UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and technology

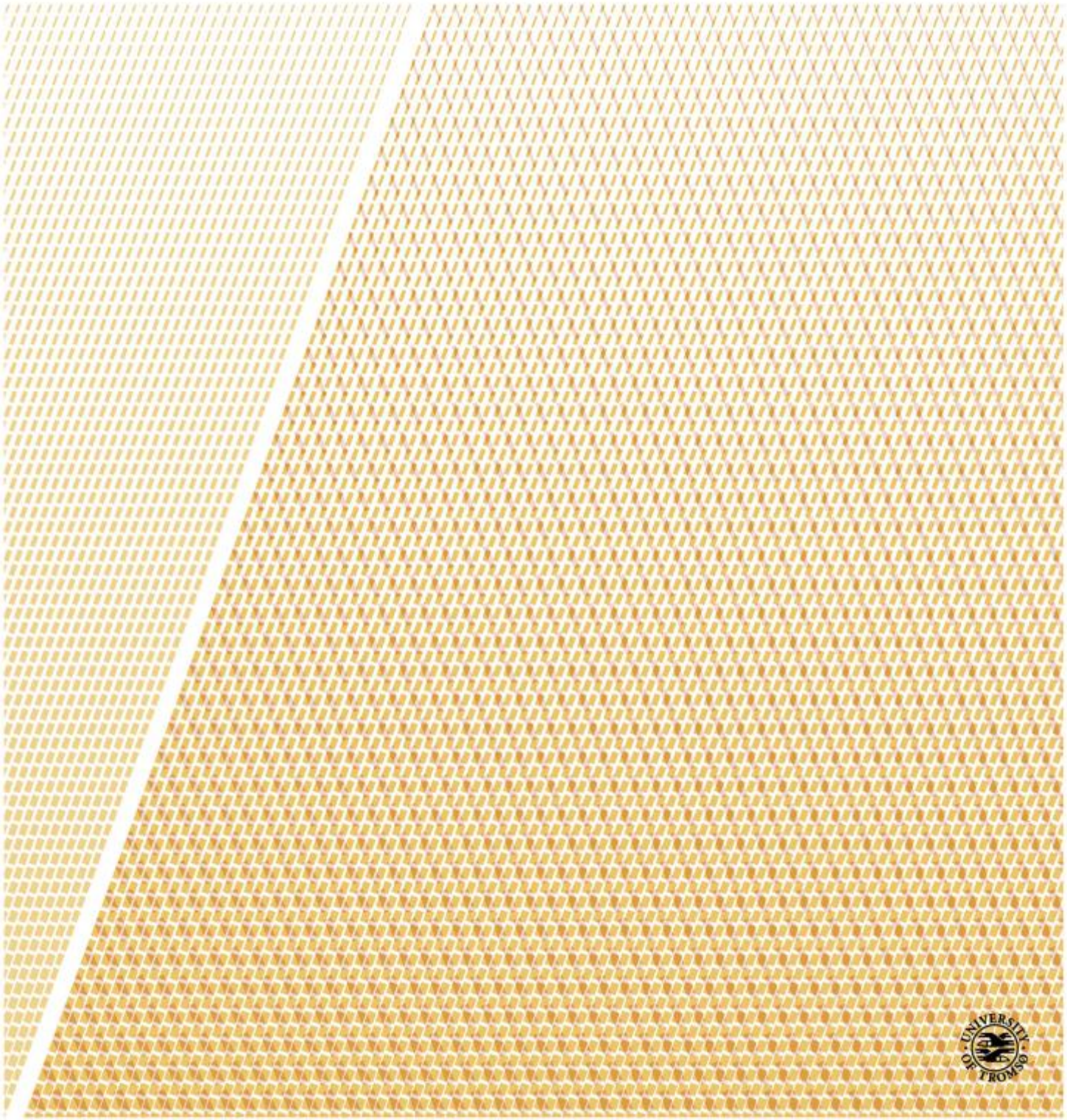
Department of Mathematics and Statistics

Long-range memory in Earth's climate response – analysis of paleoclimatic records and climate model simulations

—

Tine Nilsen

A dissertation for the degree of Philosophiae Doctor – October 2017



Abstract

The memory in reconstructed records of Earth's surface temperature and records from paleoclimatic model simulations is investigated in this thesis using a variety of estimation methods. For the Holocene period, the analyses reveal that many local and spatially averaged paleoclimate time series exhibit long-range memory (LRM) on timescales from a few years to centuries or millennia, with a positive spectral exponent β predominantly less than unity, corresponding to a fractional noise process. In both local and large-scale temperature reconstructions, the memory properties are likely affected by the choice of reconstruction methods, proxy records and how the reconstruction is processed to obtain even temporal resolution.

Furthermore, for ice core records extending beyond the Holocene, a second scaling regime is identified. However, the concept of using only second order statistics to describe these long records is critically examined, since the variability associated with the Dansgaard-Oeschger events and the deglaciation cannot be sufficiently described by a simple fractional noise.

Finally, as the proxy noise and reconstruction techniques may alter the memory properties inherent to temperature proxies, the ability of one selected reconstruction technique to preserve LRM-properties of proxy data has been assessed through pseudoproxy experiments. Analyses demonstrate that for the ensemble mean, LRM is best preserved in the field and spatial mean reconstructions if the input data are noise-free or weakly perturbed by noise. However, the credibility of the confidence ranges of the reconstructions are more representative of the input data for high noise levels. The reconstruction skill is found to decrease with increasing noise-levels of the input data, but is virtually insensitive to the strength of the persistence.

Acknowledgements

I would like to thank UiT - the Arctic University of Norway for giving me the opportunity to take a PhD. The position was funded by the Norwegian Research Council under grant nr. 229754 (the KLIMAFORSK programme). I am very grateful for the warm welcome and friendly work environment at the Department of Mathematics and Statistics.

Thank you Kristoffer, Dima and Martin for giving me support and supervision, and for being available at almost all possible times. Your time and effort to provide scientific training is greatly appreciated. I would not have been able to finish this PhD without your thoughtful encouragement and strong commitment. Thank you for giving me the chance to travel to conferences, Summer Schools and Workshops around the World, and for the good company.

My time as a PhD would not have been the same if the desk next to me was empty. Thank you my near and dear friend Hege Beate, for being a great inspiration and for keeping up with my occasional frustration. Greetings and many thanks also to my other former and present office mates Lene, Ola and Rebekka for the good company, also when traveling together. Tough days have become better thanks to you, Erik and Amund.

At last, I would like to thank my family: mamma, pappa and Linn, and also the rest of my family and my friends for all support, encouragement and cheering over the time i have been working on my PhD project. Kristine and Miriam: thank you for all the fun we have had that has helped me unwind from work. I could not have finished this project without the invaluable help from all of you over the last year, when times have been extra difficult. Thank you for believing in me.

Contents

Chapter 1

Introduction

"One man's noise is another man's signal".

1.1 The concept of memory in the climate system

For the last 11700 years human civilization has developed under stable and warm climatic conditions in the interglacial period called the Holocene. During the Pleistocene (2 588 000 years before present up to the Holocene), there were many glacial periods, followed by shorter interglacials. Analyses of a number of climate proxies have revealed that the climate exhibits variations on a broad range of time scales from months up to hundreds of million years. There are different drivers of climate change operating on different time scales, but the climate state is not just a simple function of the drivers, but rather the result of a complex interaction between external forcing and internal processes in the climate system.

There are numerous ways of modeling these interacting processes of the climate system. The most complex climate models simulate the dynamics and physical processes of the atmosphere and the ocean at a range of spatial scales, described by geophysical fluid dynamic equations and parameterizations for numerous processes. Simpler models represent the energy balance of the climate system in boxes and reduced number of dimensions, reducing drastically the initially infinite number of system dimensions. The climate system is complex, and simplifying assumptions must be used for all types of models when describing the dynamics.

Another viewpoint on modeling climate variability is to describe it as a noise background with superposed trends, where the noise background is described as a stochastic process. In this thesis the main focus will be on surface temperature, but the same con-

cept can be used for other variables such as e.g. precipitation. Statistical modeling is extremely simplified compared with state-of-the-art general circulation models (GCM's), and it may seem pointless to represent Earth's surface temperature by random numbers drawn from a statistical distribution. On the other hand, the parameters of such simple models are estimated from real surface temperature data, and the models thereby represent possible realizations of a process with the same statistical properties as the true climate system. One motivation for using statistical modeling is that mean estimates such as the global mean surface temperature (GMST) can be estimated on a regular desktop computer in less than a second, giving projections of future temperature changes with similar uncertainties as those achieved from GCM simulations run on supercomputers. In order to achieve such estimates it is necessary to learn about the statistics of natural climate variability and about the nature and timescales involved for the temperature response to external forcing.

The noise background for surface temperature is not white in time, as the interacting subsystems respond on different timescales and introduce inertia in the climate system. A common assumption is that the noise is red, modeled with the autoregressive model of order one, the AR(1) process (Hasselmann, 1976), (Bindoff et al., 2013, Chapter 10). The AR(1) process is dominated by strong temporal correlations on timescales shorter than some characteristic timescale τ_c. There is no temporal dependence on longer timescales. In statistical modeling terms we may describe the AR(1) process as exhibiting short-range memory (SRM) or short term persistence/correlations. Another perception on the nature of climate variability is that the noise has correlations on *all* timescales. This noise background can be modeled as a long-range memory (LRM) stochastic process, for instance the fractional Gaussian noise (fGn).

To understand the relevance of the memory in a time series, consider the significance of a trend. This is a very common issue for climate studies, which is typically also requested by stakeholders outside the scientific community. If the noise background exhibits natural variability on all timescales, it is more difficult to establish significance of a trend than if the noise is purely white. In the following, the background noise in the climate system is assumed to represent internal variability. Superposed trends arise from the response to external forcing, but also due to certain nonlinear internal dynamics such as the El Niño Southern Oscillation (ENSO).

It is well known that the fundamental assumptions and statistical methods applied to proxy data introduce uncertainties and biases that are not applicable to instrumental data. For instance, regression-based reconstruction techniques may produce reconstructions exhibiting variance loss and a mean value bias for the preindustrial period

(Christiansen et al., 2009). Other reconstruction methods are based on fundamental assumptions about the serial correlations of the data, e.g. that they follow an AR(1) structure, (Tingley and Huybers, 2010a). The persistence in proxy-based reconstructions and paleoclimate simulations will be studied in detail in the following, emphasizing how uncertainties and biases may introduce artifacts in the variability levels of time series at different time scales.

The work in this thesis is mainly focused on late-Holocene surface temperature, but in Paper 2 the full Holocene period and the last glacial period are also considered. As a rule of thumb the temperature of interest is averaged over a larger area such as the Northern Hemisphere, but sparseness of proxy data sometimes makes it necessary to study local or regionally confined data.

1.2 Thesis outline

Chapters 2 and 3 cover the background information on the most relevant subjects necessary to comprehend the information in the three papers attached. Chapters 4-7 elaborate and discuss further details that may affect the scaling properties of paleoclimate reconstructions. Relevant literature is reviewed by topic. An important criterion for choosing the topics for discussion and literature review is that they are relevant for scaling analysis specifically of paleoclimate reconstructions and/or model simulations. An extensive introduction and discussion of LRM analysis in instrumental temperatures was presented earlier in Østvand (2014).

Chapter 2 describes the fundamental statistics we rely on for time series analysis and statistical modeling of climatic records. In particular, this chapter describes the concept of stochastic processes, self-similarity and scaling. The statistical methods used for scaling analysis are also introduced.

The theory of proxies and proxy-based temperature reconstructions is introduced in Chapter 3. Some well-known proxy archives are described, which are frequently used to reconstruct surface temperature. The radiocarbon dating method is also mentioned, and the concept of larger-scale reconstructions is briefly introduced.

Paleoclimate reconstruction techniques are discussed in depth in Chapter 4. Here, some caveats of using regression-based reconstruction techniques are pointed out, which may influence the scaling properties of temperature reconstructions. The reconstruction methods applied in Mann et al. (1998, 1999, 2008, 2009); Luterbacher et al. (2016); Werner et al. (2017) are presented, as well as the reconstruction procedure of Moberg

et al. (2005).

Chapter 5, 6 and 7 summarize the three papers included in this thesis, together with a discussion and literature review of each paper. The first and second manuscripts (Østvand et al., 2014; Nilsen et al., 2016) are published in Earth System Dynamics, where the review process is public and available online. Reading the discussion documents provides supplementary insight into the issues dealt with in the papers. In particular for Paper 2 there was a lengthy review process, and the discussion reflects the different points of view on the subject within the scientific community. A summary of the review process is included in the discussion of Chapter 6. The last manuscript (Nilsen et al., 2017) is under preparation and will be submitted to Climate of the Past.

1.3 List of publications

Papers

Paper 1

Østvand, L., T. Nilsen, K. Rypdal, D. Divine, and M. Rypdal. **Long-range memory in internal and forced dynamics of millennium-long climate model simulations**, Earth Sys. Dyn., 5, 295-308, 2014. doi:10.5194/esd-5-295-2014.

Paper 2

Nilsen, T., K. Rypdal and H.-B. Fredriksen. **Are there multiple scaling regimes in Holocene temperature records?**, Earth Sys. Dyn., 7:419-439, 2016. doi:10.5194/esd-7-419-2016.

Paper 3

Nilsen. T., J. P. Werner and D. V. Divine. **How wrong are climate field reconstruction techniques in reconstructing a climate with long-range memory?**, Manuscript in preparation, to be submitted to Climate of the Past November 2017.

Other publications and presentations

First author:

Nilsen, T., K. Rypdal, H.-B. Fredriksen, D. Divine, **Is there a break in scaling on centennial time scales in Holocene temperature records?** Poster presentation at *International Partnerships in Ice Core Sciences (IPICS) second open science conference*, Hobart, March 2016.

Nilsen, T., K. Rypdal, H.-B. Fredriksen, **Little evidence for multiple scaling regimes in Holocene surface temperatures**. Oral presentation at conference on *Scales and scaling in the climate system: bridging theory, climate models and data'*, Jouvence, October 2015.

Nilsen, T., K. Rypdal, H.-B. Fredriksen, M. Rypdal, O. L øvsletten, **Is there a break in scaling on centennial time scale in Holocene temperature record?** Oral presentation at *European Geosciences Union General Assembly*, Vienna, April 2015.

Nilsen, T., **Long-range memory in Earth's climate response - analysis of paleoclimate records and climate model simulations.** Oral presentation at *Arctic Marine Geology and Geophysics Research School (AMGG) annual meeting*, Tromsø, March 2015.

Nilsen, T., H.-B. Fredriksen, K. Rypdal, **Long-range memory in temperature reconstructions from ice cores: glacial vs interglacial climate conditions.** Poster presentation at *Norwegian Research School in Climate Dynamics (ResCLIM) All Staff Meeting*, Hurtigruten, March 2015.

Nilsen, T., H.-B. Fredriksen, K. Rypdal, **Long-range memory in temperature reconstructions from ice cores: glacial vs interglacial climate conditions.** Poster presentation at *American Geophysical Union Fall Meeting*, San Francisco, December 2014.

Nilsen, T., **Long-Range Memory in Millennium-Long ESM and AOGCM Experiments.** Poster presentation at *Norwegian Research School in Climate Dynamics (ResCLIM) All Staff Meeting*, Oscarsborg, March 2014.

Co-author:

Werner, J. P., D. Divine, F. C. Ljungqvist, T. Nilsen and P. Francus, **Spatio-temporal variability of Arctic summer temperatures over the past two millennia: an overview of the last major climate anomalies**, Climate of the Past Discussions, 2017:1-43, 2017, doi:10.5194/cp-2017-29.

Rypdal, M., E. B. Myklebust, K. Rypdal, H.-B. Fredriksen, O. Løvsletten, T. Nilsen, **Early-warning signals for tipping points in strongly driven systems with many characteristic time scales.** Oral presentation at *CRITICS Workshop*, Kulhuse, September 2016.

Rypdal, K., T. Nilsen, **Observations on paleoclimatic time scales. I: Earth's climate response to a changing Sun.** Subsection in book, EDP Sciences 2015, ISBN 978-2-7598-1733-7. pp. 129-138.

Rypdal, K., M. Rypdal, T. Nilsen, H.-B. Fredriksen, **The Nature of the Macroweather-Climate Scaling Break in the Holocene.** Poster presentation at *American Geophysical Union Fall meeting*, San Francisco, December 2014.

Solhaug, R. M., T. Nilsen, H.-B. Fredriksen, K. Rypdal, C. Hall, **Hvorfor skal vi stole på klimaforskerne?** Interview in *"Labyrint"*, an UiT magazine, Tromsø, October 2014.

Rypdal, K., L. Østvand, T. Nilsen, D. Divine, **Long-range memory in millennium-long ESM and AOGCM experiments.** (Oral presentation at *European Geosciences Union General Assembly*, Vienna, April 2014.

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

Chapter 2

Persistence and time series analysis

2.1 Concepts and definitions

A time series $\mathrm{X} = (\mathrm{X}_1, \mathrm{X}_2, ...)$ is a sequence of discrete-time data with the first two moments defined as the mean $\mu \equiv \mathrm{E[X]}$, and the variance $\sigma^2 \equiv (\mathrm{E}[(\mathrm{X} - \mu)^2])$. Higher-order moments include the skewness γ and the kurtosis κ.

Time series analysis is used to extract relevant statistical information from the data, either in the time domain or in the frequency domain. Such analyses may include estimating the first central moments, studying the cyclic/quasi-periodic behavior of the record and examining the serial dependence in time. Furthermore, this information may facilitate statistical modeling of the data using stochastic processes.

A stochastic process $\{X(t)\}_{t=1}^{\infty}$ is a collection of random variables associated with an indexed set of numbers, usually interpreted as points in time. If the number of points is finite the process is said to be discrete in time, while it is continuous if the index set is considered as an interval of the real line. In the following the curly braces will be dropped when the different stochastic processes are defined. It will be made clear from the context whether it is referred to the stochastic process or the one dimensional marginals (the random variables X(t)). A variety of different stochastic processes are available, a selection is presented below and are relevant for my work:

The Gaussian white noise: $w(t) \sim \mathrm{N}(\mu, \sigma^2)$, with independent and identically distributed (iid) normal draws.

The autoregressive model of order 1 (AR(1)): $Z(t) = \phi Z(t-1) + w(t)$, where ϕ is the AR(1) parameter, and w(t) is Gaussian white noise with variance σ_w^2. Stationarity requires $|\phi| < 1$. The AR(1) process is discrete in time, while the continuous equivalent is the Ornstein-Uhlenbeck process.

The Wiener process (Brownian motion): $W(t)$ is the integral of a white noise process $w(\tau)$ on $\tau \in [0, t]$:

$$W(t) = \int_0^t w(\tau) d\tau$$

$W(t)$ has the following properties:

1. $W(0) = 0$
2. $W(t)$ has independent and Gaussian increments
3. The paths of $W(t)$ are continuous

Hence, each value of a Wiener process is given by the previous value plus a random number drawn from a normal distribution with zero mean.

Fractional brownian motion (fBm): a generalization of a Brownian motion, where the increments need not be independent. The covariance structure of the fBm $B_H(t)$ is:

$$\mathbb{E}[B_H(t)B_H(s)] = \frac{1}{2}\big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\big),$$

where H is the self-similarity exponent. For a nonstationary process, H is a real number $0 < H < 1$. The fractional Brownian motion is continuous, the sample paths are almost nowhere differentiable. The process itself is nonstationary, but the increments are stationary.

Fractional Gaussian noise (fGn): the increment process of an fBm $\mathrm{X_H}(t) = \mathrm{B_H}(t+1) - \mathrm{B_H}(t)$. This process is stationary. For the increment process, the self-similarity exponent H is called the Hurst exponent ($0 < H < 1$). In the following we will only consider persistent time series, which limits H to ($\frac{1}{2} < H < 1$). The

case $0 < H < \frac{1}{2}$ is referred to as antipersistent.

The main difference between a motion and a noise is that a motion is characterized by increased variance over time. The noise on the other hand has a stationary variance. The Gaussianity of a time series can be investigated by hypothesis testing such as the Kolmogorov-Smirnov test or the Shapiro-Wilk test, in addition to graphical inspection of the quantile-quantile plot (Q-Q plot). Establishing Gaussianity in a time series allows statistical modeling using the stochastic processes introduced above.

2.1.1 Self-similarity and memory

A self-similar object is exactly or approximately similar to a part of itself. A stochastic process $B_H(t)$ is self-similar if

$$\forall a > 0 : B_H(at) \stackrel{d}{=} a^H B_H(t),$$

Where $\stackrel{d}{=}$ means equality in distribution and H is the self-similarity exponent introduced in Sect 2.1. The fBm is self-similar. The terms *scale-invariance* or simply *scaling* are also used to describe this statistical property.

The variance σ^2 and the Hurst exponent H constitute the parameters prescribed for the fGn process. This process exhibits long-range memory (LRM), also known as long-range dependence/persistence if H is in the range $\frac{1}{2} < H < 1$. An example of a discrete-time LRM process is the fractional autoregressive integrated moving average (FARIMA) process. Other types of LRM-processes also exist, but will not be further discussed. The memory terminology is related to the fact that the value at time t depends not only on the previous value at time $t-1$ but on all previous values $[-\infty, t-1]$. LRM of fGns is characterized by an algebraically decaying autocorrelation function (ACF):

$$\lim_{t\to\infty} C(t) \propto t^{\beta-1} \tag{2.1}$$

such that $\int_0^\infty C(t)\mathrm{d}t = \infty$, i.e., $0 < \beta \leq 1$. The power spectral density (PSD) also has a power law dependence in the asymptotic limit:

$$\lim_{f \to 0} S(f) \propto f^{-\beta} \tag{2.2}$$

It can be shown that $\beta = 2\mathrm{H} - 1$ and $0 < \beta < 1$ indicates positive persistence. In the following we will use the spectral parameter β when referring to the memory parameter instead of H. $\beta = 0$ corresponds to a white noise process, exhibiting no memory. Equations 2.1 and 2.2 also hold for fBms $B_H(t)$. For an fBm we have $\beta = 2\mathrm{H} + 1$, $1 < \beta < 3$. The limit $\beta = 1$ marks the transition from a fractional Gaussian noise to a fractional Brownian motion. The value $\beta = 2$ corresponds to the Brownian motion.

2.1.2 Fractality

Another description often met when working with self-similar patterns or processes is fractality. Investigating fractality in geophysical time series is relevant when choosing the statistical model used to represent the data at hand. The need to introduce this terminology arises from the possible pitfall of erroneously categorizing geophysical time series, or their cumulative sums, as self similar, when in reality they belong to another class of processes that are not self-similar. This class includes the multifractals. Careful analysis also prevent us from making the mistake of interpreting data that are not multifractal as multifractal.

In the statistical sense, monofractality implies self similarity. The details in a pattern appear to exhibit the same statistical properties independent of the scale at which it is studied. A classical example of a monofractal structure is a coastline (Mandelbrot, 1967), where the length of the coastline increases as the length of the measuring stick is decreased.

The scaling function $\zeta(q)$ can be used to distinguish monofractal and multifractal records, given power-law shape of the structure functions $\mathrm{S}_q(\Delta t)$. To investigate the shape of the structure- and scaling function of a stationary process $X(t)$ it is necessary to form the cumulative sum

$$Y(t) = \sum_{i=1}^{N} X(i)$$

The structure functions of the nonstationary process $Y(t)$ as a function of time scale Δt is:

$$S_q(\Delta t) \equiv \mathbb{E}\,|Y(t+\Delta t) - Y(t)|^q \tag{2.3}$$

Which can be estimated as:

$$\hat{S}_q(\Delta t) = \frac{1}{(N-\Delta t)} \sum_{i=1}^{N-\Delta t} |Y(i+\Delta t) - Y(t)|^q \tag{2.4}$$

It is required that the qth moment is finite. A process is monofractal or multifractal only if the structure functions are power-law functions of time. In that case, we can define a scaling function $\zeta(q)$ by the relation:

$$\mathbb{E}\,|\mathrm{Y}_t|^q \propto t^{\zeta(q)} \tag{2.5}$$

Self-similarity implies a linear scaling function, while multifractal processes have well-defined and strictly concave scaling functions (Bacry and Muzy, 2003). Note that if the structure functions of the nonstationary process are not power-laws, then the process is neither monofractal or multifractal. Section 2.5 of Rypdal and Rypdal (2016) present an example of a process which has a power-law structure function only for $q = 2$, but not for other moments. This process is a type of Lévy noise, the jump-diffusion process. It has independent draws and is non-Gaussian, but should not be mistaken for a multifractal process since the structure functions of the cumulative sum for the first and third moment are not power-laws. This example illustrates the usefulness of employing higher-order statistics in time series analysis. Additional information is revealed about the data, and the degree of ambiguity is reduced.

The above mentioned properties are used in Sect. 3.3 of Paper 2 to demonstrate that the global mean surface temperature (GMST) record for the period 1880-2010 is monofractal. The demonstration requires that the strong trend associated with anthropogenic warming is removed, and that the cumulative sum of the time series is used. Rypdal and Rypdal (2010) showed that the GMST is Gaussian, and the fGn is therefore a suitable model for this temperature record. For data extending beyond the instrumental period, similar analyses can be done to justify the use of a monofractal Gaussian model. The multiproxy reconstruction for the Northern hemisphere (Moberg et al., 2005) is used to investigate the Gaussianity and structure functions of the temperature for the past two millennia. Figure

2.1a shows the temperature reconstruction, 2.1(b) shows the pdf of the record compared to a normal distribution plotted on a logarithmic axis. 2.1c shows the quantile-quantile plot (Q-Q plot) of the data versus a normal distribution.

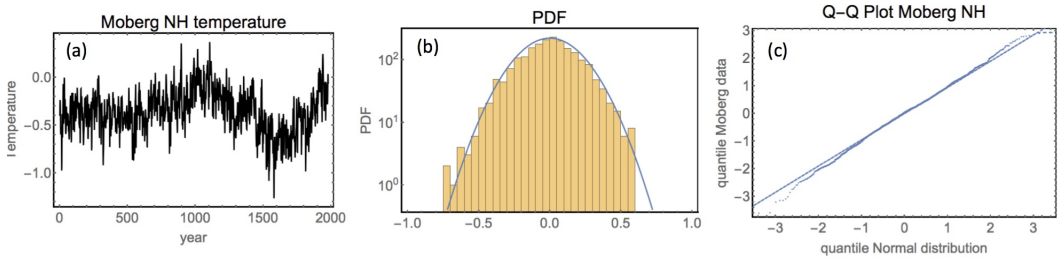


Figure 2.1: (a) The Moberg temperature time series for AD 1-1978. (b) The pdf for the record, in log plot. (c) The Q-Q plot of the record.

The pdf in Fig. 2.1b reveal a small deviation from Gaussianity for the tails of the distribution. The discrepancies are due to the volcanic responses in the temperature signal, which are manifested as sharp and abrupt drops in the temperature over 1-2 years following a volcanic eruption. The volcanic forcing is from our point of view not considered as part of the internal climate variability. The Gaussian model is therefore appropriate to represent the internal variability of for the temperature reconstruction of Moberg et al. (2005). The structure functions for moments $q = 1 - 6$ are plotted in Fig. 2.2a, which is a log-log plot. They are linear up to a time scale of approximately 300 years. The scaling function for $q = 1 - 15$ is plotted in Fig. 2.2b for time scales 4-256 years. The scaling function is close to linear, indicating that the Moberg record is monofractal and that a self-similar process is suitable for statistical modeling. The fGn is appropriate because the estimated scaling exponent $\beta = 0.74$.

On even longer time scales, the paleotemperature records from the Greenland ice core records can be used to test Gaussianity and study the structure functions. The δ^{18}O ratio is used as a paleotemperature measure, and the time series have been interpolated to obtain even time steps. The time series in Fig. 2.3a is plotted on a reversed time axis, with the abbreviation "BP" indicating before present (year 0 in this terminology is set to 1950 AD). The section of the Holocene period selected here is the past 10 000 years. The negative tail of the Q-Q plot in Fig. 2.3c is not perfectly in line with the normal distribution, but this is due to smoothing and

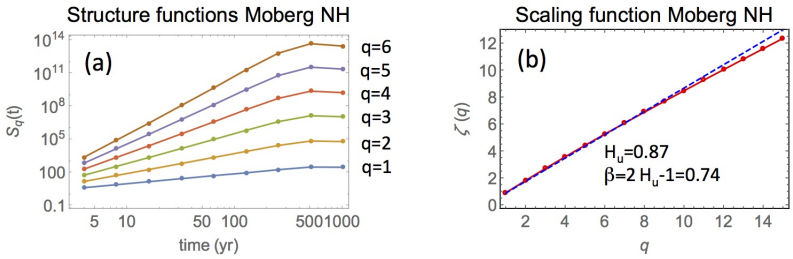


Figure 2.2: (a) The structure functions for the Moberg record for $q = 1 - 6$. (b) The scaling function.

noise effects on the high-frequency variability of the record. The deviation from normality observed in the pdf of Fig. 2.3b is to some extent due to the volcanic activity as mentioned for the Moberg et al. (2005) record above, but is also related to the cooling event observed around 8200 years BP. This event is known as the 8.2 kiloyear event, and is described further in Sect. 2.2. Truncating the GRIP Holocene record before this event gives a pdf and a q-q plot which is consistent with a Gaussian. For further details this is demonstrated in the online discussion of Paper 2 in our first response to reviewer Shaun Lovejoy.

The Gaussian model is assumed to be approximately representative for the Holocene GRIP section. The structure functions for moments $q = 1 - 7$ are plotted in the log-log plot of Fig. 2.4a. They are linear up to a time scale of approximately 1000 years. The scaling function for $q = 1 - 15$ is plotted in Fig. 2.4b for time scales 8-1024 years.

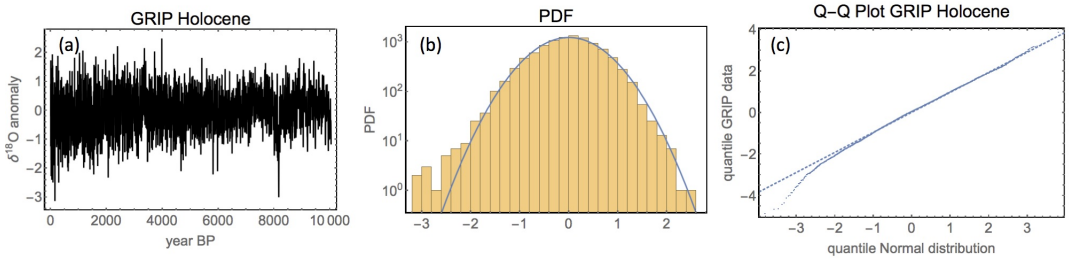


Figure 2.3: (a) The GRIP Holocene record . (b) The pdf for the record. (c) The Q-Q plot of the record.

For the last glacial period it is demonstrated that the distribution of δ^{18}O values

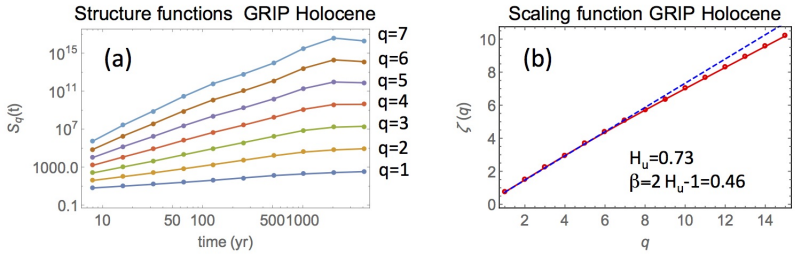


Figure 2.4: (a) The structure functions for the GRIP Holocene record for $q = 1 - 7$. (b) The scaling function.

is not Gaussian, and that the structure functions are not power-laws. Figure 2.5a shows the GRIP δ^{18}O anomaly time series for a section of the last glacial period including the last glacial maximum, spanning approximately 32 00 - 13 000 years BP. Fig. 2.5b shows the pdf, which is skewed and deviating significantly from the normal distribution. The structure functions in Fig. 2.5c are not linear in the log-log plot, and the scaling functions are therefore not shown.

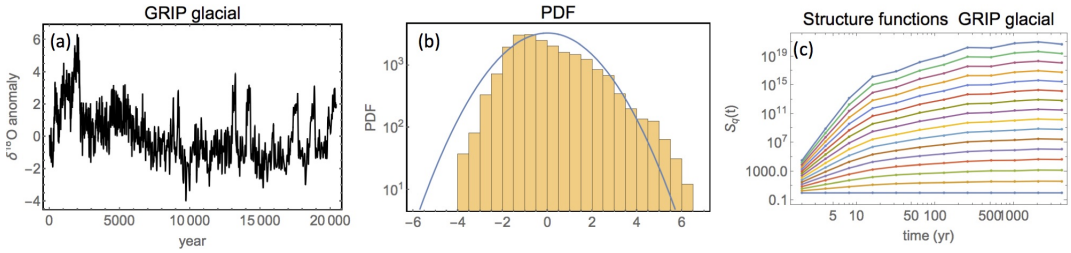


Figure 2.5: (a) The GRIP time series for a section of the last glacial period. (b) The pdf for the record. (c) The structure functions for the GRIP glacial record.

2.2 Toolbox for scaling analysis of geophysical timeseries

In principle, the information of LRM in a time series can be extracted from the fat tail of the autocorrelation function. Unfortunately, most data sets are limited in size, and hence this tail will be so contaminated by noise that a slope of this tail in a log-log plot cannot be identified. Fortunately, there are a number of other techniques available for investigating the scaling properties of a time series. A simple estimator for the power spectral density has been used in all three papers, Paper 1 and 2 also use other estimators to investigate the scaling properties. The methods

have individual strengths and sensitivities, it might therefore be advantageous to use a multi-method approach for specific data sets.

All methods study in some sense the same measure, namely how the variability in the record at hand varies on a range of time scales. The spectrum presents this variability in the frequency domain. Proxy-based reconstructions are sometimes available only with an age model with uneven time steps, making interpolation to annual or otherwise even sampling necessary. The techniques listed below are designed to be used for data with even time steps. There are also possibilities for analyzing records with uneven time-steps directly, although these techniques are less well established in the community working with climatic time series.

To estimate the **power spectral density** we use the periodogram, which for the evenly sampled time series $x(1), x(2), \ldots, x(n)$ is defined in terms of the discrete Fourier transform H_m as

$$S_m = \frac{2|H_m|^2}{N}, \; m = 1, 2, \ldots, N/2,$$

where N is the length of the time series. The frequency measured in cycles per time unit is $f_m = m/N$. The smallest frequency which can be represented in the spectrum, and the frequency resolution, is $1/N$. The periodogram is described in Sect. 2.1 of Paper 1, Sect. 3.1 of Paper 2 and Sect. 2.3 of Paper 3. Power spectra are visualized in log-log plots. If the record exhibit scaling, the scaling parameter β can be estimated through a linear fit. If the record has short-range memory (SRM) and follow an AR(1) structure, the PSD is Lorentzian in shape. The raw and log-binned periodograms are plotted, where log-binning implies that the spectral power is averaged over frequency ranges that are equidistant. The log-binned spectrum is less noisy than the raw counterpart, and β is therefore estimated from the log-binned periodogram unless otherwise stated. The raw periodogram is very noisy, and for many types of spectral analyses it is common to use windowing techniques to improve the statistical properties. This is relevant for instance if the task is to determine significance of spectral peaks against the background continuum. However, for scaling analyses we are interested in the overall shape of the spectrum, and the noise is reduced when we use log-binning to estimate β. The **Lomb-Scargle periodogram** is introduced in paper 2 as an alternative spectral technique that can be used directly on the unevenly spaced ice-core paleotemperature records. This method was found to perform well but not significantly better

than if interpolation and the standard periodogram was used.

All remaining techniques used for scaling purposes are defined in the time domain. The **wavelet variance technique** is described in Sect. 2.3 of Paper 1. This technique is sensitive to oscillations if the mother wavelet is oscillatory, and may be used specifically to identify such variability in the records studied. The resulting wavelet variance will vary depending on which mother wavelet is used. The Mexican Hat wavelet (second derivative of a Gaussian) is real valued and often used for analysis of geophysical data sets. Methods for handling missing data are presented in Mondal and Percival (2010), but were not used further in my work. The **wavelet scalogram** is described in Sect. 3.2 of Paper 2. This measure is plotted against time and time scale, and thereby provides supplementary information to the spectral analysis. The supplement is particularly useful if there are time-localized features in the time series that may cause increased/decreased power on specific frequencies. This is true for instance for the GRIP δ^{18}O record for the Holocene period, illustrated in Fig. 2.4. The abrupt decrease in δ^{18}O around 8200 years BP is related to dynamics affecting temperatures in the North Atlantic Ocean. Probably, the event was caused by a large pulse of freshwater into the North Atlantic associated with the collapse of the Laurentide ice sheet (Alley and Ágústdóttir, 2005). The effect of this event on the low-frequency spectrum power is shown by the wavelet scalogram in Fig. 11 of Paper 2.

The **detrended fluctuation analysis (DFA)** is introduced in Sect. 2.2 of Paper 1, and in Sect. 3.5 of Paper 2 it is demonstrated that the technique is unsuitable for detecting breaks in the scaling. The method will essentially shift a break to longer time scales, as illustrated in Fig. 4 of Paper 2. There exists a multifractal version of the DFA analysis, which has been used e.g. on climate data from the last glacial period (Shao and Ditlevsen, 2016). Løvsletten (2017) presents a modified DFA fluctuation function that handles missing data.

Finally, the **Haar fluctuation function** is described in Sect. 3.4 of Paper 2 and in Lovejoy and Schertzer (2012a,b). The standard fluctuation function is the square root of the second order structure function (Rypdal et al., 2013). The fluctuations are scaling with scale τ if the fluctuation function $\mathrm{F}(\tau) \sim \tau^{(\beta-1)/2}$. The Haar fluctuation analysis is a modification where the data record is convolved with the Haar wavelet.

Since there is only one realization of each climatic observation series available, which is finite and discretized, all estimators used to analyze scaling properties are subject to *finite size effects* when used on single time series. By this we refer to the fact that that the variance estimates on the longest time scales/lowest frequencies are based on very few data points, which results in larger uncertainties than on shorter time scales /higher frequencies where more data is available. The finite size effect is observed as a widening of the estimated measure for longer time scales/lower frequencies, as illustrated by Fig. 2.6b and 2.6c. The first panel shows an arbitrary realization of a synthetic fGn with $\beta = 0.7$ and n=2000 data points. The middle panel shows the 95% confidence range of the log-binned PSD for a Monte Carlo ensemble of realizations such as in (a). Panel (c) shows a similar log-log plot but for the DFA2 fluctuation function.

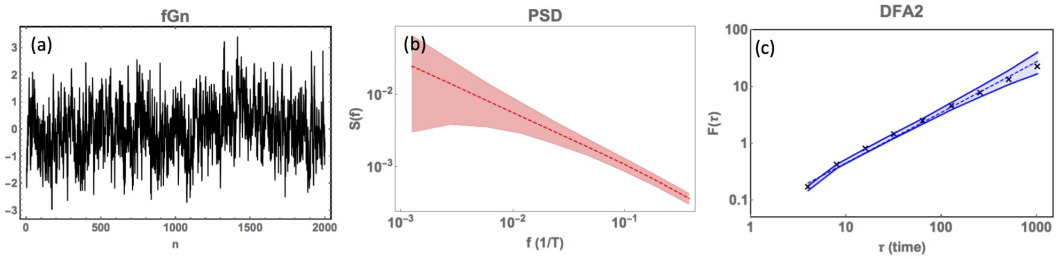


Figure 2.6: (a) An fGn time series with scaling exponent $\beta = 0.7$ and n=2000 data points. (b) The log-binned power spectra for the 95% confidence range of a Monte Carlo ensemble of fGn generated as in (a). The red, dashed line is the ensemble mean. (c) The DFA fluctuation function for the 95% confidence range of a Monte Carlo ensemble of fGn estimated as in a. The blue, dashed line is the ensemble mean, and the black crosses are fluctuation function values for a random realization of fGn.

By comparing Fig. 2.6b and 2.6c it appears that the finite size effect is more pronounced for the PSD than for the DFA function. However, we stress that the DFA variance associated with a time scale τ does not measure the variance specifically at this time scale, but is rather a weighted sum of the variances on time scales shorter than τ. The shorter time scales have more data, and hence the low uncertainties result from this artifact.

Chapter 3

Proxies and proxy-based temperature reconstructions

To learn more about the complex interplay of the climate system, we need reliable quantitative estimates of climates of the past. This chapter deals with the most used proxies involved in such studies, and the methods employed to extract climate-related information from them. The following text is partly based on Lowe and Walker (1997) and de Wit et al. (2015) [Chapter 3.1].

Apart from instrumental data and historical documents, the available data on climate variability in the past is limited to proxies that contain indirect information on environmental variables of interest. In relation to climate, the word "proxy" is often used interchangeably for *a natural archive* such as an ice core, or a *physical measurement* made from the archive such as a geochemical analysis. In fact, a number of proxy records may be extracted from a single archive in addition to an age model. Figure 3.1a shows the cross section of an ice core, which is divided into a number of segments used for a variety of analyses. This type of cut plan is typical for large ice coring projects. Figure 3.1b shows an ice core segment with visible annual layers and a volcanic ash layer that settled on the ice sheet approximately 21 000 years ago (National Science Foundation, 2015).

The proxy growth/deposition is influenced by climatic conditions, but the precise relationship is generally unknown. The hypothesis of a relationship between a proxy and an environmental variable is therefore built on scientific knowledge and a number of simplifying assumptions. Uncertainties arise from a number of sources, but first and foremost from the lack of understanding of the physical mechanisms

building the proxies. It is generally unknown how these mechanisms and processes have changed in the past. The uniformitarian principle is therefore fundamental for interpreting proxies, implying that the physical relationship between a proxy and an environmental variable is stationary in time. This is a an important source of error, the principle has certainly been violated as observed from evolution of living organisms and landscapes. The violation of the uniformitarian principle is nevertheless ignored for most proxy-variable relationships, because the evolution is hard to identify and quantify on the shorter time scales.

A proxy is considered an imperfect recorder of environmental conditions, capturing some aspects of variability but with possible discontinuities and time-dependent sensitivity. In addition, the proxy signal is noisy, meaning it reflects climate conditions in combination with some aspect of local weather variability and/or non-climate effects such as food supply and biological factors. In general, proxies must be calibrated against modern instrumental records to yield a quantitative reconstruction of past climate.

3.1 Annually banded archives

Selected proxy archives grow or are deposited following the seasonal cycle. They develop seasonal/annual bands that can be distinguished with the naked eye. The bands can be counted manually or using automated software, resulting in remarkably accurate chronologies. Such high-resolution records are highly valued as they can be used for cross-verification with other records where the age-model is less well constrained.

3.1.1 Tree rings

Dendroclimatology is the science of determining past climates from properties of the annual tree rings. Rings are wider when conditions favor growth, narrower for poor growth seasons. Another property of the annual rings, the maximum latewood density is also used as a proxy in addition to simple ring width. Using tree rings, local climates can be reconstructed for hundreds to thousands of years.

3.1.2 Ice cores

Ice cores are recovered by drilling through the Greenland and Antarctic ice sheets, glaciers in North American regions, islands of the North Atlantic and Arctic

Oceans, and alpine, tropical and sub-tropical locations. Measuring oxygen isotope ratios in water molecules allows estimation of past temperatures and snow accumulations. Isotopic fractionation makes the heavier isotope ^{18}O precipitate more easily as temperatures decrease than the lighter isotope ^{16}O. In addition to oxygen isotopes, water contains the hydrogen isotopes ^{1}H and ^{2}H, which are also used as temperature proxies. The best dated series are based on sub-annual sampling of ice cores and the counting of seasonal ice layers. Such series may have absolute dating errors as small as a few years in a millennium. Absolute age-markers in ice cores include volcanic ash layers, these are used as tie-points when the age-models are generated.

3.1.3 Corals

Paleoclimate reconstructions from corals provide insights into the past variability of the tropical and sub-tropical oceans and atmosphere, making them a key addition to terrestrial information. The corals used for paleoclimate reconstruction grow throughout the tropics in relatively shallow waters, often living for several centuries. Accurate annual age estimates are possible for most sites using a combination of annual variations in skeletal density and geochemical parameters. Paleoclimate reconstructions from corals generally rely on geochemical characteristics of the coral skeleton such as temporal variations in trace elements or stable isotopes.

3.2 Archives with dating uncertainties

A number of proxy archives are controlled by other factors than the seasonal cycle, and thereby exhibit larger age uncertainties. Typically, the deposition is irregular and simplifying assumptions are used to quantify the sedimentation/growth rate.

3.2.1 Sediment cores

Marine sediment cores are widely used for reconstructing past climate. One of the common approaches is to extract and study the marine microfossils that are preserved in the sediments. Carbonate deposits from foraminifera and coccolithophores are examples of abundant microfossils found in seafloor sediments that are good indicators of past environmental conditions. Diatoms are also of great importance for reconstructing past climate. They are unicellular, photosynthetic

algae with a siliceous shell. The general assumption is that the down-core composition of diatomic microfossil assemblages is related to past environmental conditions at the core site. A number of statistical techniques are elaborated to convert assemblages to past estimates of hydrographic conditions, including sea-surface temperature at the study site. In lake sediment cores, remains of microorganisms such as diatoms, foraminifera, microbiota, and pollen within sediment can indicate changes in past climate, since each species has a limited range of habitable conditions.

3.2.2 Speleothems

Speleothems are mineral deposits formed from groundwater within underground caverns. Stalagmites, stalactites, and other forms may be annually banded or contain compounds that can be radiometrically dated. Thickness of depositional layers or isotopic records can be used to determine past climate conditions.

3.2.3 Borehole measurements

Borehole data are direct measurements of temperature from boreholes drilled into the Earth's crust. Departures from the expected increase in temperature with depth can be interpreted in terms of changes in temperature at the surface in the past, which have slowly diffused downward, warming or cooling layers below the surface. Reconstructions show substantial sensitivity to assumptions that are needed to convert the temperature profiles to ground surface temperature changes, hence borehole data are most useful for climate reconstructions over the last five centuries.

3.3 Dating

Determining the age of paleoclimate proxy samples is based on either radiometric dating techniques, and/or incremental dating for proxies with seasonal or annual layers. Furthermore, Stratigraphic age markers such as layers of tephra (volcanic ash) can be used to constrain age models. With radiometric dating, point estimates are obtained that are typically interpolated to construct a complete age model. A number of naturally occurring radioactive isotopes exist in nature, and some are incorporated into proxy material. Radiocarbon (^{14}C) and Uranium-series dating are two well-known methods used to construct paleoclimatic age models. The

choice of dating method depends on the available material and also on the relevant timescales for the study.

3.3.1 Radiocarbon dating

^{14}C is constantly replenished in the atmosphere through cosmic rays. The ratio of $^{14}C/^{12}C$ in the atmosphere is therefore known, and through biological/geochemical processes the carbon is incorporated into living organisms. When the organisms die, the ratio of $^{14}C/^{12}C$ will decrease because the isotope ^{14}C is subject to radioactive decay. Geochemical measurements of ^{14}C in a sample of organic material of known volume are converted to age estimates with associated uncertainties by using a calibration curve. The calibration takes into account a number of factors, including the fact that the ratio of $^{14}C/^{12}C$ in the atmosphere has not been constant over time. This is an example where the violation of the uniformitarian principle is taken into account, and it has been shown to improve the age model considerably. The term radiocarbon years is used for ages estimated without this correction, and they may differ significantly from calendar years.

Radiocarbon dating require the presence of organic material, which is not always preserved in proxy archives such as e.g. ice cores. The dating technique is applicable to data with a maximum age of $\sim$45 000 years, since the radiocarbon decays over this time range. Probabilistic methods for calibration have been developed for radiocarbon dating, because the calibrated calendar ages do not have normally distributed PDF's. Bayesian statistical models are formulated so that relative dating information, if available, is implemented in the prior distribution, while the radiometric dating is expressed through the likelihood. Finding appropriate priors is the main challenge for this type of probabilistic models. Details can be found in Ramsey (2009), where the methodology used in the OxCal computer software is described. In this computer package, drawing from the posterior is done by using the MCMC algorithm called Metropolis-Hastings.

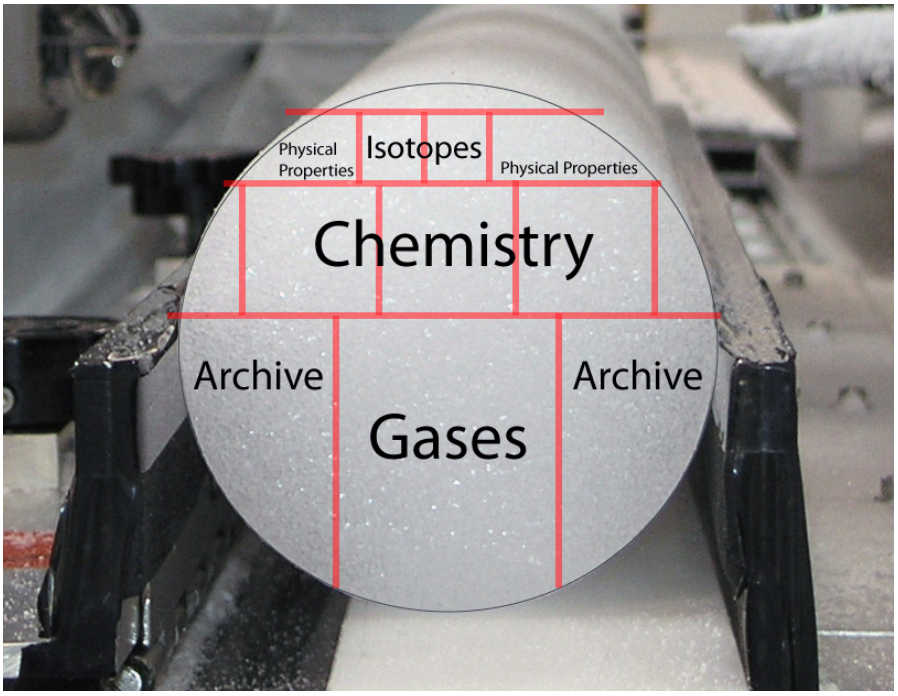
3.4 Larger scale reconstructions

If we want to understand the climate impact of external forcings like solar irradiance and volcanic aerosols, we cannot be satisfied with local reconstructions of the climate at the proxy sites. It is necessary to reconstruct a large-scale climate field and/or or mean index by employing reconstruction techniques such as mul-

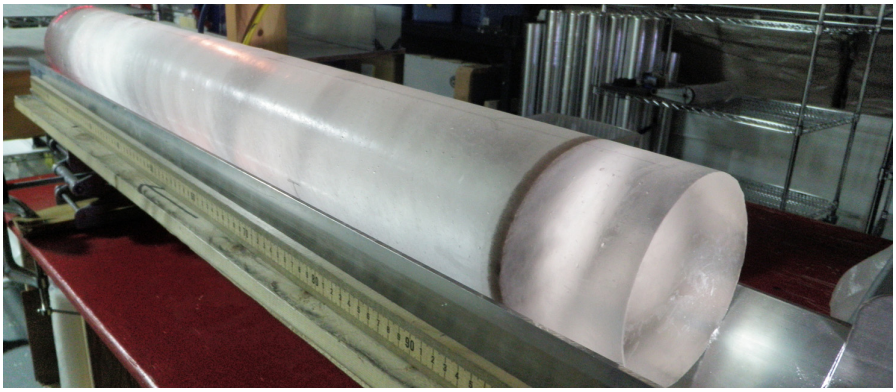
tivariate regression methods or Bayesian models. Up till recently, reconstructions of climate fields and mean temperatures on global scale have mainly been limited to the Northern hemisphere (NH). This is because the main landmasses and the majority of proxy archives are located in this hemisphere. The first NH reconstructions published in the late 1990s were obtained by techniques involving Principal Component Analysis (PCA) and relied heavily on tree-ring series. They were criticized for suppressing variance on long time scales. This was an issue because the methods could potentially suppress the temperature difference between the Medieval Warm Period (AD 800 - 1100) and the Little Ice Age (AD 1550 - 1850). A large number of millennium-long NH reconstructions based on different sets of proxy archives and statistical methods have been published later.

The past 2000 years of climate change have been reconstructed in more detail than ever before by the PAGES 2k project. The 2k Network of the Past Global Changes (PAGES) project aims to generate a globally encompassing, high-resolution regional synthesis of climate variability for the last 2000 years. The results presented in PAGES 2k Consortium (2013) reveal interesting regional differences between the different continents, but also important common trends. The global average of the global reconstruction is based on 511 climate archives from around the world. The two main results are a confirmation that current global surface temperatures are higher than at any time in the past 1400 years, and that the Medieval Warm Period and Little Ice Age were not globally synchronized events. The period from around AD 830 to 1100 generally encompassed a sustained warm interval in the Northern Hemisphere, but in South America and Australasia it occurred from around AD 1160 to 1370.

A new generation of PAGES2k reconstructions are currently under development, now generated from 692 records as described in PAGES 2k Consortium (2017). The reconstruction techniques have been updated and are even more sophisticated. The reconstructions are published in Climate of the Past special issue (McGregor et al., 2016).



(a) Cross section of an ice core with core processing line cut plan.



(b) Ice core segment with visible annual layers and volcanic ash deposits

Figure 3.1: Source: U.S. National Ice Core Laboratory

Chapter 4

Paleoclimate reconstruction techniques

4.1 Overview

There are two main types of paleoclimatic reconstructions available: those based on climate field reconstruction (CFR) techniques and those based on index methods. CFR methods generate a spatiotemporal field of the reconstructed variable of interest, here surface temperature. The index methods generate a mean value for the considered area. Common for all methods is that the reconstruction is calibrated against instrumental temperature for a common period of overlap, known as the calibration period. A segment of the instrumental period which is withdrawn from the calibration can be used for verification/validation purposes. However, this period is short and associated with the strong anthropogenic warming. The use of pseudoproxies is therefore useful for additional testing of reconstruction techniques. More on this in Sect. 7.1.

Many index and CFR techniques are based on univariate, multivariate, or multiple regression techniques. Proxies are regressed onto temperatures, or opposite. The direct regression using proxies as the independent variable is simpler, but is in conflict with the underlying physics: that the proxies are functions of temperature and not the other way around. As elaborated below, this choice is important, and the effect of error in variables (EIV) poses an additional challenge. Some types of proxy records are known to have a nonlinear relationship with temperature, for instance varve thickness of lake sediment records (Emile-Geay and Tingley, 2016). Transformation algorithms are developed that takes this nonlinearity into account,

so that transformed records can be used for linear regression with temperature.

4.2 Notes on regression-based reconstruction methods

Ordinary linear regression relates a dependent variable y to an independent variable x through:

$$y = \lambda x + \epsilon \tag{4.1}$$

Where ϵ is an error term. Whether direct or indirect regression is chosen for reconstruction purposes has effects on the resulting reconstruction, as stated in e.g. Christiansen et al. (2009); Christiansen (2011). The direct regression effectively underestimates the regression slope and the magnitude of the predicted values due to errors in the independent variable, producing a reconstruction which is biased towards zero and with underestimated low-frequency variance. This bias is known as regression dilution, and can be corrected for univariate regression but not multiple regression by an "inflation" technique. The inflation method compares the variance of the reconstruction and the instrumental observations during the calibration interval, and scales the reconstruction with a factor depending on this ratio. Indirect regression is more physically intuitive, but involves an extra step of inversion. The high-frequency variability with small amplitudes will, however, be more influenced by noise than the direct regression counterpart. An additional choice to make is whether to perform the indirect regression between each proxy record and the local or the global temperature. The former approach is used in Christiansen (2011), where the large-scale reconstruction is taken as the average of the local reconstructions. On the other hand, Mann et al. (1998, 1999) perform the regression between each proxy record and the large-scale patterns of global temperature.

For the temperature reconstruction problem, the inherent errors in both the dependent and independent variables can be partly circumvented by applying error-in-variables (EIV) regression techniques. This methodology uses the observed noisy measures of proxies and instrumental temperatures instead of the true variables. The EIV model is not identifiable under normality, i.e. additional assumptions or information is needed to correct for the underestimated regression slope. This information may concern the fraction of variance for the proxy and instrumental temperature error terms. The total least squares (TLS) regression is an example of a regression model that takes into account the error in both

dependent and independent variables, and is used for reconstruction of the Northern hemisphere decadal temperature by Hegerl et al. (2007); Mann et al. (2008). Mann et al. (2008) present one global field and two different variants of spatial mean reconstructions with decadal resolution.

Last but not least, the regression-based reconstruction problem is further complicated by the effect of the anthropogenic warming trend in the calibration period. Common statistical practice is to remove nonstationarities from the data before processing, but a caveat is that the trend could contain important information about low-frequency variability. The effect of removing this trend from the data prior to the calibration has been tested by Christiansen et al. (2009). Detrending of proxies and instrumental temperatures results in substantially worse performance for all of the seven reconstruction techniques considered in that paper.

4.3 Index reconstruction methods

The simplest index reconstructions are composite-plus scale (CPS) methods, where a collection of temperature proxies are averaged and then scaled so that the mean and variance is consistent with that of instrumental data for an overlapping calibration period. The contribution from each proxy is typically weighted according to the area or according to the correlation with local or large scale instrumental temperature. Other index methods use regression to estimate the scaling factors.

Examples of large-scale surface temperature reconstructions constructed by some form of the CPS method are Jones et al. (1998); Esper et al. (2002); Moberg et al. (2005) and one of the reconstructions in Mann et al. (2008).

4.4 CFR reconstruction techniques

Most CFR techniques are based on multivariate linear regression. For surface temperature, these techniques relate a matrix of temperature proxies to a matrix of instrumental temperature data during a calibration interval.

If $\mathbf{P} \in \mathbb{R}^{m \times n}$ is the matrix of the proxy network, and $\mathbf{T} \in \mathbb{R}^{r \times n}$ is the matrix of the instrumental temperatures, the relationship found through direct multivariate linear regression is

$$\mathbf{T}' = \mathbf{B}\mathbf{P}' + \epsilon \tag{4.2}$$

Matrix dimension m is the number of proxies, n is the length of the instrumental

period and r is the number of spatial locations in the instrumental field. The prime indicates that $\mathbf{P}$ and $\mathbf{T}$ are standardized by subtracting the mean and normalizing by the standard deviation. $\mathbf{B}$ is a matrix of regression coefficients, and ϵ is the residual error. Using the ordinary least-squares (OLS) approach we seek to minimize the mean squared error by choosing B as:

$$\mathbf{B} = (\mathbf{T}'\mathbf{P}'^{\mathrm{T}})(\mathbf{P}'\mathbf{P}'^{\mathrm{T}})^{-1} \tag{4.3}$$

Where T indicates the matrix transpose. In CFR applications, the number of spatial grid points is typically larger than the temporal dimension in the instrumental period, making the system of equations underdetermined. In this case, the inverted matrix $(\mathbf{P}'\mathbf{P}'^{\mathrm{T}})^{-1}$ does not exist without regularization. The available CFR techniques based on multivariate regression differ in how they regularize the problem. Principal component analysis (PCA) is one approach for regularization, described further in Sect. 4.5.

For the ridge regression method (Tikhonov and Arsenin, 1977), a regularization term $\mathbf{\Gamma}$ is added so that the matrix $(\mathbf{P}'\mathbf{P}'^{\mathrm{T}})$ is invertible:

$$\mathbf{B} = (\mathbf{T}'\mathbf{P}'^{\mathrm{T}})(\mathbf{P}'\mathbf{P}'^{\mathrm{T}} + \mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma})^{-1} \tag{4.4}$$

The regularization term is often chosen to be a multiple of the identity matrix. The Ridge regression technique is presented in e.g. Schneider (2001) as a means of regularization in the multivariate linear regression technique for climate field reconstruction known as RegEM, discussed in further detail in Sect. 4.6.

Another CFR method include the canonical correlation analysis (CCA) (Smerdon et al., 2011). The CCA method is related to the PCA in the sense that while PCA constructs a new orthogonal coordinate system through a transformation which optimally describes variance in one data set, the CCA method defines coordinate systems that optimally describe the cross-covariance between two data sets.

Yet another category of CFR reconstruction techniques does not rely on multivariate linear regression, but is based on Bayesian hierarchical modeling and inference, introduced in Sect. 4.7. The BARCAST reconstruction technique (Tingley and Huybers, 2010a) was used in pseudoproxy experiments in Paper 3.

4.5 Regression using principal component analysis (PCA)

The matrices $(\mathbf{P}'\mathbf{P}'^{\mathrm{T}})$ and $(\mathbf{T}'\mathbf{T}'^{\mathrm{T}})$ are the covariance matrices of the proxies and instrumental temperatures, respectively. One form of regularization for regression methods is to exploit the information in the latter matrix so that only the major patterns of variability is retained and used in the regression. The "hockey stick" temperature reconstructions in Mann et al. (1998, 1999) are based on this PCA approach. A truncated set of principal components for the instrumental temperature field is used as the independent variable, and the set of proxies is the dependent variable. First, the temperature at each grid point is standardized by subtracting the mean and normalizing by the standard deviation. The standardized matrix $\mathbf{T}$ is then constructed by weighting each grid point by the cosine of the central latitude. The orthogonal transformation results in:

$$\mathbf{T} = \sum_{k=1}^{K} \lambda_k \mathbf{u}_k \mathbf{v}_k \tag{4.5}$$

Where the empirical orthogonal function (EOF) $\mathbf{v}_k$ describes the spatial pattern of eigenvector k. The principal component (PC) $\mathbf{u}_k$ describes the variability in the time domain. The scalar λ_k comprises the fraction of resolved data variance which is standardized and weighted.

The selection of which eigenvectors to retain for the regression analysis is based on the ability to explain variance in the proxy network during a calibration interval, taking into account the spatial correlation with the multiproxy data set. The N_{EOFs} eigenvectors are then trained against each single proxy indicator using multiple regression during the calibration period, see further details in (Mann et al., 1998). The resulting reconstructed principal components are converted to reconstructed temperatures through eigenvector expansion:

$$\hat{\mathbf{T}} = \sum_{k=1}^{N_{\mathrm{EOFs}}} \lambda_k \mathbf{u}_k \mathbf{v}_k. \tag{4.6}$$

The temperature reconstruction in Mann et al. (1998) covers the period AD 1400-1995. The best spatial coverage of proxy data is in the Northern Hemisphere. Mann et al. (1999) extend the reconstruction to AD 1000. Later studies have pointed to a weakness in the reconstruction technique, causing underestimated

low-frequency variability (Zorita et al., 2003; von Storch et al., 2004; Bürger et al., 2006; Zorita et al., 2007).

4.6 RegEM CFR - regularized expectation maximum climate field reconstruction

Schneider (2001) presents an CFR technique which also exploits the proxy data prior to the instrumental period, in addition to using a calibration interval for estimation of regression coefficients. The expectation maximization method estimates parameters from incomplete data such that the likelihood of the available data is maximized iteratively given that the data is a function of the model parameters. The RegEM algorithm is used for estimation of the mean and covariance matrix for the incomplete data set, and imputation of missing values. A linear regression model relates missing "m" and available "a" values. Each record (instrumental grid cell or proxy record) $\mathbf{x}$ (consisting of missing and available values) is represented as a row vector within a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ that describes the full multivariate data set with n records and p variables. Missing values are related to available values within the same record or in other records by:

$$\mathbf{x}_m = \mu_m + (\mathbf{x}_a - \mu_a)\mathbf{B} + \mathbf{e}$$

where $\mathbf{B}$ is the matrix of regression coefficients relating available and missing values within the multivariate data set. The residual vector $\mathbf{e}$ is a random error vector with mean zero and covariance matrix $\mathbf{C}$ to be determined.

The rows $\mathbf{x}$ of the data matrix $\mathbf{X}$ can be weighted to account for differing area representation of grid box data, or differing error variances.

The iterative algorithm can be summed up as follows:

1. Missing values are filled into the matrix $\mathbf{X}$ based on some initial guess.
2. Estimates of the mean and covariance matrix are calculated, based on the complete dataset.
3. New estimates for the infilled values are inserted.

4. Iteration until convergence

Because the number of grid cells is typically greater than the number of records when reconstructing climate variables, regularization is necessary. Ridge regression is applied in the standard RegEM procedure (Schneider, 2001). The conditional maximum likelihood estimate of $\mathbf{B}$ is replaced by a regularized estimate.

When using the full RegEM CFR procedure, one starts by filling in the missing data points in the instrumental dataset using the RegEM algorithm. The full RegEM CFR algorithm is then applied to the combined proxy- and infilled instrumental data set. A calibration interval is chosen, and the temperature reconstructions from the proxies are calibrated to correspond with the instrumental data. (e.g. 1856-1995). Moving backwards in time, the RegEM algorithm is first used to fill in missing values for the time period 1800-1995, then 1700-1799 and stepwise further back in time until the desired start of the reconstruction. Note that RegEM requires that there is at least one observation at each reconstructed grid cell.

The standard RegEM procedure was used in a pseudoproxy study by Mann and Rutherford (2002). It was later modified and used in Rutherford et al. (2005); Mann et al. (2005) in a so-called "hybrid" variant, where the variability in the calibration period is separated at a time scale of 20 years. The low- and high-frequency components are subsequently composited into a single reconstruction. Furthermore, Mann et al. (2007, 2008) has abandoned the ridge regression regularization in favor of the truncated total-least squares (TTLS) approach. Both ridge regression and TTLS account for observational errors in the data. The TTLS method computes the regression coefficients by means of principal components of the covariance matrix $\mathbf{\Sigma}$. Mann et al. (2007) argue that this regularization gives more robust estimates, and that it makes the algorithm faster.

4.7 Bayesian hierarchical modeling

Another type of CFR methods are based on Bayesian hierarchical models (BHM). The Bayesian methodology is different from the common frequentistic way of thinking in the sense that parameters are not estimated by repeated experiments, but are included in probability statements that represent the current state of knowledge of a hypothesis given data. Bayes' theorem can be stated as:

$$\mathrm{p}(\mathbf{t}|\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}|\mathbf{t})\mathrm{p}(\mathbf{t})}{\mathrm{p}(\mathbf{y})}$$

where $\mathbf{t}$ is a hypothesis, typically including a set of parameters, and $\mathbf{y}$ is the collection of observations or data available. $p(\mathbf{t})$ is the prior probability, and comprises information about the hypothesis before observations are made. $p(\mathbf{y}|\mathbf{t})$ is the likelihood function, describing the probability of the observations given a set of parameters. Finally, $p(\mathbf{t}|\mathbf{y})$ is the posterior probability describing the hypothesis given the observations. The posterior includes the probability distributions for the parameters and hypothesis under study. The available information about the prior and likelihood is typically used to form probability density functions (PDF's) for the two terms. A well-defined PDF for the posterior can in simple cases be calculated directly, but when the number of possible outcomes is very large, iterative algorithms are typically used to estimate a representative distribution for the posterior. Such algorithms are called Markov Chain Monte Carlo (MCMC) methods.

Tingley and Huybers (2010a) present a reconstruction technique called BARCAST based on a BHM. The algorithm is described in detail Sect. 2.1 of Paper 3. The BARCAST model consists of three levels. A model for the true structure of the target variable is formulated at the process level, this is the field that will be reconstructed. Details on the observational data (instrumental and proxy) is described at the data level together with formulated observation equations. The relationship between proxy and target variable is formulated through a multivariate linear regression equation. At locations where there are no observations it is assumed true target variable values. At the prior level, prior probability distributions are formulated for each parameter in addition to likelihood functions for the data given the true field values and all parameters. Bayes' rule is applied, and samples are drawn from the posterior probability distribution using an MCMC algorithm known as the Gibbs sampler until convergence of the posterior. The reconstruction skill of BARCAST was tested using synthetic proxy data (Tingley and Huybers, 2010b), and was shown to provide skillful reconstructions and outperforming RegEM. BARCAST has the opportunity to estimate a complete field of the variable of interest, while RegEM requires a minimum number of available values for each location where missing values are to be imputed. On the other hand, BARCAST is more computationally demanding, (Tingley and Huybers, 2010a).

BARCAST has been used to reconstruct surface temperature over Europe (Luterbacher et al., 2016), and the Arctic (Tingley and Huybers, 2013; Werner et al., 2017). The method has also been tested and compared with the CCA

method using pseudoproxy experiments (Werner et al., 2013).

4.8 Methodology and details using a single temperature reconstruction as an example

The scaling properties in a selection of hemispheric mean multiproxy temperature reconstructions have been analyzed in Paper 2. Earlier, Rybski et al. (2006); Mills (2007); Lovejoy and Schertzer (2012b) have also used regional and hemispheric multiproxy temperature reconstructions for LRM analyses. An important caveat when performing such analyses is that the variability-level in the considered reconstructions may be altered in some sense due to the above mentioned regression dilution, or the selection of proxy records, or due to filtering techniques used. These effects are not related to the climate variability, and one should therefore be careful when interpreting the resulting curves as simply scaling or non-scaling. In fact, the curves could reflect climate variability in combination with other artifacts due to biases and noise. To gain insight into these details, the section below describe the reconstruction procedure for a temperature reconstructions which has been central in my work, Moberg et al. (2005).

4.8.1 The Northern hemisphere index temperature reconstruction by Moberg et al. (2005)

The reconstructed temperature presented in Moberg et al. (2005) is a Northern Hemisphere reconstruction covering the time period 1-1979 AD. The reconstruction was constructed to preserve low-frequency, and circumvents the regression dilution problem by applying a wavelet filtering technique. It is created from 11 low-resolution proxy time series (e.g. ice-melt records, pollen data, lake sediments, 1-180 year resolution) and 7 tree-ring records (annual resolution). The algorithm for reconstruction consists of first dividing the 18 local reconstructed temperature time series into an eastern and a western part. Linear interpolation was then applied to all time series in order to create annual mean values. The beginning and end of the time series were padded with surrogate data so that they all covered the time period; 300 BC - 2300 AD. The Mexican hat wavelet with 22 scales was then applied to each series, creating 22 time series with wavelet transforms (WT). For each scale 1-9 (Fourier timescales $<$80 years), the WT from the tree-ring proxy series were averaged. For the scales 10-22 (Fourier timescales $>$80 years), the WT

from the low-resolution proxy series were averaged. Scale 1-22 were then merged, creating two full WT time series, one for the Eastern and one for the Western Northern hemisphere. The two subsets were then averaged, and the inverse WT was calculated, creating a dimensionless NH temperature reconstruction. Finally, the mean and variance of the reconstructed temperature time series were calibrated to correspond to the instrumental data available for the time period 1856-1978.

The wavelet-filtering technique used in Moberg et al. (2005) results in a reconstruction with effective temporal resolution of four years. The high-frequency variability corresponding to time scales 1-4 years should therefore not be used for scaling analysis, as it is smoothed and not representative of the true variability on these time scales. as shown in Paper 1 and 2, the record shows excellent scaling with $\beta \sim 0.7$ on time scales from 4 years up to centennial or millennial time scales. However, caution should be taken when interpreting this reconstruction as representative for the full Northern hemisphere, as the majority of the local records are terrestrial, and only a handful represent coastal marine sites. Furthermore, (Mann et al., 2005) criticized the reconstruction and suggested that the low-frequency variability may be inflated. Moberg et al. (2008) used pseudoproxy experiments to test the reconstruction skill after this criticism. The reconstruction technique was compared against the CPS procedure, and the results showed that the wavelet-based reconstruction method performed better for pseudoproxies generated from a forced climate simulation than for a control run simulation. The results also depended on the noise type and level, which is natural and expected. After the reconstruction was first published in 2005, the understanding about proxy-temperature relationships have been improved. The proxy data quality and dating has also been refined, and the overall proxy database has been updated and expanded. The Moberg et al. (2005) temperature reconstruction is still considered as an important contribution to the ensemble of millennium-long reconstructions, but some of the proxy records it is based on are less relevant today.

Chapter 5

Discussion and summary of Paper 1

5.1 Paleoclimate model simulations

State-of-the-art climate models are based on physical processes observed in the era of instrumental observations, and parameterizations are calibrated from present climate conditions. It is therefore important to test the skill of such models for conditions that are significantly different from the present climate. The purpose is to find out if the models respond correctly and are able to simulate future climate change in a realistic manner. Paleoclimate simulations are used for such testing since they can be evaluated and compared with proxy-based observations. Since it became known in the 1980s that response to forcing is model dependent, systematic model intercomparison became an important part of the model evaluation and development. Earlier intercomparison projects for paleoclimate simulations focused on the last glacial maximum (LGM) and the Mid-Holocene period, where climate conditions were significantly different from the present climate. The most recent Paleoclimate Modelling Intercomparison Project (PMIP3) also include forced simulations for the last millennium (Braconnot et al., 2012). The climate during this time period was more similar to the present, but the intercomparison value of these simulations is that the modeled variability can be compared with a wide range of accurately dated proxy-based reconstructions. Inconsistencies between simulations and observations may indicate model deficiencies or errors in the reconstructions. There is a wide range of complexity levels for models used for paleoclimate simulations. Potential sources of discrepancies include e.g. errors in the forcings used by the models, incorrect representation of modeled responses and feedbacks, or the interpretation of proxies dominated by noise/uncertainties. Standardization

of the forcing data sets used in all model runs of PMIP3 facilitates intercomparison between models and thereby quantifies potential model-dependent uncertainties

5.2 Paper one: Long-range memory in internal and forced dynamics of millennium-long climate model simulations

5.2.1 Summary

The memory properties in selected paleoclimate simulations were investigated in a systematic manner in this paper, and compared with the memory properties of climate reconstructions. The different simulations were based on GCM models or models of intermediate complexity, with and without Earth-system modules and external forcing. The representation of physical processes and interactions in the climate system must necessarily be simplified in climate models. The degree of simplification is dependent on the model complexity. It is important to investigate if and what types of model configurations that facilitate LRM, and if the level of complexity and/or the inclusion of Earth System Modules or external forcing is essential for the memory. Vyushin et al. (2004) studied ensemble experiments of the NCAR PCM using a variety of forcing configurations, and found that inclusion of volcanic forcing greatly improved the model scaling behavior.

The advantage of working with data from climate model simulations is that the data sets are provided with high spatiotemporal resolution, they cover long periods in time and have no gaps in space or time. Two proxy-based temperature reconstructions were also analyzed with the same techniques as the model data for comparison. One of them represent the Northern Hemisphere (Moberg et al., 2005), and is a multiproxy temperature reconstruction based on seven tree-ring records and 11 low-resolution records. The other reconstruction is based on diatoms from marine sediments, Miettinen et al. (2012), and represent a local site in the Northern North Atlantic. The simulated temperature data were extracted to replicate the reconstruction regions.

Four model configurations were selected where one is included in PMIP3, (HadCM3). Three of the models are fully coupled atmosphere-ocean general circulation models (AOGCMs): COSMOS ESM, ECHO-G and HadCM3, and the last one is an Earth system model of intermediate complexity (EMIC): LOVECLIM. For COSMOS and ECHO-G there was a control run available in addition to the

forced run.

The scaling parameter β was estimated using three different techniques: the power spectrum, DFA and wavelet variance analysis. The results showed scaling in all simulated data sets, and no systematic difference between control run/forced run or the level of complexity. For the forced runs we found $0.8 < \beta < 1.2$ on timescales from decades to centuries. The corresponding results for the Moberg et al. (2005) multiproxy temperature reconstruction were $0.6 < \beta < 0.69$. The control runs had scaling parameters $0.82 < \beta < 0.87$. The scaling analyses were also conducted on residual data sets, where the theory of the stochastic-dynamic model described in Rypdal and Rypdal (2014) was applied to the model timeseries and the Moberg et al. (2005) temperature reconstruction. The general idea of the stochastic-dynamic model is that the temperature signal is a superposition of a linear response to external forcing, and a response to stochastic forcing from unresolved spatiotemporal turbulence. The latter term represent internal variability in the climate system, which can be modeled as a scaling process such as the fGn using an LRM hypothesis. Subtracting the response to external forcing from the full temperature signal leaves a detrended residual. The scaling of the residual can then be investigated. This was done for the forced runs in our study and the Moberg et al. (2005) reconstruction. The temperature response to the external forcing was estimated using available forcing timeseries. All residuals were scaling with $0.6 < \beta < 1.1$. The individual scaling parameters are lower than what was estimated from the forced runs, but the small differences simply indicates a strengthening of the memory when external forcing is included.

Through our study it has been demonstrated that state-of-the-art AOGCMs and EMICs are able to generate output temperature data exhibiting LRM. The memory is intrinsic in the modeled internal variability, but increase slightly when taking into account the temperature response to external forcing. The strength of the memory is similar to that observed in the milliennium-long mean temperature reconstruction for the Northern Hemisphere (Moberg et al., 2005). The scaling parameter estimated from DFA-2 of the local SST reconstruction of Miettinen et al. (2012) was also compared with that estimated from modeled temperature of ECHO-G from the same location. The scaling was found to be similar on timescales from years to centuries, with $0.45 < \beta < 0.67$.

5.3 Climate response in proxy-based climate reconstructions and climate simulations

This section summarises additional relevant aspects of the scientific discussion around scaling in paleoclimate model simulations versus observations. The choice of literature reflect the many sources of uncertainty that needs to be considered when interpreting LRM in model simulations.

5.3.1 The role of volcanic forcing

There is no clear consensus yet on the role of volcanic eruptions in the generation of LRM. The global instrumental temperature response to large and explosive tropical volcanic eruptions is apparent as a sharp peak in the temperature record over the subsequent two years only, (Robock and Mao, 1995). The long-term effects are less pronounced, but there are indications that the surface temperature may also be more indirectly affected for decades due to shifts in oceanic circulation (Miller et al., 2012; Schleussner et al., 2015). Such atmosphere-ocean interaction is an example of a mechanism that introduce LRM into the climate system. As mentioned, Vyushin et al. (2004) found that volcanic forcing improved the scaling in model temperature record significantly, while Østvand et al. (2014) found that the external forcing did not have any substantial influence on the scaling. The contradicting results gives reason to elaborate more systematically on the model skill in representing the temperature response to strong, volcanic eruptions.

Many of the small-scale physical processes associated with volcanic eruptions have traditionally been parameterized in climate models. The actual dynamical processes are highly complex and involve atmospheric and oceanic interaction and teleconnections. The spatiotemporal evolution of the temperature change associated with a specific eruption depends on a number of factors such as the season and location of the eruption, the chemical composition of the ash plume, and the size distribution of sulphate particles. Observational data from large, tropical, explosive volcanic eruptions with global impact on the climate is needed to model the dynamical processes and calibrate the response in climate models. For obvious reasons, the amount of such observational data is limited. The Pinatubo eruption in 1991 was monitored closely using modern in-situ and remote-sensing instruments, but further back in time the observational data basis is sparse and based on indirect measurements.

Driscoll et al. (2012) demonstrated discrepancies in the spatially distributed surface temperature anomaly between CMIP5 model simulations and instrumental observations after recent, major tropical volcanic eruptions. The reanalysis data set of the 20th Century version 2 (20CRv2) Compo et al. (2011), was used to represent observations. The study points to well-known dynamical features that the models fail to reproduce after large eruptions, including a strengthened NH polar vortex, positive NAO and a pattern of warming across Eurasia. The spatial pattern of modeled cooling was shown to be substantially stronger than the observed.

Mann et al. (2012) revealed a similar problem for historical eruptions by comparing the Northern Hemisphere mean surface temperature for one AOGCM, one EBM and one tree-ring reconstruction. For three major eruptions the modeled temperature decrease was significantly stronger than the reconstructed, e.g. $\sim 2°$C cooling versus $\sim 0.6°$ for the 1258/1259 eruption. It is not trivial to find out how such large volcanic spikes in the temperature response affect the LRM. Mann et al. (2012, 2013) hypothesized that trees near the tree-line might have a reduced sensitivity to cooling, so that the strong cooling was simply not recorded in the tree-rings after abrupt, explosive volcanic eruptions. This effect was further demonstrated using a biological growth model. However, Anchukaitis et al. (2012) pointed to a number of factors that contradict this hypothesis, including the improper choice of tree-ring growth model and assumptions about temperature thresholds for growth and length of growing season. By changing the tree-growth model and replacing the parameters with more realistic measures, they demonstrated that trees near the tree-line do indeed record the temperature change associated with explosive volcanic eruptions. It was pointed out that large-scale tree-ring based reconstructions may suffer from lagged effects from prior-year weather on subsequent ring formation, and also that the spatial distribution of the tree-ring network may influence the mean temperature variability.

Another possible explanation for the deviation between the modeled temperature response and the observed temperature data is that volcanic forcing for the past millennium is generally poorly constrained with respect to magnitude prior to the instrumental period (Fernández-Donado et al., 2013). The volcanic forcing is primarily reconstructed from sulphate in ice cores. However, the ice core records have higher age uncertainties than tree-ring records, where the enrichment of ^{14}C in tree-rings following an eruption provides an absolute age marker (Sigl et al., 2015).

Consistency between tree-ring and ice-core chronologies is therefore crucial to construct a coherent age model for the volcanic forcing. The exact timing of volcanic eruptions prior to 1250 CE is debated because of such chronological discrepancies between ice cores and tree ring records. Sigl et al. (2015) constrained and revised the Greenland and Antarctic ice core chronologies using age markers linking the age models of ice-cores and tree-rings. The revision of the ice core chronologies before 1250 AD implies necessary updates to the reconstructed volcanic forcing, and it also means that the magnitude of high-to mid frequency temperature variability (annual to a few years) may be biased in multiproxy climate reconstructions where ice core records are used together with other proxy types.

If the volcanic forcing used as input in climate models is poorly constrained as described in Fernández-Donado et al. (2013); Sigl et al. (2015), and the models cannot simulate the dynamics associated with volcanic eruptions, (Driscoll et al., 2012), it seems challenging to test the long-term temperature response to a single eruption. A new generation of climate model simulations is currently under planning and preparation (Zanchettin et al., 2016), the "Model intercomparison project on the climatic response to volcanic forcing", (VolMIP). This set of experiments is part of the CMIP6 initiative, and includes short simulations to investigate the short-term seasonal to interannual dynamical response to a Pinatubo-like eruption, longer simulations to study long-term effects of a single eruption, and experiments on volcanic clusters. By performing idealized, ensemble-based experiments it is possible to pin down the climate response in different layers of the atmosphere and the ocean, and to observe how the internal variability contributes to the response. The intention of the VolMIP experiments is to assess the coherence between climate models in simulating the response to volcanic forcing starting from similar initial conditions and using the same set of external forcing, and identify causes why such robustness is not achieved.

5.3.2 Multi-scale climate variability in proxy-based reconstructions and paleoclimate simulations

Laepple and Huybers (2014a) demonstrate significant differences in the magnitude of regional sea surface temperature (SST) variability on multidecadal time scales for climate models and instrumental measurements. The power spectrum shows that models have substantially underestimated power on decadal time scales and longer compared with observation temperature. The instrumental data are cor-

rected for different sources of noise prior to spectral analyses. The authors suggest that the deviation is a result of models systematically underestimating regional sea surface temperature variability, and that the mismatch is most evident on long time scales. On the global level the consistency is much better, which may be due to stronger spatial covariance in the models than in the observations (PAGES 2k-PMIP3 group, 2015).

Laepple and Huybers (2014b) compare the level of variability for SST in proxy-based reconstructions and simulations. Again, the spectra show a clear difference on the multidecadal and longer time scales, but longer time periods are available for study than in Laepple and Huybers (2014a). For the proxy records, a continued increase in power is observed for frequencies corresponding to decadal time scales and longer, while the models show a flat spectrum on the lowest frequencies available. The proxy records are corrected for different sources of noise before comparison, which are different and less well constrained than the sources of instrumental noise. The proposed explanations are that modeled internal variability may be underestimated, or the external forcing or the responses in the models may be too weak.

The issue of continental-scale temperature variability in models and proxy reconstructions was the focus of (PAGES 2k-PMIP3 group, 2015). The variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions were compared for the past millennium. Reconstructions cover the landmasses of the Arctic, Europe, North America, Asia, South America, Australasia and Antarctica. The simulations were required to fulfill certain criteria, leading to exclusion of some PMIP3 past millennium simulations and inclusion of some pre-PMIP3 simulations. Direct observation of the time series indicate stronger long-term trends and more variability on centennial and longer time scales in the reconstructions than in the simulations. Spectral analysis shows that most of the models exhibit more variance than the reconstructions at high frequencies, and for three out of seven regions the models show less variability than the reconstructions on multidecadal time scales and longer. At the same time, the spatial correlation structure was demonstrated to be considerably stronger for the simulations than the reconstructions. Response to volcanic forcing is considerably stronger in the simulations than in the reconstruction, with cooling of -0.5 to -1 °C in the models in contrast to -0.1 to -0.25 °C for the reconstructions.

Franke et al. (2013) estimate local spectral exponents from power spectra for

instrumental, reanalysis, proxy and simulated data. The proxy data material is dominated by tree-ring width (TRW) and tree ring density (MXD), and it is demonstrated that the proxy data have a larger spread in beta-values than the instrumental data, biased towards higher values. It is suggested that the low-frequency variability in tree-ring records is overestimated compared with instrumental, reanalysis and model temperature. The bias in local beta values may propagate into multiproxy reconstructions that heavily depends on tree-ring data. This is illustrated by estimating local spectral exponents for the global climate field reconstruction by Mann et al. (2009). This reconstruction is dominated by tree-ring records, and local values of beta are mainly $1 < \beta < 2$ (Franke et al., 2013), supplementary material. In contrast, local β values for the CRU TS3 dataset are $0 < \beta < 1$.

Chapter 6

Discussion and summary of Paper 2

6.1 Paper two: Are there multiple scaling regimes in Holocene temperature records?

There are different views on how scaling models can be used to describe the Earth's surface temperature. One perception is presented in Rypdal (2012); Rypdal and Rypdal (2014); Rypdal et al. (2015); Rypdal and Rypdal (2016). The surface temperature is considered as a superposition of internal variability (the response to stochastic forcing) and a forced signal which is the linear response to external forcing. The background variability may be modeled using a stochastic process such as the fGn or fBm, or a different process that incorporates non-gaussianity if this is considered more appropriate. The response to external forcing can be modeled using periodic functions with superposed trends.

The concept of a break in scaling implies a different model for surface temperature. Spectral analysis of temperature proxy data representing time periods longer than the Holocene typically gives a spectrum with a characteristic break in scaling around centennial timescales. Lovejoy and Schertzer (2012b) use this property to construct a conceptual framework where temperature variability is divided into different categories associated with the time scales involved. Temperature variability on time scales shorter than 10 days is "weather", while variability between 10 days to a century is "macroweather". For time scales longer than centennial the variability is categorized as "climate". So far, these terms have not gained general acceptance. The variability associated with "climate" is typically dominated by data from the last glacial period or even further back in time. The

Earth's climate is nonstationary over such long timescales, and the concept of scaling may be considered less meaningful for these types of records without an associated statistical model. For the present, Holocene climate there is no scale break in detrended temperature records, and it may be argued that the simpler model with one single scaling regime is more suitable for describing variability up to a few millennia. From an information criteria perspective the simpler model has fewer parameters, and should therefore be preferred. it is possible to use a single model with separate scaling regimes to describe the Earth's surface temperature over a range of time scales. However, the proxy temperature records from the last glacial climate state are characterized by strong nonlinearities. The same is true for the interglacial-glacial transitions. Describing this variability by a single scaling regime is an extreme simplification, and it can therefore be debated how useful the scale-break model is.

6.1.1 Summary

The Earth's surface temperature record is highly nonstationary on timescales beyond the Holocene. Ditlevsen et al. (1996); Huybers and Curry (2006) and Lovejoy and Schertzer (2012b) all present log-log power spectra for temperature and/or temperature proxies covering the Holocene and the last glacial period, with estimated scaling parameters β. The three studies all identify a break in scaling at centennial time scales, and β changes from approximately $0 < \beta < 0.6$ on the shorter scales to $\beta > 1$ on the longer. Huybers and Curry (2006) suggest that the power-law continuum in the spectrum of surface temperature on time scales between one year and a century is a result of an inverse cascade in frequency space driven by the seasonal cycle forcing.

The alternative view of Rypdal (2012); Rypdal and Rypdal (2014); Rypdal et al. (2015); Rypdal and Rypdal (2016) is that a linear power-law response to stochastic forcing yields the $\beta < 1$ scaling, and the power-law response is interpreted from an energy balance perspective where energy is exchanged between different parts of the climate system with different heat capacities and response times. The fGn has been frequently used to model the Earth's surface temperature. When $0 < \beta < 1$, this process is persistent. If we want to model a process with $\beta > 1$, an fBm is used.

It is necessary to study proxy records that extend into the last glacial period to detect the observed scaling transition around centennial time scales with

significance. In Huybers and Curry (2006); Lovejoy and Schertzer (2012b), the scale break appears in composite power spectra. The high-frequency variability is represented by recent high-resolution temperature records and -proxies, while low-frequency variability is represented by temperature proxies with lower temporal resolution covering time periods longer than the Holocene. This means the scale-break only appears when considering proxy types of different nature and two climate states that are very different from each other.

If the scale-break is a universal feature, it should also be observed in records for the Holocene only. Assume that the centennial time scale is chosen as an arbitrary boundary between high- and low frequency temperature variability. Then, due to the temporal coverage of instrumental and reanalysis temperature records, and the nature of proxy-based reconstructions, most records represent *either* high- *or* low frequency variability, but very few can be used to actually identify the claimed scale break at this transition.

The existence of a potential scale-break at centennial time scales in the Holocene was investigated using statistical hypothesis testing. This involves formulating a conservative null hypothesis, in our case that the Holocene temperature variability can be described using a stochastic process with a single scaling regime. The alternative hypothesis is that a two-regime model with a break at centennial time scales is a better description. A significance level of 5% is chosen for rejection of the null hypothesis. If the temperature records are unlikely to be realizations of fGn processes with a single scaling regime, the null hypothesis is rejected.

The Greenland and Antarctic ice core records have high enough temporal resolution, and long enough coverage in time to detect and test the significance of a potential scale break. For simplicity, and due to the sampling rate, data from the Holocene and the last glacial period was analyzed, but not further back in time. Two different approaches were used; to separate the ice core records into a Holocene and a glacial part, and to analyze the full record in search for a significant break. The results showed that for the Holocene-only records, an increase in spectral power was observed at millennial time scales, but not at centennial. On the other hand, a centennial scale break was detected in the glacial record, and also in the full record. This scale-break was significant, and appears to be associated with the difference in glacial/interglacial variability, manifested for instance as the presence of Dansgaard-Oeschger (DO) events in Greenland and North Atlantic records from the last glacial period.

Six late-Holocene multiproxy reconstructions were also analyzed in search for a potential scale break, but the records were too short to claim any significant break. One multiproxy reconstruction covering the full Holocene period was also included. However, it was demonstrated that the high-frequency variability was artificially low in this record due to the selection of proxy data and the reconstruction technique.

Our conclusion is that the available temperature data from the Holocene is insufficient to justify the notion of separate scaling regimes for Earth's surface temperature on time scales longer and shorter than centennial. Only when data from the last glacial period were included in the analyses, the scale-break appeared to be significant. The notion of multiple scaling regimes is illustrative of the temperature variability on a large span of time scales: epochs, periods, eras and eons. The Earth's surface temperature variability is non-stationary over this range of time scales, so the scaling properties will depend on the time period, and what the dominating dynamics are. As a simple example, consider a conceptual model of the Greenland paleotemperature records for the Holocene and the last glacial period. The Holocene is similar to a white noise process of length N, connected at the end with the last glacial period, represented as a highly intermittent record of length $\sim$ 10*N with a higher standard deviation. It is possible to estimate the power spectrum of this combined record and study the scaling properties, but note that the very different characteristics of the two time periods will be merged together. The power spectrum will be dominated by the variability in the last glacial period, and the shorter Holocene period will essentially be suppressed. This is acceptable only if the longer time scales are considered more important. If the two-regime model is considered universal, prediction based on this model would include the transition to a new glacial period at centennial time scales, simply because the model required this. This does not comply with recent findings summarised in Masson-Delmotte et al. (2013, Chapter 5), where it is stated that "It is therefore *virtually certain* that orbital forcing will not trigger a glacial inception before the end of the next millennium".

In the end, the problem at hand is about model selection. The problem-solving principle known as "Occam's razor" states that "Among competing hypotheses, the one with the fewest assumptions should be selected". Without doubt, the simplest model for the Holocene climate is the one with a single, monofractal scaling regime, which should be used as the null hypothesis.

6.1.2 The review process

The manuscript was submitted to Earth System Dynamics. Since there was a lengthy review process and numerous disagreements, the discussion phase is summarized below.

6.1.3 The scope

The centennial scale-break appeared in our analyses of ice core records only when the last glacial period was included. We did, however, observe an increase in power at the lowest frequencies for some of the Holocene records, but this was not a statistically significant break in scaling. To test the statistical significance we performed numerical experiments with idealized long-range memory processes. The null hypothesis of a single scaling regime could not be rejected. It was also demonstrated that spurious increases/decreases in power appeared even in realisations of idealized processes defined by a single scaling parameter. Such deviations from scaling occur simply as artifacts.

Our conclusion was therefore that the one- and two-regime model were both possible descriptions for the Holocene temperature, but the one-regime model should be preferred. information-theoretic criteria are used for model selection in modern science, with the intention of finding a suitable description that does not lead to overfitting. Our main results were therefore of a negative type, which is unconventional in the peer-reviewed literature. The results revealed possible type 1 statistical errors (false positives) in the existing literature on this specific topic.

It appears that the intermittency associated with deglaciation and the DO events/Southern Hemisphere teleconnections causes the centennial scale break observed in the ice core records. This variability associated with nonlinear dynamics is important, but we question the relevance to describe it as a second scaling regime. The physical processes are poorly understood, and assigning a scaling parameter has little value without an associated statistical model. In Rypdal and Rypdal (2016), the scaling in the Greenland ice core records for the last glacial period are investigated. It is suggested that the underlying climate variability can be described as a 1/f noise, that is, when only the intervals between the DO-events are analyzed. The events themselves, in addition to the deglaciation, are separated from the noise background.

6.1.4 Remarks about method selection

Time series for proxy reconstructions typically have uneven sampling in time, which cannot be handled directly by standard techniques for scaling analysis. Linear interpolation can be used to obtain even sampling, or alternatively one can use a method that works for unevenly spaced data. Linear interpolation will by definition change the PSD of the signal, in some cases producing underestimated high-frequency variability. The "red-bias" in the power spectrum is an artifact which is observed as a steep decrease in spectral power at high frequencies in a log-log plot, and may be misinterpreted as a separate scaling regime. In our work it was important to distinguish true scale breaks in the spectra from artifacts and biases. For some of the evenly sampled time series such a bias was already observed, such as Moberg et al. (2005); Mann et al. (2009). A method known as the Lomb-Scargle periodogram was investigated since it can be used directly on unevenly sampled data. We were not certain beforehand that the method would be suitable for our purpose, since it was originally designed to handle periodic data with random missing values. There were few earlier studies where this method was used for paleoclimate data, one example is (Pelletier, 1998). After testing the Lomb-Scargle periodogram and the first round of reviews, it was decided to use interpolation and the standard periodogram in the main paper, and include the test results for the Lomb-Scargle periodogram in supplementary material. Contrary to the red bias observed in the standard periodogram, the Lomb-Scargle periodogram exhibit a blue bias for high frequencies. The skill of this estimator depends on the skewness of the distribution of sampling intervals. It was found to perform well but not significantly different from using linear interpolation in combination with the standard periodogram.

The Mexican Hat and Morlet wavelet scalograms were plotted for illustrative purposes for some of the interpolated time series versus time and time scale. Using this technique, it is possible to detect systematic differences in power over the length of the record. This was the case for the Marcott et al. (2013) reconstruction. The wavelet scalogram was also used to pin down time-localised events that caused changes in power in one ice core record. We demonstrated that the increase in power observed at millennial time scales in the Holocene GRIP record occurred due to the 8.2 kiloyear event. This was an abrupt cooling event caused by sudden drainage of the proglacial lake Agassiz into the North Atlantic (Alley and Ágústdóttir, 2005).

In the accepted manuscript, only selected wavelet scalograms were included. These plots were used as supplements to the main results from the spectral analyses.

DFA of order n was shown to be insensitive to scale breaks, and to shift the scale where breaks are observed. The performance of the Haar wavelet fluctuation analysis was tested and compared to the power spectrum, and the two methods appeared to give similar results and uncertainties. However, breaks in scaling appeared more pronounced in plots of the Haar fluctuation function versus the power spectrum, even if they were not statistically significant. Structure- and scaling functions were introduced to demonstrate that the Holocene temperature records at hand are monofractal. The fGn model was thereby justified, and using the Haar fluctuation it was shown that a one-regime model plus a trend has significantly lower uncertainties on the longest time scales than the two-regime model.

6.1.5 How should proxy-based reconstructions be used for scaling analyses?

When working with scaling in paleoclimatic temperature, the limitations and uncertainties of proxies and proxy-based reconstructions have to be taken into account. The proxy-based data go through a number of processing steps and are based on fundamental assumptions not relevant to instrumental temperature. These assumptions arise from e.g. the complexity of physical and/or biological processes associated with the climate proxy formation. Although physically or biologically based models (forward proxy models) have been evolving during the recent years, the relation between the proxy and the variable it is assumed to record is generally unknown. It is also unknown how the proxy-variable relation changes with time. The uniformitarian principle is therefore used, i.e. that the relation is assumed to be stationary in time. The scaling does not necessarily reflect only the true climate variability, and the low versus high frequency variability is often misrepresented. In addition, the sampling site of local/regional proxy records is generally carefully chosen to investigate a specific phenomena or type of variability, and is not necessarily representative of the variability over a larger region. All things considered, the proxy-related uncertainties makes it harder to draw firm conclusions about scaling from these types of data than from instrumental measurements. Proxy-based reconstructions should not be used uncritically for scaling

analyses, as is demonstrated in this paper. Nevertheless, even if the reconstructions are poor compared to actual instrumental measurements, the scaling might provide useful information. We argue that the noise and uncertainties associated with the proxies makes it even harder to justify a scaling model with two regimes, separated at centennial time scales. Given the uncertainties, inference based on proxy-based reconstructions should be taken with more caution than the one inferred from instrumental data.

6.1.6 Principles of data selection

The scaling properties in paleotemperature reconstructions have been analyzed in e.g. Pelletier (1998); Blender et al. (2006); Rybski et al. (2006) in addition to the papers mentioned above. Some of the data sets were re-analyzed since we were interested in a potential scale-break in addition to general scaling properties. Six high-resolution ice core paleotemperature records from Greenland and Antarctica were selected: GRIP, GISP2, NGRIP, EPICA Dome C, Taylor Dome and Vostok. In addition, six proxy/multiproxy reconstructions from the late Holocene and one multiproxy reconstruction covering the full Holocene period were analyzed: Jones et al. (1998); Briffa et al. (2001); Esper et al. (2002); Huang (2004); Moberg et al. (2005); Mann et al. (2008); Neukom et al. (2014). The late-Holocene records were too short to detect any significant break, while an increase in power at millennial time scales could be detected in the Holocene part of the ice core records. For the full-Holocene multiproxy record of Marcott et al. (2013) it was demonstrated that the high-frequency variability is incorrectly represented in the earlier part of the record. This is an example of a bias associated directly with the particular multiproxy reconstruction procedure and data selection. The bias should be accounted for before doing scaling analysis on the record.

Lovejoy and Schertzer (2012b) claimed that the Holocene paleotemperatures from Greenland ice cores are exceptionally stable, and should therefore be given less credit. He refers to a paleotemperature record in Berner et al. (2008) that should be considered more realistic for Northern Hemisphere, Holocene climate. This record has considerably higher fluctuation levels, and scaling analysis give $\beta > 1$ as opposed to the ice core data which have $\beta \sim 0$. The reconstruction referred to is based on diatoms from a marine sediment core, sampled at Reykjanes Ridge Southwest of Iceland. Lovejoy and Schertzer (2012b) do not mention that the sampling site in Berner et al. (2008) was chosen specifically because it is

optimal for detecting effects of quasi-cycle ice rafting pulses, sometimes denoted as Bond events. The authors report that such periodic cooling events are indeed detected in the record. They are known to be localised to the North Atlantic region and are therefore not global/hemispheric features. Even if the sampling sites of the ice cores and marine sediment cores are just 1500 km apart, they belong to different realms and have different seasonality. The observed differences in the scaling parameters for the ice core/marine sediment data is therefore unsurprising. In particular, Fredriksen and Rypdal (2016) demonstrate that the land/ocean difference in heat capacity influences local values of the scaling parameter. Specifically, $\beta_{sea} > \beta_{land}$. This does not necessarily involve the existence of two different scaling regimes on different timescales.

6.1.7 Concluding discussion about the scale break and its validity

Though our study we have demonstrated that the scale break observed at centennial time scales in composite spectra in Ditlevsen et al. (1996); Huybers and Curry (2006); Lovejoy and Schertzer (2012b) appears to be an artifact of compositing high-resolution temperature records from the present, Holocene climate, and low-resolution proxy records from the last glacial period and even further back in time. On longer timescales than 10 00 years the Earth's surface temperature variability represents a manifestation of a nonstationary and nonlinear process with the nonlinearity being largely a result of transitions between different climate attractors. Using scaling to describe such an evolving system implies a substantial simplification with a loss of crucial information on the mechanisms and behaviour of the climate system. Due to a substantial difference in the variance of the climate related signal for the last glacial period and the Holocene, fitting a single parameter model to a composite record will result in a dominance of the glacial period data in the model fitting procedure.

Statistical hypothesis testing and employing information principles cannot be used to favor a statistical model with two scaling regimes using available temperature records from the Holocene alone. Furthermore, the estimation of different scaling parameters from different records does not indicate different scaling regimes, but may very well be a result of different climate dynamics dominating locally, as well as effects of uncertainties in proxy-based reconstructions.

Chapter 7

Discussion and summary of Paper 3

7.1 Testing the skill of reconstruction methods

The "skill" of a reconstruction method is commonly quantified through statistical measures related to the mean-squared error between reconstructed and true target variables during a validation interval. If instrumental data represent the target data used for validation, then the testing is performed over a very short interval in time, typically less than 100 years that are withheld from the calibration period. Another option is to use synthetic temperature data for these validation exercises, performing so-called *pseudoproxy experiments*. Target values are typically extracted from millennium-long paleoclimate GCM simulations or from instrumental data sets, and pseudoproxies are constructed by adding synthetic noise to simulate real proxy uncertainties. The target values are extracted in a spatial pattern simulating real-world multiproxy networks. The calibration interval is again chosen to be in modern time, but instead of true instrumental data it is also possible to use pseudoinstrumental data for regression purposes. Pseudoinstrumental data are obtained by extracting modeled target data from a spatial complete field for the modern period. The pseudoproxy-based reconstruction can be compared with the true model target for the remaining reconstruction period, which is now typically hundreds of years long. A pseudoproxy experiment can be used to assess/compare skill of one or a number of reconstruction methods, or to validate a particular reconstruction. A review on the topic is found in (Smerdon, 2012).

Through different experiment designs, pseudoproxy experiments are used to

investigate the sensitivity of a reconstruction method to some types of uncertainties in an idealized fashion. A number of experiments are typically conducted for a single model simulation and target variable, and the reconstruction skill for each experiment is estimated. For instance, a known proxy uncertainty is noise color and level. In pseudoproxy experiments, separate experiments can be conducted using white or red proxy noise, and the signal to noise ratio (SNR) can be varied between ∞ and zero, spanning the range from a perfect proxy signal to a signal masked completely by noise. Varying the SNR is how proxy data quality is modeled. The SNR is related to the correlation between proxy and target variable:

$$\mathrm{SNR} = \frac{|\rho|}{\sqrt{1-\rho^2}}$$

Where ρ is the Pearson's correlation coefficient. Note that different correlation measures can be applied. The local correlation refers to the correlation between a proxy signal and the instrumental variable in close proximity. In real-world proxy networks, the SNR in proxies has been estimated to 0.5-0.25 based on local correlation, (Smerdon, 2012). Maximum correlation refers to the highest correlation estimated between a proxy and any of the spatially distributed instrumental variables for a given year, (Wang et al., 2014). Other pseuodoproxy experiments handle uncertainties in the spatiotemporal distribution of the proxy network, seasonality or nonlinear aspects of the proxy-variable relationship. It is possible to use different models and/or model realizations to test robustness of the results. Because reconstruction methods may be sensitive to the individual noise realizations used for the pseudoproxies, an ensemble of the same experiment is typically used to obtain statistical uncertainties associated with this factor.

7.2 Paper 3: How wrong are climate field reconstruction techniques in reconstructing a climate with long-range memory?

7.2.1 Summary

In Paper 3 we study the performance of the Bayesian hierarchical model BARCAST described in Tingley and Huybers (2010a) through pseuodoproxy experiments. Target data are generated as stochastic LRM fields with prescribed spatiotemporal covariance structures. In contrast to the LRM structure of the input

data, the underlying assumption of the BARCAST reconstruction method to be tested is that the data follow an AR(1) structure in time. Pseudoinstrumental and pseudoproxy data were constructed by sampling the data in a spatiotemporal idealized pattern, and adding different levels of white noise to generate pseudoproxy data. The data were then used as input for the BARCAST reconstruction technique, and a number of analyses were performed on the resulting ensemble of reconstructions to compare it with the target.

One of the goals of this study is to investigate if BARCAST preserves the LRM properties of the target data. For this purpose, the spectra of the local and spatial mean reconstructions and target data are estimated using the periodogram, and hypothesis testing is performed. The reconstructed spectra are estimated on an ensemble member basis, and it is tested whether the average spectrum is consistent with the spectrum of the theoretical LRM model for the target data. The results show that the LRM-scaling of the local reconstructions are slightly better if measured between proxy sites than if measured directly at a proxy site. When the local reconstructions are based on noisy input data, they are not consistent with the LRM-hypothesis at the proxy site. However, for low noise-levels they are scaling between proxy sites. The spatial mean reconstructions show better scaling properties than the local reconstructions. Hypothesis testing was also performed in the spectral domain using the null hypothesis that the reconstructed data follow the AR(1) model with parameters estimated from BARCAST. This null-hypothesis was rejected for all experiments both for the local and spatial mean reconstructions.

The reconstruction skill was additionally tested using selected skill measures: the correlation, the root mean squared error and the continuous ranked probability score. The skill was found to be positive for all experiments both for local reconstructions and the spatial mean, and decreasing in the expected manner when the signal-to-noise ratio was decreased.

7.3 Other studies using pseudoproxy experiments

The Mann et al. (2007) pseudoproxy study includes a suite of experiments performed to validate the RegEM method. Target data is extracted from two millennium-long GCM simulations. Addition of white/red proxy noise is tested, and also varying the SNR and using temporally variable number of proxy records. The

study is a response to former criticism of CFR methods, and it is concluded in the pseudoproxy study that RegEM provide robust results.

Tingley and Huybers (2010a) use real instrumental values as targets for pseudoproxies and thereby restrict the reconstruction of spatial fields of surface temperature over North America to cover the instrumental period. Experiments are conducted using varying SNR and varying the total number of pseudoproxy records.

Wang et al. (2014) compares four different CFR methodologies for modeling surface temperature using PPE's. Target data is extracted from one model, and in order to compare with past PP studies, this is the same model as most of the results and discussion in Mann et al. (2007) are based on. The CFR techniques tested in Wang et al. (2014) are: the RegEM algorithm described by Schneider (2001) (regularized by ridge regression), the hybrid version of RegEM applied in Mann et al. (2009) (regularized by TTLS), canonical correlation analysis (CCA) and GraphEM. Experiment design involves varying SNR, but only using white proxy noise. Separate experiments are performed for networks with uniform/varying number of records in time. Using pseudoproxy records that are uniform in time is highly unrealistic, as observed in Fig.1 (b) in Wang et al. (2014). The unknown proxy-target variable relationship was also tested, and modeled differently in two additional experiments. Testing involved calculating the correlation coefficient ρ between each proxy in the Mann et al. (2008) network and the instrumental surface temperature field for each year between 1850-2006 AD. The results showed that some types of proxies were highly correlated with the local target, while a large fraction of other proxies had maximum correlation with a target value at great distance from the proxy location. To account for this bimodal distribution, separate experiments were conducted where the pseudoproxy networks were spatially distributed to simulate the locations with local/maximum correlation, respectively. The latter type of correlation exploits teleconnections or long-range dependence in space.

The results from Wang et al. (2014) show that when SNR is estimated from the maximum correlation network, average SNR is 0.47, but a few high-quality proxies have SNR higher than 1 and contribute largely to improving the total reconstruction skill. When temporal heterogeneities of the proxy network are taken into account, the reconstruction skill does not decrease monotonically back in time as might be expected. The type of climate variability (forced/internal) is also important for the skill, where for instance the higher solar/volcanic forcing during

the medieval warm period (MWP) contributes to increased skill. Regarding spatial skill, modeled teleconnections are shown to contribute to higher skill than expected in regions with low proxy coverage. Sensitivity to the reconstruction methodology is found to be low when data quality is high, and data quality is generally more important for the skill than varying availability of proxies with time. The Mann et al. (2009) implementation of RegEM provides the most skillful estimate of global mean surface temperature for both the local and maximum correlation experiments. For the spatial skill, CCA and GraphEM have the highest skill throughout the reconstruction period for the maximum correlation experiment, while for the local correlation, the Mann et al. (2009) and the TTLS RegEM implementations were most skillful. The variable skill estimates reflect that the four techniques have different sensitivity to input data and how the data are processed differently. Since there are many uncertainties associated with the proxies in a particular network, it is therefore suggested to use different techniques for reconstructing a climate field from a single proxy network.

Concluding remarks

Data from forced runs and control runs of paleoclimate model simulations exhibit LRM properties on time scales from years to centuries, with a spectral exponent β ranging from $0.6 < \beta < 1.1$. No systematic differences in scaling were found between forced runs and control runs of the millennium-long GCM and EMIC model simulations. The strength of persistence estimated from the model simulation time series is in line with that estimated from instrumental temperature records and paleoclimate reconstructions.

Analyzing selected temperature reconstructions, a centennial-scale change in scaling properties from $\beta \sim 0$ to $\beta > 1$ was not found for data covering only the Holocene time period. However, the break was evident when including data from the last glacial period. The issues of trends and universal scaling are problematic, since scaling analysis then must be performed on a signal which has time-dependent statistical properties.

Fields of LRM-processes with prescribed spatiotemporal covariance structure have been constructed using a novel technique, and pseudoproxy experiments show that a state-of-the-art climate field reconstruction (CFR) technique contributed to altered scaling properties in the resulting field and spatial mean reconstruction. The background assumption of the Bayesian BARCAST CFR technique, that the temperature follows an AR(1) structure, results in biased parameter estimates and thereby partly incorrect temperature reconstructions both locally and for the global mean.

Bibliography

Alley, R. B. and Ágústdóttir, A. M. The 8k event: cause and consequences of a major Holocene abrupt climate change. *Quaternary Science Reviews*, 24(10), 1123 – 1149, 2005. doi:http://dx.doi.org/10.1016/j.quascirev.2004.12.004.

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Buntgen, U., Cook, E. R., D'Arrigo, R. D., Esper, J., Evans, M. N., Frank, D., Grudd, H., Gunnarson, B. E., Hughes, M. K., Kirdyanov, A. V., Korner, C., Krusic, P. J., Luckman, B., Melvin, T. M., Salzer, M. W., Shashkin, A. V., Timmreck, C., Vaganov, E. A. and Wilson, R. J. S. Tree rings and volcanic cooling. *Nature Geosci.*, 5(12), 836–837, 2012. doi:10.1038/ngeo1645.

Bacry, E. and Muzy, J. Log-Infinitely Divisible Multifractal Processes. *Communications in Mathematical Physics*, 236(3), 449–475, 2003. doi:10.1007/s00220-003-0827-3.

Berner, K. S., Koc, N., Divine, D., Godtliebsen, F. and Moros, M. A decadal-scale Holocene sea surface temperature record from the subpolar North Aatlantic constructed using diatoms and statistics and its relation to other climate parameters. *Paleoceanography*, 23, 2008. doi:10.1029/2006PA001339.

Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R. and Zhang, X. Detection and Attribution of Climate Change: from Global to Regional, book section 10, pp. 867–952. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. doi:10.1017/CBO9781107415324.013. URL `www.climatechange2013.org`.

Blender, R., Fraedrich, K. and Hunt, B. Millennial climate variability: GCM-

simulation and Greenland ice cores. *Geophys. Res. Lett.*, 33(4), 2006. doi: 10.1029/2005GL024919.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B. and Zhao, Y. Evaluation of climate models using palaeoclimatic data. *Nature Clim. Change*, 2(6), 417–424, 2012. doi: 10.1038/NCLIMATE1456.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G. and Vaganov, E. A. Low-frequency temperature variations from a northern tree ring density network. *J. Geophys. Res.-Atmos*, 106(D3), 2929–2941, 2001. doi:10.1029/2000JD900617.

Bürger, G., Fast, I. and Cubasch, U. Climate reconstruction by regression - 32 variations on a theme. *Tellus*, 58, 227–235, 2006. doi:10.1111/j.1600-0870.2006.00164.x.

Christiansen, B. Reconstructing the NH Mean Temperature: Can Underestimation of Trends and Variability Be Avoided? *Journal of Climate*, 24(3), 674–692, 2011. doi:10.1175/2010JCLI3646.1.

Christiansen, B., Schmith, T. and Thejll, P. A Surrogate Ensemble Study of Climate Reconstruction Methods: Stochasticity and Robustness. *Journal of Climate*, 22(4), 951–976, 2009. doi:10.1175/2008JCLI2301.1.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D. and Worley, S. J. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654), 1–28, 2011. doi:10.1002/qj.776.

de Wit, T. D., Ermolli, I., Haberreiter, M., Kambezidis, H., Lam, M. M., Lilensten, J., Matthes, K., Mironova, I., Schmidt, H., Seppälä, A., Tanskanen, E., Tourpali, K. and Yair, Y., editors. Earth's climate response to a changing Sun: A review of the current understanding by the European research group TOSCA. EDP Sciences, Les Ulis Cedex A, France, 2015. 345 pp.

Ditlevsen, P. D., Svensmark, H. and Johnsen, S. Contrasting atmospheric and climate dynamics of the last-glacial and Holocene periods. *Nature*, 379(6434), 810–812, 1996. doi:10.1038/379810a0.

Driscoll, S., Bozzo, A., Gray, L. J., Robock, A. and Stenchikov, G. Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, 117(D17), n/a–n/a, 2012. doi:10.1029/2012JD017607.

Emile-Geay, J. and Tingley, M. Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies. *Climate of the Past*, 12(1), 31–50, 2016. doi:10.5194/cp-12-31-2016.

Esper, J., Cook, E. R. and Schweingruber, F. H. Low-Frequency Signals in Long Tree-Ring Chronologies for Reconstructing Past Temperature Variability. *Science*, 295(5563), 2250–2253, 2002. doi:10.1126/science.1066208.

Fernández-Donado, L., González-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P. and Zorita, E. Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Clim. Past*, 9(1), 393–421, 2013. doi:10.5194/cp-9-393-2013.

Franke, J., Frank, D., Raible, C. C., Esper, J. and Bronnimann, S. Spectral biases in tree-ring climate proxies. *Nature Clim. Change*, 3(4), 360–364, 2013. doi:10.1038/NCLIMATE1816.

Fredriksen, H.-B. and Rypdal, K. Spectral Characteristics of Instrumental and Climate Model Surface Temperatures. *Journal of Climate*, 29(4), 1253–1268, 2016. doi:10.1175/JCLI-D-15-0457.1.

Hasselmann, K. Stochastic climate models Part I. Theory. *Tellus*, 28(6), 473–485, 1976. doi:10.1111/j.2153-3490.1976.tb00696.x.

Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J. and Zorita, E. Detection of Human Influence on a New, Validated 1500-Year Temperature Reconstruction. *Journal of Climate*, 20(4), 650–666, 2007. doi:10.1175/JCLI4011.1.

Huang, S. Merging information from different resources for new insights into climate change in the past and future. *Geophys. Res. Lett.*, 31, 2004. doi: 10.1029/2004GL019781.

Huybers, P. and Curry, W. Links between annual, Milankovitch and continuum temperature variability. *Nature*, 441(7091), 2006. doi:10.1038/nature04745.

Jones, P. D., Briffa, K. R., Barnett, T. P. and Tett, S. F. B. 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. doi:10.1191/095968398667194956.

Laepple, T. and Huybers, P. Global and regional variability in marine surface temperatures. *Geophysical Research Letters*, 41(7), 2528–2534, 2014a. doi:10.1002/2014GL059345.

Laepple, T. and Huybers, P. Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *P. Natl. A. Sci.*, 111(47), 16682–16687, 2014b. doi:10.1073/pnas.1412077111.

Lovejoy, S. and Schertzer, D. Haar wavelets, fluctuations and structure functions: convenient choices for geophysics. *Nonlinear Proc. Geoph*, 19(5), 513–527, 2012a. doi:10.5194/npg-19-513-2012.

Lovejoy, S. and Schertzer, D. Low Frequency Weather and the Emergence of the Climate, pp. 231–254. 196. American Geophysical Union, 2012b. doi:10.1029/2011GM001087.

Løvsletten, O. Consistency of detrended fluctuation analysis. *Phys. Rev. E*, 96, 012141, 2017. doi:10.1103/PhysRevE.96.012141.

Lowe, J. J. and Walker, M. J. C. Recontructing Quaternary Environments. Addison Wesley Longman, Harlow, England, 1997. 446 pp.

Luterbacher, J., Werner, J. P., Smerdon, J. E., Fernández-Donado, L., González-Rouco, F. J., Barriopedro, D., Ljungqvist, F. C., Büntgen, U., Zorita, E., Wagner, S., Esper, J., McCarroll, D., Toreti, A., Frank, D., Jungclaus, J. H., Barriendos, M., Bertolin, C., Bothe, O., Brázdil, R., Camuffo, D., Dobrovolný, P., Gagen, M., García-Bustamante, E., Ge, Q., Gómez-Navarro, J. J., Guiot, J.,

Hao, Z., Hegerl, G. C., Holmgren, K., Klimenko, V. V., Martín-Chivelet, J., Pfister, C., Roberts, N., Schindler, A., Schurer, A., Solomina, O., von Gunten, L., Wahl, E., Wanner, H., Wetter, O., Xoplaki, E., Yuan, N., Zanchettin, D., Zhang, H. and Zerefos, C. European summer temperatures since Roman times. *Environmental Research Letters*, 11(2), 2016. doi:10.1088/1748-9326/11/2/024001.

Mandelbrot, B. How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science*, 156(3775), 636–638, 1967.

Mann, M. E., Bradley, R. S. and Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392(6678), 779–787, 1998. doi:10.1038/33859.

Mann, M. E., Bradley, R. S. and Hughes, M. K. Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.*, 26(6), 759–762, 1999. doi:10.1029/1999GL900070.

Mann, M. E., Fuentes, J. D. and Rutherford, S. Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nature Geosci*, 107(3-4), 202–205, 2012. doi:10.1038/NGEO1394.

Mann, M. E. and Rutherford, S. Climate reconstruction using ?Pseudoproxies? *Geophysical Research Letters*, 29(10), 139–1–139–4, 2002. doi:10.1029/2001GL014554.

Mann, M. E., Rutherford, S., Schurer, A., Tett, S. F. and Fuentes, J. D. Discrepancies between the modeled and proxy-reconstructed response to volcanic forcing over the past millennium: Implications and possible mechanisms. *Journal of Geophysical Research: Atmospheres*, 118(14), 7617–7627, 2013. doi:10.1002/jgrd.50609.

Mann, M. E., Rutherford, S., Wahl, E. and Ammann, C. Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate. *Journal of Climate*, 18(20), 4097–4107, 2005. doi:10.1175/JCLI3564.1.

Mann, M. E., Rutherford, S., Wahl, E. and Ammann, C. Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, 112, 2007.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S. and Ni, F. Proxy-based reconstructions of hemispheric and global sur-

face temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 105(36), 13252–13257, 2008. doi:10.1073/pnas.0805721105.

Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G. and Ni, F. Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, 326, 1256–1260, 2009. doi:10.1126/science.1177303.

Marcott, S. A. a. D. S., Clark, P. U. and Mix, A. C. A Reconstruction of Regional and Global Temperature for the Past 11,300 Years. *Science*, 339(6124), 1198–1201, 2013. doi:10.1126/science.1228026.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González-Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X. and Timmermann, A. Information from Paleoclimate Archives, book section 5, pp. 383–464. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN ISBN 978-1-107-66182-0. doi:10.1017/CBO9781107415324.013. URL `www.climatechange2013.org`.

McGregor, H. V., Francus, P., Abram, N., Evans, M. N., Goosse, H., von Gunten, L., Kaufman, D., Linderholm, H., Loutre, M. F., Neukom, R. and Turney, C., editors. Climate of the past 2000 years: regional and trans-regional syntheses. Climate of the Past, issued by Copernicus Publications on behalf of the European Geosciences Union, Göttingen, Germany, 2016.

Miettinen, A., Divine, D., Koç, N., Godtliebsen, F. and Hall, I. R. Multicentennial Variability of the Sea Surface Temperature Gradient across the Subpolar North Atlantic over the Last 2.8 kyr. *Journal of Climate*, 25(12), 4205–4219, 2012. doi:10.1175/JCLI-D-11-00581.1.

Miller, G. H., Geirsdóttir, A., Zhong, Y., Larsen, D. J., Otto-Bliesner, B. L., Holland, M. M., Bailey, D. A., Refsnider, K. A., Lehman, S. J., Southon, J. R., Anderson, C., Björnsson, H. and Thordarson, T. Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophysical Research Letters*, 39(2), n/a–n/a, 2012. ISSN 1944-8007. doi:10.1029/2011GL050168. L02708, URL `http://dx.doi.org/10.1029/2011GL050168`.

Mills, T. C. Time series modelling of two millennia of northern hemisphere temperatures: long memory or shifting trends? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1), 83–94, 2007. doi: 10.1111/j.1467-985X.2006.00443.x.

Moberg, A., Mohammad, R. and Mauritsen, T. Analysis of the Moberg et al. (2005) hemispheric temperature reconstruction. *Climate Dynamics*, 31(7), 957–971, 2008.

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M. and Karlén, W. Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data. *Nature*, 433, 613–617, 2005. doi: 10.1038/nature03265.

Mondal, D. and Percival, D. B. Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics*, 62(5), 943–966, 2010. doi: 10.1007/s10463-008-0195-z.

National Science Foundation. NSF Press release 15-048. 2015. Accessed: 08-09-2017, URL `http://icecores.org/indepth/2015/spring/antarctic-ice-core-reveals-how-sudden-climate-changes-in-north-atlantic-moved-south.shtml`.

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., Gonzalez-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., Raible, C. C., Steig, E. J., van Ommen, T., Vance, T., Villalba, R., Zinke, J. and Frank, D. Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change*, 4(5), 362–367, 2014. doi:10.1038/nclimate2174.

Nilsen, T., Divine, D. and Werner, J. P. Pseudoproxy experiments using LRM processes, 2017. To be submitted to Climate of the Past.

Nilsen, T., Rypdal, K. and Fredriksen, H.-B. Are there multiple scaling regimes in Holocene temperature records? *Earth System Dynamics*, 7(2), 419–439, 2016. doi:10.5194/esd-7-419-2016. URL `https://www.earth-syst-dynam.net/7/419/2016/`.

Østvand, L. Long Range Memory in Time Series of Earth Surface Temperature. Ph.D. thesis, University of Tromsø, the Arctic University of Norway, 2014.

Østvand, L., Nilsen, T., Rypdal, K., Divine, D. and Rypdal, M. Long-range memory in internal and forced dynamics of millennium-long climate model simulations. *Earth Sys. Dyn.*, 5, 295–308, 2014. doi:10.5194/esd-5-295-2014.

PAGES 2k Consortium. Continental-scale temperature variability during the past two millennia. *Nature Geosci*, 6(5), 339–346, 2013. doi:10.1038/ngeo1797.

PAGES 2k Consortium. A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, 4, 170088 EP –, 2017. doi:10.1038/sdata.2017.88.

PAGES 2k-PMIP3 group. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Clim. Past*, 11(12), 1673–1699, 2015. doi:10.5194/cp-11-1673-2015.

Pelletier, J. D. The power spectral density of atmosphereic temperature from time scales of 10^{-2} to 10^{6} yr. *Earth and Planetary Science Letters*, 158, 157–164, 1998. doi:10.1016/S0012-821X(98)00051-X.

Robock, A. and Mao, J. The Volcanic Signal in Surface Temperature Observations. *Journal of Climate*, 8(5), 1086–1103, 1995. doi:10.1175/1520-0442(1995)008<1086:TVSIST>2.0.CO;2.

Rutherford, S., Mann, M. E., Osborn, T. J., Briffa, K. R., Jones, P., Bradley, R. S. and Hughes, M. K. Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain. *Journal of Climate*, 18(13), 2308–2329, 2005. doi:10.1175/JCLI3351.1.

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

Rypdal, K. Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions. *J. Geophys. Res.*, 117, D06115, 2012. doi:10.1029/2011JA017283.

Rypdal, K., Østvand, L. and Rypdal, M. Long-range memory in Earth's surface temperature on time scales from months to centuries. *J. Geophys. Res.*, 118(13), 7046–7062, 2013. doi:10.1002/jgrd.50399.

Rypdal, K., Rypdal, M. and Fredriksen, H.-B. Spatiotemporal Long-Range Persistence in Earth's Temperature Field: Analysis of Stochastic–Diffusive Energy Balance Models. *Journal of Climate*, 28(21), 8379–8395, 2015. doi: 10.1175/JCLI-D-15-0183.1.

Rypdal, M. and Rypdal, K. Testing Hypotheses about Sun-Climate Complexity Linking. *Phys. Rev. Lett.*, 104(12), 128501–4, 2010. doi:10.1103/PhysRevLett. 104.128501.

Rypdal, M. and Rypdal, K. Long-memory effects in linear-response models of Earth's temperature and implications for future global warming. *J. Climate*, 27(14), 2014. doi:10.1175/JCLI-D-13-00296.1.

Rypdal, M. and Rypdal, K. Late Quaternary temperature variability described as abrupt transitions on a $1/f$ noise background. *Earth Sys. Dynam.*, 7, 281–293, 2016. doi:10.5194/esd-7-281-2016.

Schleussner, C.-F., Divine, D. V., Donges, J. F., Miettinen, A. and Donner, R. V. Indications for a North Atlantic ocean circulation regime shift at the onset of the Little Ice Age. *Climate Dynamics*, 45(11), 3623–3633, 2015. doi:10.1007/s00382-015-2561-x.

Schneider, T. Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *Journal of Climate*, 14(5), 853–871, 2001. doi:10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2.

Shao, Z.-G. and Ditlevsen, P. D. Contrasting scaling properties of interglacial and glacial climates. *Nature Communications*, 7, 2016. doi:10.1038/ncomms10951.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Buntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schupbach, S., Steffensen, J. P., Vinther, B. M. and Woodruff, T. E. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, 523(7562), 543–549, 2015. doi: 10.1038/nature14565.

Smerdon, J. E. Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *Wiley Interdisciplinary Reviews: Climate Change*, 3(1), 63–77, 2012. doi:10.1002/wcc.149.

Smerdon, J. E., Kaplan, A., Chang, D. and Evans, M. N. A Pseudoproxy Evaluation of the CCA and RegEM Methods for Reconstructing Climate Fields of the Last Millennium. *Journal of Climate*, 24(4), 1284–1309, 2011. doi: 10.1175/2010JCLI4110.1.

Tikhonov, A. and Arsenin, V. Solutions of ill-posed problems. Scripta series in mathematics. Winston, 1977. ISBN 9780470991244.

Tingley, M. P. and Huybers, P. A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part I: Development and Applications to Paleoclimate Reconstruction Problems. *Journal of Climate*, 23(10), 2759–2781, 2010a. doi:10.1175/2009JCLI3015.1.

Tingley, M. P. and Huybers, P. A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part II: Comparison with the Regularized Expectation Maximization Algorithm. *Journal of Climate*, 23(10), 2782–2800, 2010b. doi:10.1175/2009JCLI3016.1.

Tingley, M. P. and Huybers, P. Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature*, 496(7444), 201–205, 2013.

von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F. and Tett, S. F. B. Reconstructing Past Climate from Noisy Data. *Science*, 306(5696), 679–682, 2004. doi:10.1126/science.1096109.

Vyushin, D., Zhidkov, I., Havlin, S., Bunde, A. and Brenner, S. Volcanic forcing improves Atmosphere-Ocean Coupled General Circulation Model scaling performance. *Geophysical Research Letters*, 31(10), n/a–n/a, 2004. doi: 10.1029/2004GL019499. L10206.

Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E. and Rajaratnam, B. Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate of the Past*, 10(1), 1–19, 2014. doi:10.5194/cp-10-1-2014.

Werner, J. P., Divine, D. V., Ljungqvist, F. C., Nilsen, T. and Francus, P. Spatio-temporal variability of Arctic summer temperatures over the past two millennia: an overview of the last major climate anomalies. *Climate of the Past Discussions*, 2017, 1–43, 2017. doi:10.5194/cp-2017-29.

Werner, J. P., Luterbacher, J. and Smerdon, J. E. A Pseudoproxy Evaluation of Bayesian Hierarchical Modeling and Canonical Correlation Analysis for Climate Field Reconstructions over Europe*. *J. Climate*, 26(3), 851–867, 2013. doi: 10.1175/JCLI-D-12-00016.1.

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., Hegerl, G., Robock, A., Pausata, F. S. R., Ball, W. T., Bauer, S. E., Bekki, S., Dhomse, S. S., LeGrande, A. N., Mann, G. W., Marshall, L., Mills, M., Marchand, M., Niemeier, U., Poulain, V., Rozanov, E., Rubino, A., Stenke, A., Tsigaridis, K. and Tummon, F. The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): experimental design and forcing input data for CMIP6. *Geoscientific Model Development*, 9(8), 2701–2719, 2016. doi:10.5194/gmd-9-2701-2016.

Zorita, E., González-Rouco, F. and Legutke, S. Testing the Approach to Paleoclimate Reconstructions in the Context of a 1000-Yr Control Simulation with the ECHO-G Coupled Climate Model. *Journal of Climate*, 16(9), 1378–1390, 2003. doi:10.1175/1520-0442(2003)16<1378:TTMEAA>2.0.CO;2.

Zorita, E., González-Rouco, F. and von Storch, H. Comments on "Testing the Fidelity of Methods Used in Proxy-Based Reconstructions of Past Climate". *Journal of Climate*, 20(14), 3693–3698, 2007. doi:10.1175/JCLI4171.1.