperature

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

Abstract	i
Acknowledgements	iii
Table of Contents	v
1 Introduction	1
1.1 Motivation	1
1.2 Thesis Outline	2
1.3 List of Publications	2
2 Short- and Long-Range Memory	5
2.1 Short-Range Memory	5
2.2 Long-Range Memory	7
3 Trends	9
3.1 Trends and noise	9
3.2 Trend estimation	10
3.3 Response Model	13
4 Methods	15
4.1 Short-Range Memory	15
4.2 Long-Range Memory	16
4.3 Comparing Short-Range and Long-Range Memory Processes	24
5 Literature Review	31
5.1 Instrumental Data	31
5.2 Simulated Temperature From Model Experiments	37
5.3 Reconstructed Temperature	40
6 Summary of Papers	41
7 Concluding Remarks	43

Bibliography	45
Paper I	53
Paper II	73
Paper III	87

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 Wojdyło*, 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 CO₂ 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 $\int_0^\infty \rho(t)dt$ 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*)

$$x(t) = \phi x(t-1) + \sigma w(t), \quad t = 1, 2, \dots, \quad (2.1)$$

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

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

AR(p) has an autocorrelation function

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

and power spectral density (PSD)

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

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

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

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

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

$$\tau_c = -\frac{1}{\log \phi}. \quad (2.7)$$

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 Lorentzian, 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:

$$dx(t) + \frac{1}{\tau_c}x(t)dt = \sigma dB(t), \quad (2.8)$$

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 $t \rightarrow \infty$, with $0 \leq \gamma < 1$. In this range of γ we have that

$$\int_0^{\infty} \rho(t) dt = \infty. \quad (2.9)$$

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

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

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

$$H = 1 - \frac{\gamma}{2} = \frac{\beta + 1}{2}. \quad (2.11)$$

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

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

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

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

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

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

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

$$\phi(B)x(t) = \psi(B)w(t), \quad (2.14)$$

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 d th difference $(1 - B)^d X_t$,

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

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

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

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

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

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 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

$$x(t) = T(A;t) + \sigma w(t), \quad (3.1)$$

where $T(A;t)$ is the trend model with parameters $A = (A_1, \dots, A_m)$ 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:

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

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 $\theta = (\theta_1, \dots, \theta_n)$ 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., $x(t) = T(A;t) + \epsilon(\theta;t)$. 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.

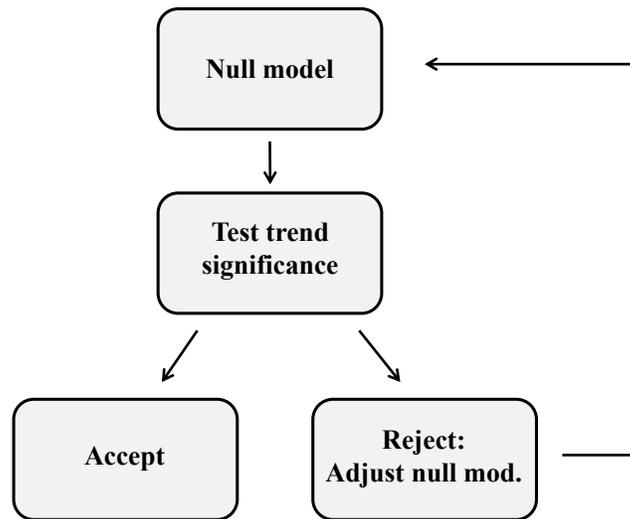


Figure 3.1: Flowchart of hypothesis testing

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

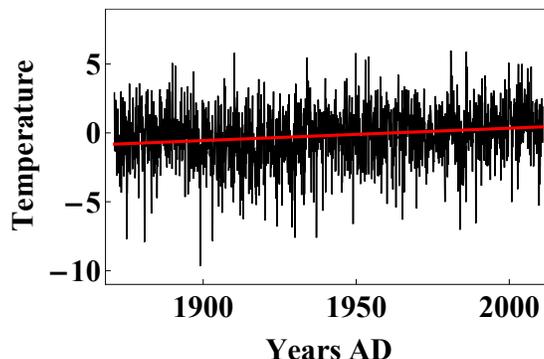


Figure 3.2: Temperature anomaly at Cheyenne. The red line is the linear trend fitted to the record.

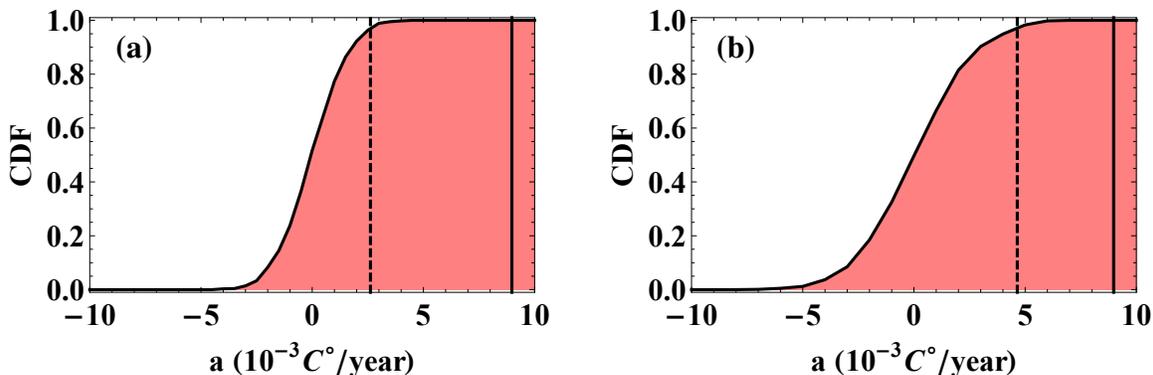


Figure 3.3: Significance of trends under (a) AR(1) null hypothesis and (b) fGn null hypothesis. Ensembles of synthetic noises are produced with the same parameters as those estimated under each hypothesis by maximum likelihood estimation (MLE). The pseudotrend parameters are then estimated by least-square 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),

$$\frac{dQ(t)}{dt} + \frac{1}{S_{\text{eq}}}T(t) = F(t). \quad (3.3)$$

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 ($dQ = CdT$) and the time constant $\tau_c = CS_{\text{eq}}$, this can be rewritten to

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

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,

$$T(t) = \int_{-\infty}^t G(t-s)F(s)ds. \quad (3.5)$$

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

$$T(t) = \underbrace{\int_{-\infty}^t G(t-s)F(s)ds}_{\text{deterministic response}} + \underbrace{\sigma \int_{-\infty}^t G(t-s)dB(s)}_{\text{stochastic response}}, \quad (3.6)$$

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 $\xi \equiv 1 \text{ Km}^2/\text{J}$ 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.

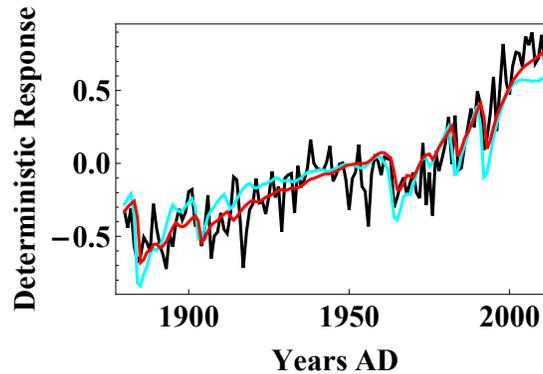


Figure 3.4: Yearly averaged global land temperature (black), and deterministic response using the exponential response model (cyan) and scale-free response model (red).

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);

$$\delta(t) = x(t) - \phi x(t - 1) = \sigma w(t), \quad (4.1)$$

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,

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

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

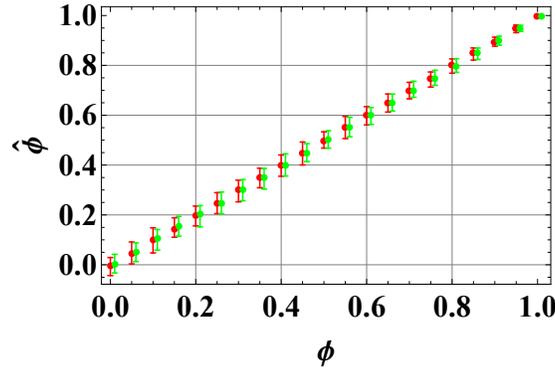


Figure 4.1: Bias and error bars for ACF (red) and MLE (green).

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 Wojdyło, 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 too 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

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

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*),

$$\gamma[k] \sim k^{2H} \quad (4.5)$$

Periodogram

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

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

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 $f = m/N$, $m = 1, 2, \dots, N/2$. For these frequencies, a self-affine time series scales as a power-law (e.g., *Voss, 1986*)

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

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

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

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

$$\int_{-\infty}^{\infty} \Psi(t') dt' = 0. \quad (4.9)$$

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*).

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

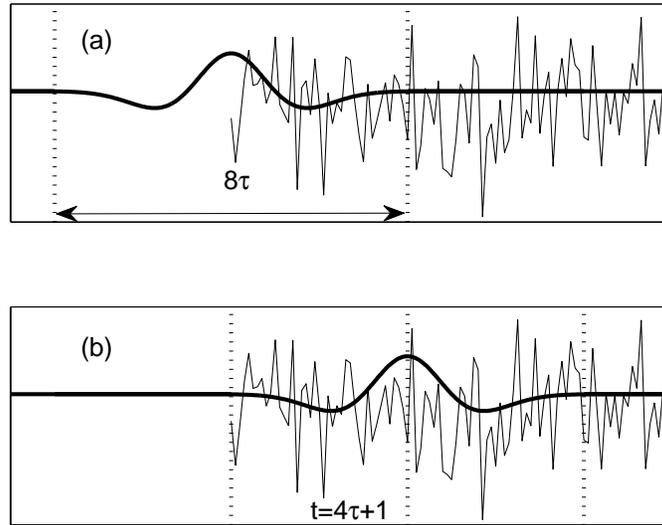


Figure 4.2: The time series is covered by the wavelet (a) for only the last half of the time when the wavelet is centred in $t = 1$ and (b) at all times when it is centred in $t = 4\tau + 1$.

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 n th derivative of Gaussian wavelet,

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

and use the notation WVA_n 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,

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

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 n th order polynomial is computed and subtracted for each segment,

$$Y_\tau(i) = Y(i) - p_\nu(i), \quad (4.13)$$

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

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

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

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

The scaling exponent is defined by the relation

$$F(\tau) \propto \tau^\alpha. \quad (4.16)$$

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, $F_{\text{FA}}^2(\nu, \tau) = [Y(\nu\tau) - Y((\nu - 1)\tau)]^2$, 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

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

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

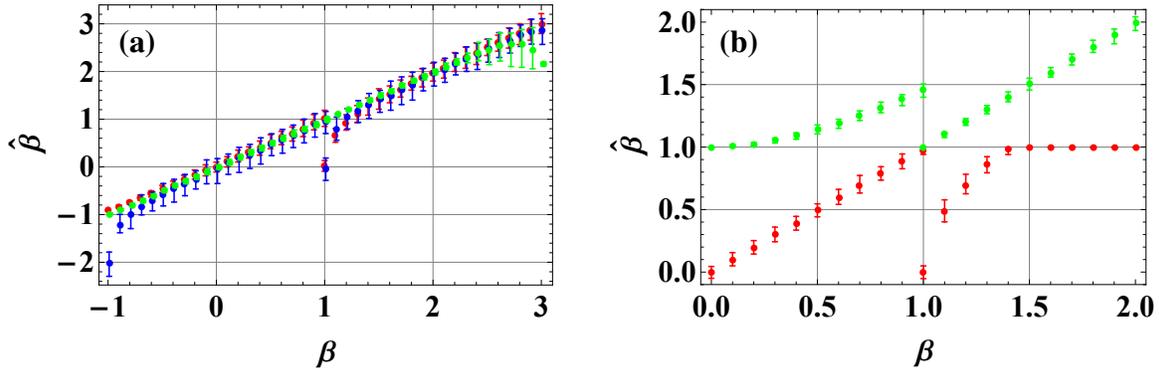


Figure 4.3: (a) Bias and error bars for DFA2 (red), WVA2 (blue) and MLE (green). (b) Bias and error bars for MLE. fGn and fBm with different β analysed as fGn (red) and fBm (green). analysing as fBm means that the increments are analysed as fGn and $\hat{\beta} = \hat{\beta}_{\text{incr}} + 2$.

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 $\beta \rightarrow 1^+$, and WVA2 has a similar bias when $\beta \rightarrow -1^+$ 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, $\hat{\beta} \rightarrow 1$ as β increases, but smaller values are found when $\beta \rightarrow 1^+$. When the increments of fGns are studied, $\hat{\beta} \rightarrow 1$ as β decreases, but larger values are found when $\beta \rightarrow 1^-$.

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

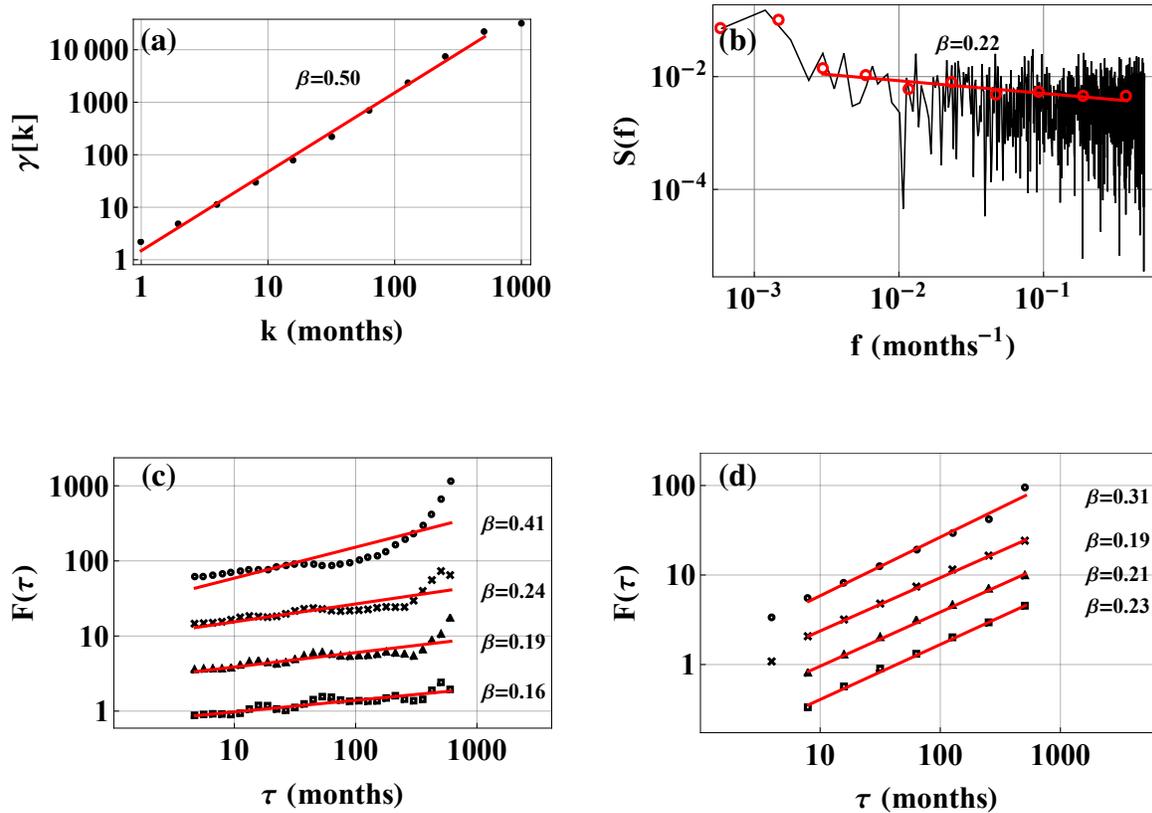


Figure 4.4: Temperature anomaly at Cheyenne analysed with (a) semivariogram, (b) periodogram, (c) WVA of order 1-4 from top to bottom and (d) DFA of order 1-4 from top to bottom. The red lines indicate the scaling range used to estimate β . The red circles in panel (b) are the result of log-binning.

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 .

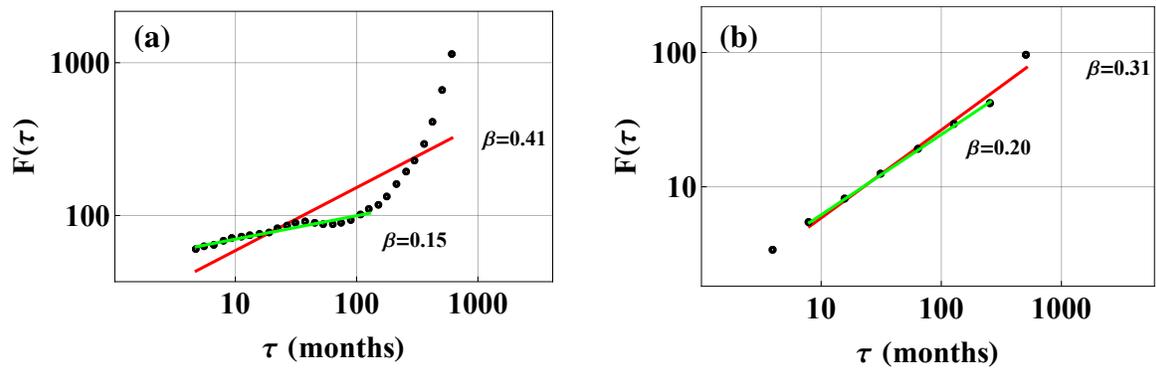


Figure 4.5: Temperature anomaly at Cheyenne analysed with (a) WVA1 and (b) DFA1. The red lines indicate the scaling region when all scales up to about $1/10$ of the record length are used to estimate β . The green lines show the scaling region when only the scales up to the assumed cross-over scale where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

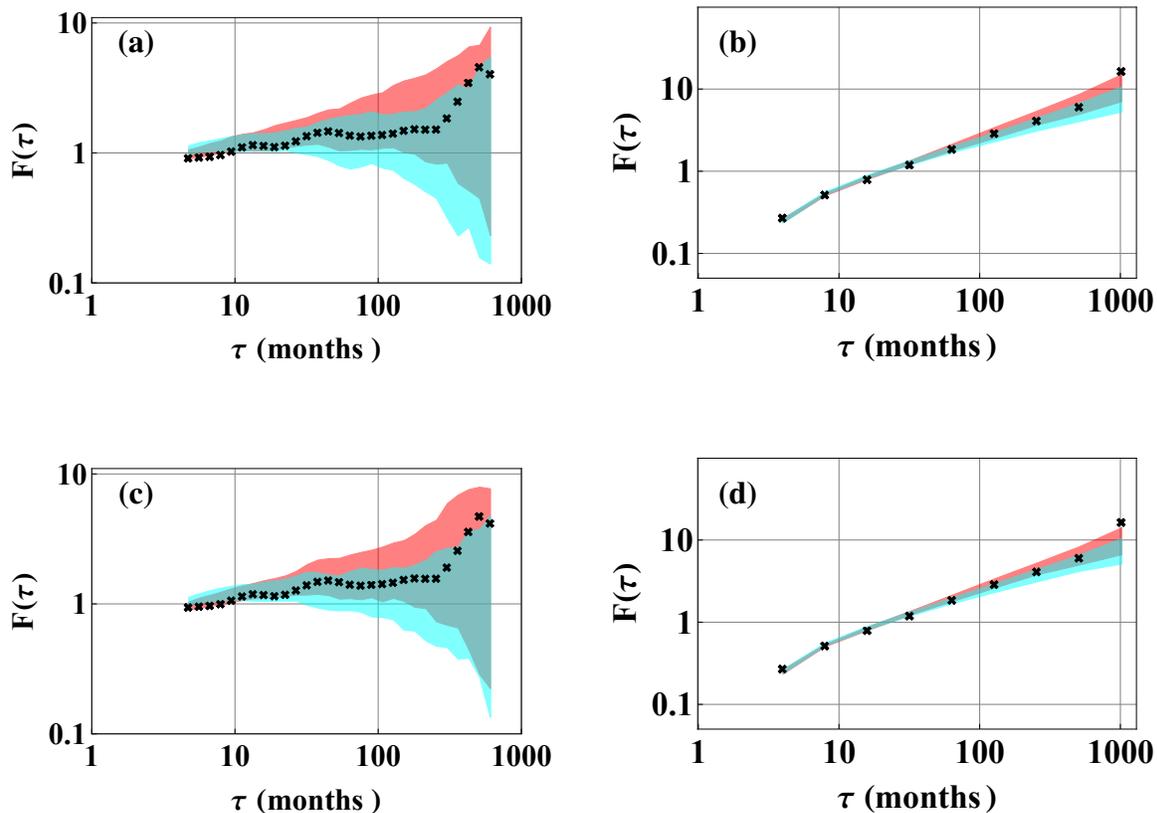


Figure 4.6: (a) WVA2 and (b) DFA2 applied to the temperature anomaly at Cheyenne (black crosses). Ensembles are generated of synthetic realizations of two different stochastic processes: An AR(1) process (cyan) and fGns (red). The synthetic processes are generated with parameters estimated from the observed 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

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

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

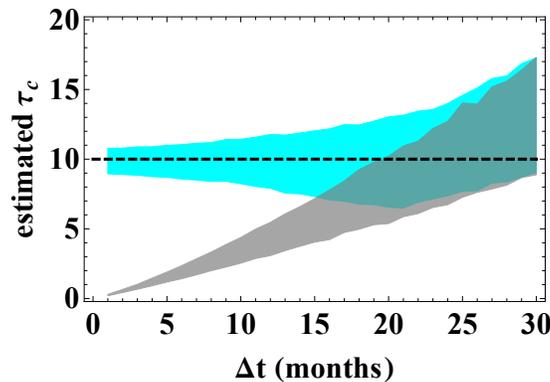


Figure 4.7: Estimated τ_c as a function of Δt for an Ornstein-Uhlenbeck process with $\tau = 10$ (cyan) and white noise (grey).

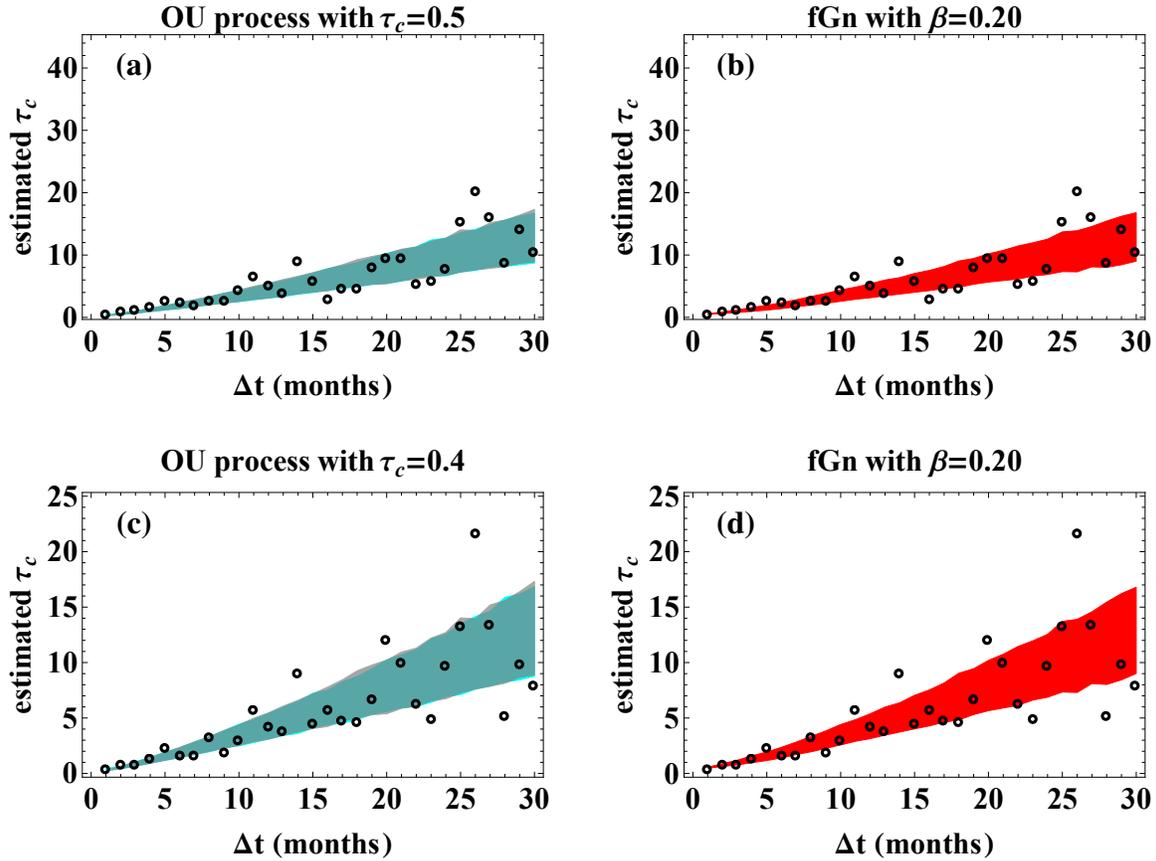


Figure 4.8: Panel (a) and (b) shows the estimated decorrelation time τ_c as a function of Δt for the temperature anomaly at Cheyenne as black circles. Ensembles are generated of synthetic realizations of two different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) in panel (b). 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

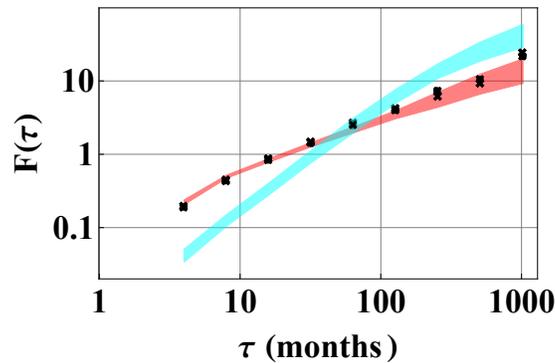


Figure 4.9: DFA2 applied to the residual from the deterministic response of the exponential response model (black circles) and scale-free response model (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn.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

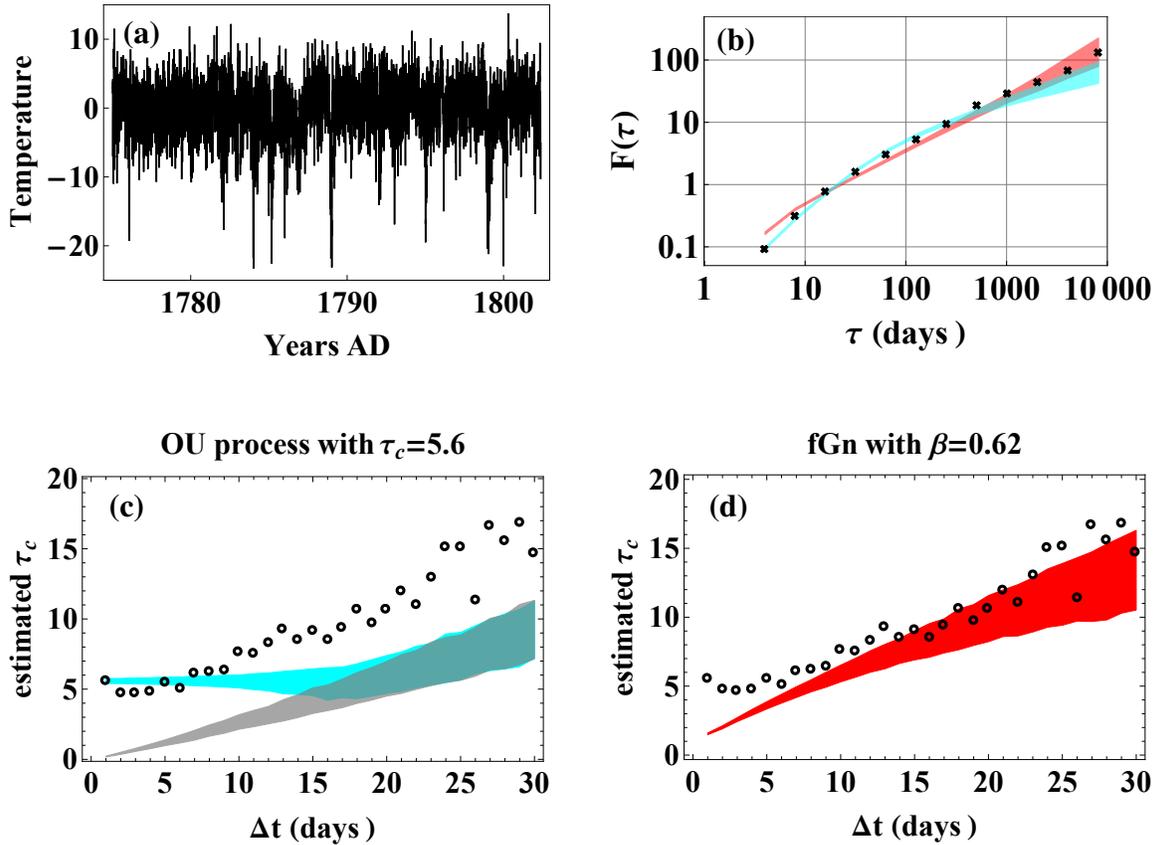


Figure 4.10: (a) 10 000 data points (~ 28 years) for Prague, daily temperature. (b) DFA2 applied to the record (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn. Panel (c) and (d) shows the estimated decorrelation 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

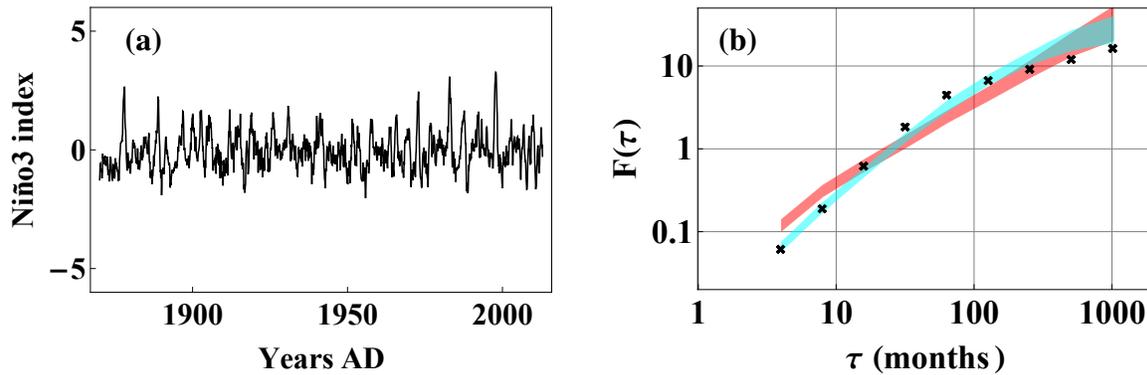


Figure 4.11: (a) The deseasonalized Niño3 index. (b) DFA2 applied to the record (black crosses). The cyan area is the 95% confidence for realizations of AR(1) and the red area is the 95% confidence area for realizations of fGn. For AR(1), $\tau_c = 11.8$ was found with MLE, while $\beta = 0.62$ was found 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 based on the isotopic fractions between ^{18}O and ^2H 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 Eurasian 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, CO₂ 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 János* (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 $\delta^{18}\text{O}$ 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 CO₂ 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 CSIROCM2 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 CO₂ 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 CO₂ makes the temperature from the model experiments lose their memory, and that the results may be improved by changing the method for adding CO₂.

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 CO₂ 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](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](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](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 climate 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. Kugiuntzis. 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. Wojdyło. 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](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](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](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](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](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.

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*.

Paper III

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

Submitted to *Earth System Dynamics Discussions*.

