The Structure of Climate Variability Across Scales

Journal Item

The Structure of Climate Variability Across Scales

Christian L. E. Franzke[1,2], Susana Barbosa[3], Richard Blender[1,2], Hege-Beate Fredriksen[4], Thomas Laepple[5], Fabrice Lambert[6], Tine Nilsen[7,8], Kristoffer Rypdal[7], Martin Rypdal[7], Manuel G. Scotto[9], Stéphane Vannitsem[10], Nicholas W. Watkins[11,12,13], Lichao Yang[14,2], Naiming Yuan[15]

[1]Meteorological Institute, University of Hamburg, Hamburg, Germany

[2]Center for Earth System Research and Sustainability, University of Hamburg, Hamburg, Germany

[3]INESC TEC, Centre for Information Systems and Computer Graphics, Porto, Portugal

[4]Department of Physics and Technology, UiT the Arctic University of Norway, Tromso, Norway

[5]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

[6]Geography Institute, Pontificia Universidad Catolica de Chile, Santiago, Chile

[7]Department of Mathematics and Statistics, UiT the Arctic University of Norway, Tromso, Norway

[8]Department of Geography, Justus-Liebig University of Giessen, Germany

[9]CEMAT and Department of Mathematics, IST University of Lisbon, Lisbon, Portugal

[10]Royal Meteorological Institute of Belgium, Brussels, Belgium

[11]London School of Economics, London, UK

[12]Open University, Milton Keynes, UK

[13]University of Warwick, Coventry, UK

[14]Department of Atmospheric and Oceanic Sciences, Peking University, Beijing, China

[15]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Key Points:

- Climate variability operates on a continuum of spatial and temporal scales in such a way that the variability exhibits scaling relationships
- Climatologically relevant imprints of scaling include Long-Range Dependence and non-Gaussian fluctuations
- Scaling has implications for trend detection, climate sensitivity and predictability

Corresponding author: Christian L. E. Franzke, christian.franzke@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019RG000657

Abstract

One of the most intriguing facets of the climate system is that it exhibits variability across all temporal and spatial scales; pronounced examples are temperature and precipitation. The structure of this variability, however, is not arbitrary. Over certain spatial and temporal ranges it can be described by scaling relationships in the form of power-laws in probability density distributions and autocorrelation functions. These scaling relationships can be quantified by scaling exponents which measure how the variability changes across scales and how the intensity changes with frequency of occurrence. Scaling determines the relative magnitudes and persistence of natural climate fluctuations. Here, we review various scaling mechanisms and their relevance for the climate system. We show observational evidence of scaling and discuss the application of scaling properties and methods in trend detection, climate sensitivity analyses, and climate prediction.

Plain Language Summary

Climate variables are related over long times and large distances. This shows up as correlations for averages on long intervals or between distant areas. An important finding is that the majority of correlations in climate can be described by a simple mathematical relationship. We present such correlations for temperature on long times. Similarly, the intensity of precipitation events depends on their frequency in a simple manner. A useful concept is scaling where a scale denotes the width of an average. Scaling says that averages on different scales are related by a simple function – mathematically this is a power-law with the scaling exponent as a characteristic number. Scaling has impacts on predictability, temperature trends and the assessment of future climate changes caused by anthropogenic forcing.

1 Introduction

An emerging topic in climate science is the systematic change of the temporal and spatial structure of climate variability seen across a multitude of spatial and temporal scales, in particular power-law behavior [e.g. *Hurst*, 1951; *Mandelbrot and Wallis*, 1968; *Huybers and Curry*, 2006; *Lovejoy and Schertzer*, 2013; *Graves et al.*, 2017a]. The intensity distribution of climate variables in relation to their frequency of occurrence also shows such power-law behavior. It is of importance to improve our understanding of the underlying structure of climate variability since this may potentially allow us not only to

improve our predictive capabilities but also contribute to an improved overall understanding of the complex Earth system as a whole. The presence of power-law behavior in both the temporal and spatial domains and in intensities can reveal aspects of the underlying dynamics of the Earth system such as climate sensitivity and predictability.

This behavior can be illustrated with two climatological time series (Fig. 1). Our choice of precipitation data (Fig. 1a) exhibits the typical intermittent behavior with no or only very little precipitation on most days interspersed with an occasional extreme event. Hence, precipitation is a climatological variable that is highly episodic. Consequently, the distribution of precipitation is much more heavy-tailed than a Gaussian distribution (Fig. 1b). Thus, large values are much more likely than in the case of variables that are Gaussian distributed; The Gaussian distribution decays much faster than a power-law. The tails of many precipitation distributions, as well as of other climatological quantities, decay according to a power-law (see Section 1.2 for details). This power-law relation between intensity and probability of occurrence constitutes a scaling relationship.

As a second time series we present the Central England Temperature (CET) [*Parker et al.*, 1992] time series for the period 1772-2017. The CET consists of observations from stations located throughout central England. In Fig. 1c we show the annual mean time series overlayed by an 11-year running mean and the non-linearly filtered decadal-scale CET data using Empirical Mode Decomposition (EMD) [*Huang et al.*, 1998; *Huang and Wu*, 2008; *Franzke*, 2009]. EMD allows for a systematic decomposition of time series into dynamically relevant oscillatory modes and a non-linear trend. The CET time series exhibits decadal-scale variations about an instantaneous mean [*Franzke*, 2009]. The observed decadal-scale variability is a visible imprint of the scaling and Long-Range Dependence (LRD) [e.g. *Gil-Alana*, 2008; *Graves et al.*, 2015]. Intuitively, Long-Range Dependence has the property that spatially coherent anomalies persist for a long time, e.g., heat waves or droughts may last for many years [*Cook et al.*, 2015], which is indicative of a decay of serial correlation which is slower than exponential, e.g. power-law decay. Long-range dependence means that positive (negative) anomalies are very likely followed by positive (negative) anomalies for long periods of time. The decay of serial correlations of long-range dependent systems behaves according to a power-law (Fig. 1d and e) as can be shown by an analysis using Detrended Fluctuation Analysis (see Sec. 2.6.3). This approach provides more robust estimates than the standard autocorrelation function, which can be noisy at long lags (Fig. 1e). In brief, this method computes the variance for

moving windows of different sizes which yields a scaling relationship for the correlation strength of values at different times.

To summarize, many climatological time series exhibit a power-law behavior in their amplitudes or their autocorrelations or both. This behavior is an imprint of scaling, which is a fundamental property of many physical and biological systems and has also been discovered in financial and socio-economic data as well as in information networks [*Mandelbrot*, 1963; *Mantegna and Stanley*, 1999; *Clauset et al.*, 2009; *Willinger et al.*, 2004; *Saichev et al.*, 2009; *Ball*, 2003]. While the power-law has no preferred scale, the exponential function, also ubiquitous in physical and biological systems, does have a preferred scale, namely the e-folding scale, i.e., the amount by which its magnitude has decayed by a factor of $\exp(-1)$. For example, the average height of humans is a good predictor for the height of the next person you meet as there are no humans that are 10 times larger or smaller than you. However, the average wealth of people is not a good predictor for the wealth of the next person you meet as there are people who can be more than a 1000 times richer or poorer than you are. Hence, the height of people is well described by a Gaussian distribution, while the wealth of people follows a power-law [*Newman*, 2005].

Furthermore, an fascinating aspect of scaling in the climate system is that it occurs in many different characteristics of climate variables. As demonstrated above it exists in time and intensity and, as we will discuss below, in space. For instance, negative vorticity anomalies, such as blocking can be very persistent [e.g. *Feldstein and Franzke*, 2017], while positive vorticity anomalies, such as storms, have a heavy-tailed probability distribution of intensities [*Corral et al.*, 2010; *Blender et al.*, 2016] and heavy-tailed waiting time distributions [*Franzke*, 2013; *Yang et al.*, 2019]. Persistence and heavy-tailed distributions are described by scaling relationships. Different dynamical regimes are likely causing the scaling properties in the intensity, time and space. In section 2.5 we discuss potential physical mechanisms which can explain scaling in the climate system. While there have been many mechanisms discussed in the literature [e.g. *Beran*, 1994; *Beran et al.*, 2013], their applicability to the climate system is still an open question.

While the existence of scaling has been known for a long time and across many scientific areas, it had been largely ignored for an almost equally long time in the analysis of climate data, with some exceptions [e.g. *Gil-Alana*, 2003; *Vyushin et al.*, 2004; *Koscielny-Bunde et al.*, 1998; *Blender and Fraedrich*, 2003; *Mann*, 2011; *Franzke*, 2012; *Dangendorf*

et al., 2014; *Becker et al.*, 2014]. Only recently has its usefulness been more widely appreciated in climate science, partly due to its inclusion in text books (e.g. *Chandler and Scott* [2011]; *Mudelsee* [2013]; *Lovejoy and Schertzer* [2013]; *Schmitt and Huang* [2016]) and partly due to the establishment of working groups such as Climate Variability Across Scales (CVAS), part of Past Global Changes (PAGES), who employ scaling approaches to improve our understanding of the complexities of the Earth system [see e.g. *Crucifix et al.*, 2017].

These scaling ideas enter the climate sciences from theoretical physics, applied mathematics, statistics, and theoretical climatology. They are rarely taught in standard meteorology, oceanography, or climate science courses. Here, we aim to bridge these disciplinary gaps by introducing the main ideas in a manner that is accessible and applicable for climate scientists.

1.1 Scales in the climate system

One of the fascinating aspects of the climate system is the close relationship between the spatial and temporal scales of the relevant physical processes. This accounts for the success of scaling analyses of the equations of motion and the systematic derivation of simplified versions of the primitive equations, such as the quasi-geostrophic or the shallow-water equations [e.g. *Vallis*, 2017; *Majda and Wang*, 2006; *Klein*, 2010; *Franzke et al.*, 2019]. For instance, the quasi-geostrophic equations are valid in the limit of a small Rossby number [*Vallis*, 2017] and describe Rossby and synoptic-scale waves and, thus, provide an excellent conceptual model to understand many important aspects of the atmosphere and ocean.

The many physical processes in the Earth's climate system span a vast dynamic range, both in space (from 10^{-3} m to 10^{7}m) and time (from seconds to millions of years) (Fig. 2). *Williams et al.* [2017] provide a census of atmospheric processes, the variability of which range from seconds to decades. In the climate system, we typically deal with the following physical processes and associated scales: turbulent eddies on time scales of a few seconds and length scales of millimeters to centimeters; convective activity on temporal scales of hours and spatial scales of hundreds of meters to a few kilometers; synoptic weather systems varying diurnally on spatial scales of hundreds to thousands of kilometers; large-scale teleconnection patterns with an intra-seasonal to inter-annual temporal

variability and spatial scales that can span an entire hemisphere; the coupled atmosphere-ocean system which varies from decadal to centennial time scales and a global spatial scale; the ice ages represent global variations on millennial time scales (Fig. 2). The main four components of the climate system (atmosphere, ocean, land, and cryosphere) tend to operate on different time scales which interact non-linearly with each other creating a plethora of interesting effects and feedbacks [*Rial et al.*, 2004; *Peters et al.*, 2004; *Williams et al.*, 2017].

An intriguing property of the climate system is that despite the fact that we have to deal with many different physical processes, the variability constitutes a continuum of fluctuations, i.e., while the variability spectrum may be interspersed by spikes belonging to some particular and well defined forcing process (e.g., daily, annual, or Milankovich cycles) the vast part of the spectrum is continuous and scales over large ranges.

1.2 Power-Law Scaling

By scaling we mean the power-law relationship between the amplitude of fluctuations and their probability of occurrence on a given temporal or spatial scale:

$$f(ay) = a^{\gamma} f(y) \tag{1}$$

where f is an arbitrary function which can either be deterministic or stochastic, y is a climate variable or time, and γ denotes the scaling exponent, a factor which allows us to zoom in and out. In case of f being a stochastic function the equality has to be interpreted as equality in distribution. When considering a time series, f is a stochastic process and Eq. (1) implies that the variability of short time scales is statistically similar to the variability on longer time scales. This also implies that no preferred time scale exists. Furthermore, this equation describes a self-similar process [*Lamperti*, 1962]; if y would denote time, then Eq. (1) would imply that the variance would go to infinity for increasing time scales. Furthermore, the fact that climate data exhibit scaling, indicates that the statistical properties remain independent of the scale [*Kolmogorov*, 1940; *Hurst*, 1951; *Lamperti*, 1962; *Mandelbrot and Van Ness*, 1968; *Mandelbrot*, 1982; *Feder*, 1988; *Franzke et al.*, 2012; *Taqqu*, 2013] as is the case for fractals [*Feder*, 1988]. The scaling property might already be a familiar concept from power spectrum analyses where, in addition to pronounced peaks, one also examines for the existence of linear slopes in a double logarithmic scale representation [e.g. *Wunsch*, 2003; *Huybers and Curry*, 2006].

In Fig. 3 we display time series sample paths in order to illustrate the scaling property; these were generated from an Autoregressive Fractional Integrated Moving Average (ARFIMA) scaling model (see Section 2.4 and appendix A:). the displayed Long-Range Dependence process has a slope of 0.75 in a log-log plot of fluctuation function versus scale while a Short-Range Dependence (SRD) process has a slope of 0.5 at long time scales. A slope of 0.5 corresponds to white noise which means that the process is uncorrelated (Fig. 3c). The power spectrum (Figs. 3d) exhibits the corresponding behavior of increasing power for lower frequencies (with a singularity at zero) of a Long-Range Dependence process exhibiting while the SRD spectrum is flat at low frequencies. Scaling in intensities is displayed in Fig. 4 for the α-stable distribution.

1.3 Climate variability across scales

The first attempt to conceptualize atmospheric variability over a wide range of scales has been made by *Mitchell* [1976]. Mitchell's ambitious composite spectrum (Fig. 5) ranged from hours to the age of the Earth and focused on the peaks in the power spectrum, thus emphasizing the quasi-periodic phenomena in the climate system and its forcings. Although *Mitchell* [1976] made a candid admission that his spectrum was mostly an 'educated guess', and despite subsequent improvements in climate and paleoclimate data, the original work has achieved almost iconic status.

Mitchell's scale-bound view led to a climate dynamics framework that emphasizes the importance of numerous processes occurring at well-defined time scales and the separation into quasi-periodic 'foreground' processes (illustrated as sharp peaks in Fig. 5) and the 'unimportant background noise'. We argue that while this division is not wrong per se, it can only explain a small fraction of the overall variability and the underlying climate system dynamics. *Wunsch* [2003] showed that the quasi-periodic signals represent only a small fraction of the total variability which is more akin to a Lorentzian spectrum of an autoregressive process while *Pelletier* [1997] and *Huybers and Curry* [2006] put an emphasis on the power-law behavior of the background spectrum.

Lovejoy and Schertzer [2013] and *Lovejoy* [2015a] postulated the existence of 5 distinct power-law scaling regimes. These regimes are based on different scaling exponents for the relationship $E(\omega) \sim \omega^{-\beta}$, where E denotes the spectral energy and ω frequency [*Huybers and Curry*, 2006]. The proposed regimes are

1. the weather regime with time scales from 6 hours up to 20 days with an exponent of $\beta \approx 1.8$
2. the macro-weather regime with time scales between 20 days and 50 years and β=0.2
3. the climate regime with time scales between 50 and 80,000 years (includes glacial-interglacial cycles) and β=1.8
4. the macro-climate regime between 80,000 and 500,000 years and β=-0.6
5. the mega-climate regime for time scales larger than 500,000 years which takes us to the limit of reliable proxies [*Lovejoy and Schertzer*, 2013] and β=1.8.

See Fig. 2a of *Lovejoy* [2015a] for an illustration of the scaling regimes.

Some recent studies focused more on the continuum aspects of the spectra [*Pelletier*, 1998; *Paillard*, 2001; *Huybers and Curry*, 2006]. For instance, *Huybers and Curry* [2006] reported qualitatively similar results for the macro-weather and climate regimes, while *Nilsen et al.* [2016] provided quantitative evidence that supports the hypothesis of just one scaling regime at least for the Holocene. *Nilsen et al.* [2016] also question whether it is meaningful to classify climate variability into universal regimes on time scales where we observe forced global climate changes, and in particular geological time scales. The reason is that the variability on the long time scales is fundamentally forced by time-dependent external processes, e.g. the Milankovich cycle, hence its statistics are time-varying [*Nilsen et al.*, 2016]. On shorter temporal scales, on the other hand, scaling is better established in many climatic data sets for a wide range of spatial, and intensity ranges. Furthermore, it has been recognized that quasi-periodic signals represent only a small fraction of the total climate variability, and while many studies have focused on understanding these quasi-periodic signals, we argue that the continuous variance spectrum is of equal significance and deserving of future research efforts.

1.4 Scope of the review

Because scales and scaling properties in the climate system are hard to adequately cover in a single paper, we will restrict this review to topics relevant to the interpretation and reconstruction of time series and to the impacts of scaling on climate variability, trends, prediction, and climate sensitivity. While we cover the potential physical mechanisms behind scaling, we can only provide a broad and non-rigorous introduction to the mathematical framework of scaling processes. More rigorous treatments can be found

elsewhere [*Beran*, 1994; *Beran et al.*, 2013; *Baillie*, 1996; *Samorodnitsky*, 2007, 2016; *Embrechts and Maejima*, 2007; *Lovejoy and Schertzer*, 2013; *Palma*, 2007; *Guegan*, 2005; *Doukhan et al.*, 2002].

Our review is structured as follows: Section 2 covers the basic ideas of scaling and estimation methods; Section 3 provides empirical evidence of scaling in climatic time series; Section 4 discusses applications of scaling like trend detection, climate prediction, and climate sensitivity. We end with an outlook and open research questions in Section 5.

2 Basic Concepts Related to Scaling Relationships

In this section we provide a brief review of the mathematical and physical background to scaling, with an emphasis on an intuitive understanding of the main ideas, leaving the details to the specialist literature.

2.1 Scaling and power laws

2.1.1 Scaling from dimensional analysis

In the physical sciences, scaling is a well-known and long established concept [*Longair*, 2003; *Bolster et al.*, 2011; *Watkins et al.*, 2016]. For instance, scaling can be used to explain: (i) how a pendulum's angular frequency depends on its length, or (ii) how the gravitational force between two bodies depends on their distance from one another.

In the first example, the angular frequency ω depends on the length l as

$$\omega = 2\pi\sqrt{\frac{g}{l}} \sim l^{-\frac{1}{2}} \tag{2}$$

where g is the gravitational acceleration. In the second example, Newton's law of universal gravitation states that the gravitational force, F, between two bodies with masses m_1 and m_2, is inversely proportional to the square of the distance between their centers, r, as:

$$F = G\frac{m_1 m_2}{r^2} \sim r^{-2}. \tag{3}$$

where G is the universal gravitational constant. The scaling property in both examples is so well established that it can be used to extrapolate and to test the behavior of systems outside their initial observable range. It can easily be seen that Eqs. (2) and (3) are different forms of the power law from Eq. (1) with γ equal to $-\frac{1}{2}$ and -2, respectively.

While originally a result of empirical observation, the above equations can also be derived from dimensional analyses. This embodies the physical principle of similarity, which requires that (natural) physical laws should be independent of (human) physical units used to describe a system. According to Buckingham's Π theorem [*Buckingham*, 1914; *Meinsma*, 2019], dimensional analysis can be used to show that any physical equation involving n variables can be rewritten using $n - m$ dimensionless parameters, where $m \geq 0$, thus revealing possible scaling relations which can then be empirically tested [*Bolster et al.*, 2011].

Dimensional analysis remains a very powerful technique for systems which resist analytic or numerical treatment. The prime example is geophysical fluid turbulence. In 1941 Kolmogorov [*Kolmogorov*, 1991a,b] derived a scaling relationship between turbulent kinetic energy E and the horizontal scale as measured by wavenumber k for isotropic turbulence. Thereby, he derived the *Kolmogorov -5/3 spectrum* (for details and underlying assumptions see *Vallis* [e.g. 2017]):

$$E(k) \sim k^{-5/3} \tag{4}$$

While a power-law distribution of the energy spectrum has been confirmed by observational evidence in the atmosphere [*Nastrom and Gage*, 1985; *Straus and Ditlevsen*, 1999], the exact exponent is still a matter of debate [*Lovejoy et al.*, 2007; *Lovejoy and Schertzer*, 2013]. For instance, *Lovejoy et al.* [2007] have shown that the atmosphere is anisotropic with different scaling exponents in the horizontal and vertical directions; which violates Kolmogorov's assumption of isotropy. Also the theoretical $-5/3$ scaling for large horizontal scales is -2.4 according to aircraft measurements [*Lovejoy et al.*, 2009]. This does not invalidate dimensional analysis but only shows that some of the underlying assumptions made by Kolmogorov in his first model (homogeneous and isotropic 3-dimensional turbulence) describe an idealized system but are typically not valid in the real atmosphere or ocean, where vertical stratification, jet streams and the presence of boundaries prevents full isotropy and homogeneity.

Another example of scaling is the addition of N random numbers, where the standard error scales as $\sigma_N \propto N^{1/2}$, a result familiar to all scientists from the undergraduate laboratory and the treatment of experimental errors [e.g. *Wilks*, 2011]. Interestingly, this result can be connected to a physical situation, by considering the root mean square of the displacement y_N from the origin of the first N steps of a random walk, which is one of

the most basic stochastic models for a time series. In a typical one-dimensional discrete random walk, a particle may start at a location and each step moves it either to the left or to the right with equal probability. The resulting root mean square of the total displacement, y, after N steps scales with $N^{1/2}$ which can also be expressed in terms of time, t, as:

$$\sqrt{y^2 - < y >^2} \sim t^{1/2}. \tag{5}$$

This describes the growth of the diffusing edge of a particle cloud executing Brownian motion (See appendix B:) [*Bouchaud and Potters*, 2003]. The random walk model is statistically self-similar, i.e., the time series generated by a random walk looks approximately the same as parts of it. In other words, the shapes and behaviors of the time series are independent of the time scale under consideration. Mathematically, statistical self-similarity can be written as

$$X(at) \stackrel{d}{=} a^{\gamma_{SS}} X(t) \tag{6}$$

and is equivalent to the scaling relationship in Eq. 1 where $\stackrel{d}{=}$ refers to that both sides are equally distributed. Here, γ_{SS} is the self-similarity parameter. In some processes, such as fractional Brownian motion, this is identical to the Hurst exponent H. The Hurst exponent H is named after Harold Edwin Hurst who first identified a scaling relationship investigating the flow levels of the Nile river and other reservoirs [*Hurst*, 1951, 1957; *Doukhan et al.*, 2002]. He developed the R/S method (see details below in appendix C.1 and appendix D:) to estimate the scaling exponent. A list of used exponents is given in Tab. 1.

The range of problems we can handle with scaling analysis can be greatly broadened if we introduce the concept of fractals by considering scaling exponents γ which are non-rational. Just as in the integer or rational cases, there is physically instructive information in fractal exponents that can go beyond that from dimensional analysis. These non-rational exponents will play an important role from now on since they are necessary to describe the observed scaling in climate time series due to Long-Range Dependence and heavy-tailed Probability Density Functions. They will be discussed in the following subsections.

2.2 Scaling in Probability Density Functions and Non-Gaussianity

2.2.1 Non-Gaussian but stable Probability Density Functions

The Central Limit Theorem states that the sum of independent and identically distributed random variables with finite variance approaches a Gaussian distribution and re-

sults in an $N^{1/2}$ scaling, where N is the length of the sums [*von Storch and Zwiers*, 2003; *Wilks*, 2011]. However, many natural systems, e.g., precipitation (Fig. 1) [*Peters et al.*, 2001, 2010; *Yang et al.*, 2019] and the Greenland ice cores [*Ditlevsen*, 1999; *Gairing et al.*, 2017; *Peavoy and Franzke*, 2010], show more erratic fluctuations, i.e., the corresponding probability density function (PDF) decays much slower than the corresponding Gaussian distribution with the same mean and variance. Hence, such distributions have heavier tails than the corresponding Gaussian distribution and very extreme events are much more likely than in the Gaussian world.

This behavior can be explained by the generalized Central Limit Theorem [*Sornette*, 2006], a generalization of the standard Central Limit Theorem [*Wilks*, 2011] which permits the random variables to have infinite variance, which means that the sums of such random variables scale as $N^{1/\alpha}$ and follow α-stable distributions with $0 < \alpha \leq 2$ [*Doukhan et al.*, 2002; *Sornette*, 2006; *Samorodnitsky*, 2016]. For $\alpha = 2$ we recover the Gaussian case with finite variance. The Central Limit Theorem expresses the fact that sums of random variables from short-tailed PDFs converge to a fixed point, i.e., a Gaussian distribution which retains its shape and is therefore a stable distribution [*Mantegna and Stanley*, 1999]. In the case of the generalized Central Limit Theorem, there is a series of such fixed points which can be imagined as forming a line in the space of all possible distributions, with each point on the line corresponding to an exponent α in the range from 2 to 0. Hence, sums of random variables from heavy-tailed, power-law Probability Density Functions converge to a power-law distribution, the α-stable distribution; rather than being Gaussian. In general, the α-stable Probability Density Functions do not have an analytic representation except via their characteristic functions, i.e., the Fourier transform of the PDF $p(x)$ [*Gardiner*, 2009]. The α-stable distributions with $\alpha < 2$ have characteristic functions of the form $p(s) \sim e^{-s^\alpha}$ and so $p(x)$ decays asymptotically as a power law: $p(s) \sim s^{-(1+\alpha)}$ as $s \to \infty$ [*Sornette*, 2006]. Furthermore, these power-law distributions decay so slowly that for $\alpha < 2$ the variance does not exist and for $\alpha < 1$ not even the mean exists. There is a corresponding random walk with α-stable increments, often called a 'Lévy flight', whose root mean square displacement grows as $\sim t^{1/\alpha}$, which is referred to as super-diffusion [*Gardiner*, 2009].

2.2.2 Other non-Gaussian Probability Density Functions

The α-stable model is simple and, thus, economical but can have extremely wild fluctuations. The properties of observational data may motivate other models for fluctuations which are less extreme than in the α-stable model. In particular, the infinite variance property of the α-stable model may yield fluctuations with tails that are heavier than desired and observed. Thus, other non-Gaussian PDFs need to be considered, such as stretched exponentials, where the PDF is given by $p(x) \sim e^{-x^s}$ with s between 0 and 1, or a log-normal distribution. Furthermore, heavy-tailed PDFs can also originate from extreme value statistics [*Coles*, 2001] that rely on the Fisher-Tippett-Gnedenko theorem [*Coles*, 2001] which is based on the maxima of identically and independently distributed sequences of random variables, rather than their sums as in the Central Limit Theorem.

Unlike α-stable distributions, these are not stable under addition which means that they converge towards the Gaussian distribution under addition. For instance, a first order autoregressive process $x_{t+1} = ax_t + \sigma^2\zeta$ where ζ is a Gaussian-distributed random variable, with variance σ^2, is also Gaussian distributed for x. However, if ζ were assumed to be log-normal then the process distribution x would not be log-normal but asymptotically Gaussian. This suggests that also non-linear and multiplicative processes need to be considered to explain the existence of power-law probability density functions. For instance, non-Gaussian distributions can also be created by multiplicative processes, such as multiplying a state variable with Gaussian noise [*Majda et al.*, 2008, 2009; *Sardeshmukh and Sura*, 2009; *Franzke*, 2017]. Such multiplicative noise can create heavy-tailed distributions. They naturally occur in stochastic climate theory [*Sura*, 2011; *Penland and Sardeshmukh*, 2012; *Franzke et al.*, 2015a; *Sardeshmukh and Penland*, 2015; *Franzke and O'Kane*, 2017; *Gottwald et al.*, 2017]. The energy cascade in turbulence is a particularly important multiplicative physical model as it describes the non-linear interaction between different scales or waves [*Vallis*, 2017].

2.3 Long-Range Dependence

Long-Range Dependence is characterized by a slow, power-law decay, of the autocorrelation function. This implies that even long ago states still affect the current state, thus, even far apart in time states, show dependence on each other.

The most basic Long-Range Dependence model is the fractional Brownian motion (fBm) (See appendix B:). The main difference between fractional Brownian motion and regular Brownian motion is that in Brownian motion the increments are independent of each other while in fractional Brownian motion such increments are dependent in time. This dependence actually covers the whole past; that is the reason why this model is sometimes also called Long-Term Persistence or Long-Memory (for $H > 0.5$). There are different definitions of fractional Brownian motion and we refer to the specialist literature for more details [e.g. *Lévy*, 1953; *Mandelbrot and Van Ness*, 1968; *Beran*, 1994; *Beran et al.*, 2013; *Embrechts and Maejima*, 2007].

While fractional Brownian motion is a continuous-time process, the statistics literature prefers a more flexible model, the discrete time Autoregressive Fractionally Integrated Moving Average model (ARFIMA) [e.g. *Hosking*, 1981; *Granger*, 1978; *Granger and Joyeux*, 1980; *Beran*, 1994]:

$$\Phi(B)(1-B)^d X_t = \Psi(B) Z_t \tag{7}$$

where B denotes the back shift operator $BX_t = X_{t-1}, B^2 X_t = X_{t-2}, \ldots$. The polynomials Φ and Ψ are defined as $\Phi(x) := 1 - \sum_{j=1}^{p} a_j x^j$ and $\Psi(x) := 1 + \sum_{j=1}^{q} b_j x^j$, where p and q are integers and denote the order of the autoregressive Φ and moving average Ψ parts, respectively. The noise variables Z_t are assumed to be independent Gaussian distributed with zero-mean and constant variance σ_Z^2. See appendix A: for more details.

However, the Autoregressive Fractionally Integrated Moving Average model model can also be generalized to use α-stable distributed increments [*Kokoszka and Taqqu*, 1994; *Stoev and Taqqu*, 2005; *Franzke et al.*, 2012; *Graves et al.*, 2017b]. For these infinite variance models no agreed upon definition of Long-Range Dependence exists [*Samorodnitsky*, 2016]. Note that for $d = 0$ the Autoregressive Fractionally Integrated Moving Average model reduces to the Autoregressive Moving Average model (ARMA) which is a Short-Range Dependence (SRD) process. In general, Autoregressive Fractionally Integrated Moving Average models can also be driven by non-Gaussian (e.g. t-distributed) noise [*Graves et al.*, 2017b]. Autoregressive Fractionally Integrated Moving Average models are more flexible than fractional Brownian motion since they combine a Long-Range Dependence component with Short-Range Dependence behavior [*Beran*, 1994; *Beran et al.*, 2013; *Franzke et al.*, 2012; *Graves et al.*, 2015]. The R package ARFIMA can be used to estimate Autoregressive Fractional Integrated Moving Average models [*Veenstra*, 2012].

These are the two most important and widely used paradigmatic models of Long-Range Dependence, but since they were not derived from basic physical laws their use in climate research was originally, and continous to be, met with criticism [e.g. *Klemeš*, 1974; *Maraun et al.*, 2004; *Mann*, 2011]. Long-Range Dependence also implies that even the most distant past still influences the current and future climate, which appears at odds with common intuition. Many geophysical equations of motion such as the Navier-Stokes or the primitive equations are usually Markovian, i.e., their current state only depends on the immediately preceding state and not on states in the more distant past. Furthermore, they do not have memory terms [*Mori*, 1965; *Zwanzig*, 1973, 2001; *Chorin and Hald*, 2013; *Gottwald et al.*, 2017]. This fact appears to be at odds with the observed (non-Markovian) Long-Range Dependence behavior of many climate time series and has led to much debate [*Percival et al.*, 2001; *Maraun et al.*, 2004; *Cohn and Lins*, 2005; *Vyushin and Kushner*, 2009; *Mann*, 2011; *Franzke*, 2012; *Bunde et al.*, 2014]. The debate stems from the fact that the underlying equations of motion are Markovian. However, Long-Range Dependence is frequently seen in time series from an aggregated system rather than data from a less ambiguous physical variable, and so the apparent paradox may be illusory since even Markovian systems can appear non-Markovian when not observing the full system. We will discuss possible physical mechanisms to explain this behavior in section 2.5.

2.4 Multi-fractals

In the introduction, we discussed scaling in precipitation intensities and in temperature time series. For intensity fields as well as time series, there are notions of multi-fractality that generalize self-similar scaling.

Intensity fields in geophysics can have spatial characteristics that are consistent with random cascades [*Kahane and Peyriere*, 1976; *Sornette*, 2006; *Kantelhardt*, 2009]. In such cascades, the intensity in a spatial region distributes non-uniformly between its smaller-scale subregions according to multiplicative processes. The simplest example is the binomial cascade introduced by *Kahane* [1985]. This model originates in turbulence theory, as a rigorous analysis of the Kolmogorov-Obukhov model for spatial variability of the energy dissipation rate [*Obukhov*, 1962; *Kolmogorov*, 1962]. The multiplicative chaos model [*Riedi et al.*, 1999] is a modern version of the same idea.

The binomial cascade and the multiplicative chaos models define singular (non-smooth) measures. By construction, the qth moments of the region-averaged intensities are power laws in spatial scale, with exponents that depend concavely on q. Consequently, the distributions of intensities between different spatial regions become increasingly leptokurtic with decreasing scale.

A multi-fractal time series $X(t)$ is one where the qth moment of an increment $|X(t + \Delta t) - X(t)|$ scales with the time-lag Δt, with an exponent $\zeta(q)$ that depends concavely on q. The scaling function $\zeta(q)$ is linear for self-similar processes.

There are several ways to construct multi-fractal stochastic processes from multi-fractal measures. In most constructions, a multi-fractal intensity field on the time axis determines the amplitudes in the time series, analogous to how the energy dissipation rate determines the amplitude of velocity field fluctuations in turbulence theory.

For strictly concave scaling functions the distributions of increments are more leptokurtic on short time scales than on longer time scales. Consequently, all multi-fractal time series are non-Gaussian. The reverse implication does not hold. It is well known, that unless one carefully verifies scaling of higher-order moments, standard techniques for estimation of multi-fractality can lead to spurious results for time series with non-Gaussian marginal distributions.

While multi-fractals are an abstract concept, they are useful for modeling time series with volatility clustering in time series, where the serial correlations between large and small amplitude events are different.

Applications of multi-fractal models in climate science have been shown by *Schmitt et al.* [1995]. More recently, *Ashkenazy et al.* [2003] analyzed climate data from the past 100kyr and found evidence for nonlinearity and clustering of the magnitude of climatic changes, consistent with multi-fractality. Similar results have been found by *Maslov* [2014]. Evidence of multi-fractal scaling in temperature, wind, and precipitation has been found by *Gan et al.* [2007]; *Royer et al.* [2008]; *Baranowski et al.* [2015]. See appendix E: for multi-fractal estimators.

2.5 Physical scaling mechanisms

Scaling, and particularly Long-Range Dependence, is an actively discussed topic in climate research. There is no obvious physical mechanism in the climate system that would allow the distant past to directly affect the current state of the system. Since the equations of motion used in climate models are all usually Markovian and do not contain memory terms, how can we explain the presence of Long-Range Dependence, and scaling, in the climate system?

2.5.1 Model Reduction

Long-Range Dependence can be explained using the Mori-Zwanzig formalism from statistical physics [*Mori*, 1965; *Zwanzig*, 1973, 2001; *Gottwald et al.*, 2017] which rigorously demonstrates how model reduction leads to the emergence of memory terms in the reduced equations of motion. Let us consider the following example [*Zwanzig*, 2001; *Gottwald et al.*, 2017]:

$$\dot{x} = L_{11}x + L_{12}y \tag{8}$$

$$\dot{y} = L_{21}x + L_{22}y \tag{9}$$

where L_{ij} are constant parameters. If we are now only interested in the dynamics of x we can formally solve for y

$$y(t) = L_{12}e^{L_{22}t}y(0) + L_{12}\int_0^t e^{L_{22}(t-s)}L_{21}x(s)ds \tag{10}$$

which we can now insert into Eq. (8)

$$\dot{x}(t) = L_{11}x + L_{12}\int_0^t e^{L_{22}(t-s)}L_{21}x(s)ds + L_{12}e^{L_{22}t}y(0) \tag{11}$$

The first term is a Markovian term from the original equations, the second term is a memory term since it integrates over the past and the last term is the initial condition which can be considered to be random. This example explicitly shows how one gets memory terms when looking only at parts of the full state vector. Eq. (11) is still exactly equivalent to the original system.

Most of our measurements are point measurements, or just measurements of a subset of the continuous fields. In either case, their dynamics stem from a low-dimensional system embedded in a climate system of infinite dimensions. The Mori-Zwanzig formalism

shows that memory effects arise if only a small part of the full system is observed. Thus, Long-Range Dependence could be a direct result of this observation. While the memory term in Eq. (11) is fairly general – which makes it impossible to know how exactly memory decays – a power-law decay is a possibility, especially when making additional assumptions about the memory kernel. *Kupferman* [2004] approximated the memory kernel with a power-law.

2.5.2 Non-Linearity and Regimes

Lorenz put forward the idea that deterministic systems can be almost intransitive, i.e., they can exhibit long-lasting climate changes and hence no unique climate state exists [*Lorenz*, 1968, 1976]. Such long-term anomalies can be a form of scaling in that the variance increases with increasing time scale. Several studies have shown that non-linearity can lead to scaling [*Lorenz*, 1976; *Franzke et al.*, 2015b; *Mesa et al.*, 2012]. Atmospheric circulation regime behavior, a main component of the climate system [*Nicolis*, 1990; *Ghil and Robertson*, 2002; *Hannachi et al.*, 2017; *Feldstein and Franzke*, 2017; *Franzke*, 2013; *Franzke et al.*, 2011], has been suggested as a prime candidate for scaling [*Franzke et al.*, 2015b]. An example of atmospheric circulation regimes is given by the quasi-stationary circulation systems like blocking events, which are quasi-stationary high-pressure systems that can last for weeks and cause heat waves and cold spells [*Hannachi et al.*, 2017; *Feldstein and Franzke*, 2017]. It has been shown for very long but finite time series that regime behavior is a plausible mechanism for scaling because the residence times of the regimes are power-law distributed [*Franzke et al.*, 2015b; *Diebold and Inoue*, 2001]. The residence time, is the time the system stays in one regime state. If these time intervals are power-law distributed then the system can exhibit Long-Range Dependence. This implies that memory effects in the climate system may not be needed to explain the apparent scaling of variance with time scale. The origin of this scaling has been found to be associated with the coarse-graining of the dynamics into a finite number of specific regimes, leading to non-Markovian dynamics [*Nicolis*, 1990; *Nicolis and Nicolis*, 1988, 1995; *Nicolis et al.*, 1997; *Vannitsem*, 2001].

Recent model experiments suggest also another possible non-linear mechanism that could explain Long-Range Dependence: the coupling of the atmosphere with other components of the climate system that have very different characteristic time scales. A case in point is ocean-atmosphere coupling, for which a reduced order non-linear coupled model

has been developed recently [*Vannitsem et al.*, 2015; *De Cruz et al.*, 2016]. This model employs the quasi-geostrophic equations to describe the large-scale dynamics of the atmosphere and oceans in extra-tropical regions. The coupling is achieved via an energy balance scheme and momentum transfer through wind stress.

Multiple scaling regimes were found (Fig. 7) using a Haar wavelet analysis (see appendix C.1.4). Remarkably, no Low-Frequency Variability (LFV) was found in the coupled model for small friction coefficients and the moments peak at a scale of about 10 days and decrease for larger periods. By low-frequency variability we mean a set of long-periodic, attracting orbits that couple the dynamical modes of the ocean and the atmosphere in this model. If low-frequency variability develops in the system, then additional peaks emerge at 10,000 and 40,000 days. Similar to *Lovejoy* [2015a], this allows us to define different regimes based on the respective scaling exponents. The structure of the Low-Frequency Variability and Long-Range Dependence critically depends on the water depth (Figs. 7b and c). This suggests that one plausible explanation of observed scaling regimes lies in the coupling of climate sub-components. We will further discuss this coupling mechanism in a linear framework next.

2.5.3 Superposition of Linear Short-Range Dependence Models and Linear Response

Another plausible scaling mechanism is the superposition of Short-Range Dependence models such as first order autoregressive process models [*Granger*, 1980]. This approach assumes that each climate subcomponent (atmosphere, ocean, land, cryosphere, etc.) evolves according to some Short-Range Dependence process. The superposition of those climate sub-component processes can result in scaling and Long-Range Dependence behavior [*Granger*, 1980]. The plausibility of this hypothesis has been confirmed by the linear response in energy balance models [*Fredriksen and Rypdal*, 2017]. Linear model types include the vertical diffusion model of *Fraedrich et al.* [2004] for the ocean temperature. With two layers the model produces a 1/f spectral range in the mixed layer temperature for a white noise surface forcing.

Another example is the Pacific Decadal Oscillation [*Mantua and Hare*, 2002] which also shows strong Long-Range Dependence [*Yuan et al.*, 2014]. The Pacific Decadal Oscillation shows variability on inter-annual to multi-decadal time scales. The Pacific Decadal

Oscillation is not thought of being a single physical model of variability; instead it is the aggregation of several different physical processes such as ENSO teleconnections, sea surface temperature reemergence and stochastic atmospheric forcing [*Newman et al.*, 2003; *Vimont*, 2005; *Schneider and Cornuelle*, 2005; *Qiu et al.*, 2007; *Newman et al.*, 2016]. Hence, the Pacific Decadal Oscillation is rather an imprint of scaling in the climate system than its cause.

On one hand this superposition mechanism is physically plausible, on the other hand from a statistical point of view it requires the estimation of many parameters. Hence, from a model selection point of view, which favors an economical model with as few parameters as necessary over more complex models (Occam's razor principle) [*Burnham and Anderson*, 2003], the scaling models are preferable. This does not mean that they are the best representation of the underlying dynamics. This suggests that in practice one has to decide whether we want to better understand the physical processes behind certain phenomena, or want an efficient and skillful statistical model, for example for prediction purposes.

2.5.4 Non-Gaussianity and Multiplicative Noise

As discussed above, scaling can also arise from the distribution of the increments or the driving noise in a stochastic process. So far we only discussed scaling in additive noise processes which in addition may have heavy-tails. Also Gaussian noise can produce power-law PDFs when it occurs in a multiplicative or state-dependent process [*Sornette*, 2006; *Majda et al.*, 2009; *Sardeshmukh and Sura*, 2009; *Sura and Hannachi*, 2015; *Penland and Sardeshmukh*, 2012; *Franzke*, 2017; *Bódai and Franzke*, 2017]. The simplest multiplicative noise process is the Kesten process [*Sornette*, 2006], a first order autoregressive process model with random coefficients:

$$x_{n+1} = a_n x_n + b_n, \tag{12}$$

where a_n and b_n are independent random variables. Under certain conditions, the Kesten process has a process cumulative probability density function (CDF) with a power-law decay of its tails, that is

$$P(X_t > x) \sim x^{-(1+\gamma)} \tag{13}$$

where γ is the power-law exponent.

Stochastic climate theory predicts the presence of multiplicative noise in non-linear systems [*Majda et al.*, 1999, 2001, 2008, 2009; *Sardeshmukh and Sura*, 2009; *Sura and Hannachi*, 2015; *Penland and Sardeshmukh*, 2012; *Franzke*, 2017; *Franzke et al.*, 2005; *Franzke and Majda*, 2006; *Franzke et al.*, 2015a, 2019; *Gottwald et al.*, 2017]. It can also be shown that multiplicative noise leads to power-laws over some ranges in stochastic climate models [*Majda et al.*, 2009; *Sardeshmukh and Sura*, 2009; *Sura and Hannachi*, 2015]. Unlike power-law processes, stochastic climate theory also provides mechanisms to limit extremes. This power-law roll-off is due to the same non-linear interaction that causes the multiplicative noise in the first place: the non-linear interaction between slow and fast components [*Majda et al.*, 1999, 2001, 2008, 2009; *Franzke et al.*, 2005; *Franzke and Majda*, 2006; *Sardeshmukh and Sura*, 2009; *Sura and Sardeshmukh*, 2008]. This can be understood as follows: The fast components of the flow, e.g. convection, synoptic-scale weather systems, are effectively serially uncorrelated on the time scale of the slow components, e.g. Rossby waves or the ocean. This time scale separation allows us to treat the fast components effectively as a noise variable. While there are non-linear interactions between the slow and the fast components in the climate system, this can be written now as a product of the slow flow variable and a noise variable; i.e. multiplicative noise, also called state-dependent noise since the impact of the noise can be modulated by the state of the slow variable. This is consistent with the findings of *Sardeshmukh and Sura* [2009] where they found evidence in global circulation model simulations that multiplicative noise is due to turbulent adiabatic fluxes and not rapid diabatic forcing fluctuations. An example are wind gusts: if the large-scale wind speed is low then there are only weak wind gusts; on the other hand, if the large-scale wind speed is high also the wind gusts are strong. This behavior can be easily represented by a multiplicative noise where the wind gusts are computed by the product of the large-scale wind speed and a noise. The relevance of multiplicative noise has been shown for sea surface temperature variability [*Sura and Sardeshmukh*, 2008], atmospheric vorticity variability [*Sardeshmukh and Sura*, 2009], teleconnection patterns such as the North Atlantic Oscillation [*Majda et al.*, 2009; *Önskog et al.*, 2019] and extreme events [*Sura*, 2013; *Penland and Sardeshmukh*, 2012; *Franzke*, 2017].

While theoretical considerations predict a power-law, e.g. the generalized Central Limit Theorem [*Sornette*, 2006], our climate system is of finite size and thus infinitely large events cannot occur which mean that the power-laws need to cut or roll off at some

intensity or spatial size. This is also consistent with the dynamical systems theory of extremes [*Lucarini et al.*, 2016] which shows that pure power-law dynamics cannot occur at arbitrarily large intensities or sizes.

2.5.5 Non-Stationarities

While *Hurst* [1951] was the first to discover scaling in natural time series, *Kolmogorov* [1940]; *Lamperti* [1962]; *Mandelbrot* [1965]; *Mandelbrot and Wallis* [1968] developed the first mathematical Long-Range Dependence models (see above) to explain such behavior [*Graves et al.*, 2017a]. From the outset, the Long-Range Dependence concept was controversial, especially in hydrology [*Klemeš*, 1974]. *Klemeš* argued that Long-Range Dependence can be caused by non-stationarities and by random walks with an absorbing boundary. The latter is mostly relevant for natural storage systems but less so for the climate system and will therefore not be discussed here. *Klemeš* argues that Long-Range Dependence is only an apparent effect and that there is no real memory in the climate system. While it is easy to construct non-stationary models exhibiting Long-Range Dependence [*Klemeš*, 1974] they raise deep philosophical questions about how the climate system is modeled. In general, all models of natural systems are assumed to have fixed parameters stemming from the underlying physical laws and all apparent non-stationarities would be the result of non-linearities in the underlying equations of motion or due to changes in external forcing (e.g. greenhouse gas emissions, Milankovich cycles). One could design non-stationary climate models by introducing random jumps in model parameters which would lead to shifts in the mean state, as proposed for hydrology by *Klemeš* [1974]. For instance, the inclusion of volcanic activity, which is very intermittent, improves the scaling behavior of climate simulations [*Vyushin et al.*, 2004]. While the success and skill of current numerical weather and climate predictions shows the usefulness of the stationarity assumption, the question remains unresolved whether non-stationary models could provide a viable alternative.

2.5.6 Self-Organized Criticality

Self-Organized Criticality (SOC) may be another possible mechanism behind scaling [*Bak et al.*, 1987; *Bak*, 1996; *Watkins et al.*, 2016]. SOC refers to a process driven by a slow and constant energy input that leads to sudden burst behavior without any typi-

cal scale. Hence, the statistics of a SOC process are described by power-laws [*Hergarten*, 2003].

Peters et al. [2001]; *Peters and Christensen* [2006]; *Peters and Neelin* [2006] used SOC to explain the observed scaling of precipitation. The atmosphere receives energy from evaporation due to solar radiation. The water vapor is stored in the atmosphere until a dynamical threshold (saturation) is reached, at which point energy bursts out, i.e., it rains and latent heat is released. These burst events have no typical scale and are a possible explanation of the observed power-law behavior of the tail of the PDF of precipitation event sizes and durations.

Another potential mechanism for power-laws is the Highly Optimized Tolerance (HOT) framework [*Carlson and Doyle*, 1999, 2000, 2002]. This framework relates power-laws to evolving structures. However, this framework has been developed for biological and engineering systems. How well it can also be applied to the climate systems needs to be examined. A recent application was to ecosystems and wild fires [*Moritz et al.*, 2005].

2.5.7 Scaling via Turbulent Cascades

While the above approaches apply to the time domain and aim to explain the presence of Long-Range Dependence in the climate system or intensity distributions, we now discuss a theory to explain the existence of scaling in the space domain. We focus on energy spectra, i.e., on how energy is distributed with spatial scale.

At the largest scales, the atmosphere is forced in a quasi-steady manner by the solar gradient between the equator and the poles, which leads to a meridional temperature gradient. The corresponding energy flux is represented by non-linear terms in the equations of motion used in coupled atmosphere-ocean models. The non-linear interactions between different spatial scales cause large eddies to break up into smaller 'daughter eddies', transferring their energy fluxes to ever smaller scales [*Vallis*, 2017] until viscosity dissipates the energy as heat.

This process can be modeled by cascade models. In the first cascade models, the parent eddies were typically large cubes that produced smaller daughter cubes of half the parent's diameter [*Schertzer and Lovejoy*, 1987]. Now, for each daughter, one flips a coin to decide how the energy flux from the parent eddy will be transferred over to the daugh-

ter. This can be done so that some daughter eddies occasionally receive zero energy while others have their fluxes multiplicatively boosted to conserve the total energy [*Novikov and Stewart*, 1964; *Mandelbrot*, 1974; *Frisch et al.*, 1978]. The outcome of these cascades are power-laws for the distribution of the energy with spatial scale [*Nastrom and Gage*, 1985; *Straus and Ditlevsen*, 1999; *Vallis*, 2017]. These are qualitatively consistent with the theoretical power-law spectra predicted by Kolmogorov [*Kolmogorov*, 1991a,b] as discussed above.

2.6 Estimation methods for scaling exponents

A multitude of estimators have been developed over the years to provide accurate estimates of the scaling exponents and different estimators infer different aspects of the scaling properties. For instance, most estimators infer the Long-Range Dependence parameter parameter d or the Hurst exponent H of a time series, and are insensitive to non-Gaussianity of its amplitudes, which can cause them to differ from the self-similarity parameter γ_{SS}. When deciding whether or not to use a particular estimator, one should always be aware of the underlying assumptions that went into its construction.

2.6.1 Estimation of the power-law exponent

Recognizing the existence of power-law tails and estimating the corresponding tail parameter or scaling exponent of power-law PDFs are important topics. *Clauset et al.* [2009] provide a review on this topic and carefully explain the potential pitfalls. Firstly, it is important to realize that true power-laws can be hard to identify and that simple regression approaches can lead to false positive identifications [*Clauset et al.*, 2009]. *Clauset et al.* [2009] recommend the use of a Maximum Likelihood Estimator (MLE). Code for the power-law estimation for the statistical programming language R is available at `http://tuvalu.santafe.edu/~aaronc/powerlaws/plfit.r`. They also show that the widely used least-squares regression approach can lead to inaccurate estimates and cannot answer the question whether the data obey a power-law decay at all [*Clauset et al.*, 2009]. *Gerlach and Altmann* [2019] propose a different way to identify power-laws using shuffling and under-sampling of the data. This approach leads to less rejections and larger confidence intervals then the *Clauset et al.* [2009] approach and potentially to more false positive identifications. While that study is mostly concerned with power-law tails of Probability Density Functions, the Maximum Likelihood Estimator approach can also be used for es-

timating the Long-Range Dependence parameter. With an Maximum Likelihood Estimator also the parameters of other distributions such as the Generalized Extreme Value, stretched exponential or the log-normal distribution can be estimated. Most of these distributions can be estimated with standard functions or packages included in the statistical software package R.

Extreme value statistics also provides methods to estimate the tail exponent of distributions [e.g. *Beirlant et al.*, 2006; *Coles*, 2001; *Embrechts et al.*, 2013]. However, they fit an extreme value distribution, either the Generalized Extreme Value or the Generalized Pareto distribution. Those distributions can have either a power-law or an exponential decay of their tail. *Gilleland and Katz* [2016] provide a R package for the estimation of extreme value distributions.

While direct confirmation of power-laws can be difficult, it is advisable to perform a model selection exercise where different models, e.g. power-law, stretched-exponential, etc, are fitted to the data and then the best fitted model is selected using some objective criterion. *Burnham and Anderson* [2003] discuss systematic ways of model selection.

2.6.2 Estimation of the Long-Range Dependence parameter

When analyzing and modeling the temporal dependence of Long-Range Dependent time series, it is important to accurately estimate the strength of Long-Range Dependence. This can be achieved by determining the Hurst exponent, H, or the fractional integration parameter, d, arising from the Autoregressive Fractionally Integrated Moving Average (ARFIMA) class of processes [*Box et al.*, 2015; *McNeil et al.*, 2015; *Franzke et al.*, 2012; *Franzke*, 2017]. H is more commonly used in physics, while d is preferred by the statistics community (Appendix D:). In principle, one could also analyze the autocorrelation function and search for a power-law decay at large time lags. However, the estimation of the autocorrelation function suffers from sampling errors. The below described estimators provide more robust estimates of the scaling exponent then the autocorrelation function.

As emphasized by *Samorodnitsky* [2016], most definitions of Long-Range Dependence in the literature are based on the second-order (or variance) properties of the process which include the asymptotic behavior of covariances, spectral density, and variances of partial sums. Second-order properties are popular choices because they are conceptually simple and they can be easily estimated from data. However, as noted by *Samorodnitsky*

[2016] this complicates the definition of Long-Range Dependence when a process has infinite variance, because theoretical moments and sample moments will behave in strikingly different ways.

Many methods are used to estimate the value of the Hurst exponent. They can be broadly divided into time-domain methods and frequency-domain methods. Time domain methods include variance-type estimators [*Taqqu et al.*, 1995; *Giraitis et al.*, 1999], the rescaled range or R/S statistic [*Mandelbrot and Taqqu*, 1979; *Bhattacharya et al.*, 1983], least-squares regression using sub-sampling [*Higuchi*, 1990], and the variance of residuals estimators [*Peng et al.*, 1994]. Frequency domain estimators include Whittle estimators [*Fox and Taqqu*, 1986; *Dahlhaus*, 1989] and connections to Fourier spectrum decay [*Geweke and Porter-Hudak*, 1983; *Lobato and Robinson*, 1996]. Wavelet-based regression approaches have also been considered [*Percival and Guttorp*, 1994; *Abry et al.*, 2000, 1995]. Extensions of wavelet estimators to other settings such as observational noise and irregularly spaced time series have been published [*Stoev et al.*, 2006; *Gloter et al.*, 2007; *Knight et al.*, 2016]. Other works about Long-Range Dependence estimation including multi-scale approaches are *Hsu* [2006] and *Coeurjolly et al.* [2014].

For technical details of the most common estimators see appendix C: . A more detailed comparison of different estimator and a discussion of their strengths and weaknesses is given by *Schmittbuhl et al.* [e.g. 1995]; *Faÿ et al.* [e.g. 2009]; *Franzke et al.* [e.g. 2012]; *Witt and Malamud* [e.g. 2013].

2.7 Spectral analysis and its limitations

Analyzing data by Fourier decomposition is attractive because each sinusoid has an exact and unambiguous time scale: its period. Furthermore, one has to assume that the signal is statistically stationary in time or homogeneous in space for Fourier analysis to be valid. This implies that the probability laws that govern the behavior of the time series do not change over time. As a result, the phases of the Fourier modes are randomly distributed and can be neglected. However, this only applies if the time series is Gaussian because the phases are correlated otherwise; a fact that is often overlooked. If observationally data really had to meet these criteria fully then Fourier analysis would not be as useful as it has been, so we can relax these conditions in practice to some extend.

A potential problem with Fourier analysis is that the interpretation of the resulting spectra are nontrivial. To illustrate this problem, we consider surrogate time series which have the same multi-fractal characteristics as the EPICA dust series [*Lambert et al.*, 2008; *Lovejoy*, 2018]. Since the time series are generated by a multi-fractal stochastic process, even the strong peaks are randomly excited (Fig. 8). However, the presence of such strong peaks can be misinterpreted as signatures of important climatic processes [*Lovejoy*, 2018]. This example, should also be taken as a cautionary tale against the simple interpretation of peaks in Fourier analysis as quasi-oscillations resulting from physical mechanisms. The important point we want to emphasize here is not that these methods do not work, but rather that their resulting spectra need to be carefully interpreted and the model representing the underlying process carefully chosen.

3 Empirical evidence of scaling

The atmospheric near surface temperature is a relevant indicator for climatic Long-Range Dependence. Instrumental measurements reach back to the 17th century (Central England temperature) [*Parker et al.*, 1992]) and cover inhabited areas and ocean areas near ship routes densely during the last 50 to 100 years. Furthermore, temperature is reconstructed using statistical relationships with proxy data for time horizons of thousands up to a million years. In all these data, a continuous background in variability shows up, which is parsimoniously described by power laws. In all these data sets, Long-Range Dependence is considered to be present if power-law scaling reaches the longest time scale observed (in contrast to a stringent mathematical definition which requires infinity as a limit).

All observations and especially climate reconstructions based on proxy data are subject to non-climatic influences such as measurement errors or imperfect recording of the climate signal. As this can also affect the scaling [e.g. *Rust et al.*, 2008; *Franzke et al.*, 2012], it is important to consider these uncertainties in order to infer useful information about the scaling of climate variability.

3.1 Station temperatures

Significant evidence for Long-Range Dependence in station temperature have been reported in many studies [*Koscielny-Bunde et al.*, 1998; *Fraedrich and Blender*, 2003;

Franzke, 2010, 2012; *Capparelli et al.*, 2013; *Bunde et al.*, 2014; *Ludescher et al.*, 2015; *Gao and Franzke*, 2017; *Graves et al.*, 2015; *Løvsletten and Rypdal*, 2016]. The exponent $H = 0.65$ of the fluctuation function $F(\tau) \sim \tau^H$ which was determined by Detrended Fluctuation Analysis was suggested to be universal for local temperature variations. This exponent is related to the exponent $\beta = 0.3$ in a power spectrum, $\mathcal{P}(f) \sim f^{-\beta}$ by $\beta := 2H - 1$. Fig. 9 shows the results for the fluctuation function obtained by Detrended Fluctuation Analysis (see appendix C.1.3) for 12 meteorological stations distributed worldwide. In the inter-annual range (up to the total duration) the stations reveal linear slopes given by an exponent $H \approx 0.65$. It is tempting to conclude that atmospheric Long-Range Dependence could be characterized by a single universal value. Later on, it was found that the value $H = 0.65$ might be a transition phenomenon between memory-less inner continents with $H = 0.5$ and some oceanic areas with $H \approx 1$ [*Fraedrich and Blender*, 2003; *Ludescher et al.*, 2015; *Løvsletten and Rypdal*, 2016].

3.2 Near surface temperatures on land and ocean

Observed near surface temperatures are available since 1900 with a sufficient density to estimate a global pattern of Long-Range Dependence [*Jones*, 1994; *Parker et al.*, 1995; *Jones et al.*, 2001]. Fig. 10 shows the power-law exponents H for the monthly data estimated by Detrended Fluctuation Analysis with quadratic trend (DFA2). The results are concentrated in coastal regions, Europe, North America, and along ship routes. In inner-continental locations, Long-Range Dependence is negligible with $H \approx 0.5$. Along the coasts the memory increases to $H \approx [0.6, 0.8]$ and in the central North Atlantic and in the equatorial Indian ocean, the highest values of $H \approx 0.9$ are found. This pattern shows that Long-Range Dependence in the considered time range is a marine phenomenon. On land no memory is evident on these time scales far from the coasts. The finding of $H \approx 0.65$ in the stations can be explained by the locations of the stations considered in Fig. 9. Inhabited areas with observational stations are traditionally along coasts. Note that there can be a huge gradient in Long-Range Dependence as seen along the western North Pacific coast. Later on, *Fraedrich and Blender* [2003] found that this surface temperature Long-Range Dependence can be found in coupled atmosphere-ocean simulations with a dynamical ocean model, but not with a slab ocean model.

A spectrum with $\beta = 1$ is denoted as $1/f$-noise (see appendix F:). If this extends to vanishing frequency (or infinite time) stationarity is violated. Clearly, in observational data

this limit cannot be attained, and whether the sea surface temperature data are stationary remains open. However, this type of variability indicates that short term averages, as are typically used in climate science, are not defined and should be interpreted with caution.

The $1/f$-spectrum of the sea surface temperature in some oceanic regions can be obtained by a diffusion model in a vertical column with two compartments [*Fraedrich et al.*, 2004]. A decisive parameter is the diffusivity in the abyssal ocean. This parameter determines turbulent mixing and is caused by tides and the orography. Although the atmospheric forcing is white, the sea surface temperature shows Long-Range Dependence close to non-stationarity. Furthermore, wind also provides a significant part of the mechanical energy for diapycnal mixing in the ocean [*Wunsch and Ferrari*, 2004; *Munk and Wunsch*, 1998] which could also lead to this behavior as discussed in Sec. 2.5.2.

In summary, the analysis of observational data reveals that Long-Range Dependence in the lower atmosphere is predominantly found in oceanic regions where the variability is close to non-stationarity ($1/f$-spectrum). Far from coasts Long-Range Dependence on decadal time scales is not observed. In transition zones along coasts a spectrum is found which is approximately given by $\mathcal{P}(f) \sim f^{-0.3}$.

3.3 Scaling in regional and global mean temperatures

Local and regional surface temperature variations have a much greater magnitude and show a weaker scaling than global average temperature variability [e.g. *Laepple and Huybers*, 2014a]. This difference in variability is strongest for inter-annual and shorter time-scales and decreases on longer time-scales. The reduced variability of the global mean temperatures reflect cancellation of variability in the global mean and the weaker cancelation toward lower frequencies is consistent with findings that temperature anomalies have greater spatial autocorrelation toward longer timescales [*Jones et al.*, 1997]. This behavior is also reproduced in diffusive Energy Balance models [*Rypdal et al.*, 2015] where the predicted slope of the spectrum of global mean temperatures is around double the slope of regional temperatures over a large frequency range following from the horizontal diffusive coupling. Thus, it is important to consider the spatial scale analyzed when interpreting the variability and long-term memory of temperature time-series.

3.4 Paleo data and simulations

The isotope fraction δ^{18}O in ice cores can be used as a proxy for the surface temperature due to the different weights of the molecules [*Dansgaard*, 1964; *Barlow et al.*, 1997]. As the annual snowfall is preserved on the Greenland and Antarctic ice-sheets, this allows in principle to analyze climate scaling from inter-annual to multi-millennial time-scales. However, especially on the shorter time-scales, non-temperature effects such as snow redistribution [*Fisher et al.*, 1985], diffusion [*Johnsen et al.*, 2000] and aliasing effects from intermittent snowfall [*Laepple et al.*, 2017] considerably affect the recorded variability and have to be corrected for, to infer about climate scaling [*Münch and Laepple*, 2018]. The oxygen fractions measured in the Greenland ice cores GRIP and GISP2 during the last 10k years reveal estimates for H [*Blender et al.*, 2006] which are clearly above 0.5 in the millennial time range (Fig. 11). The corresponding exponents in the power spectrum $\mathcal{P}(f)$ correspond to $H \approx 0.5$. Hence, scaling can be assumed, at least approximately. It is remarkable that this result can be obtained in an extremely long coupled atmosphere-ocean simulation [*Blender et al.*, 2006] which reveals intense Long-Range Dependence south of Greenland (see Fig. 11) with exponents of similar magnitude, but much less Long-Range Dependence in other oceanic regions; the Pacific ocean reveals no Long-Range Dependence of a comparable intensity. The simulation was performed under present-day conditions, hence no external variability is necessary to explain this result. The Long-Range Dependence is related to the variability of the zonally averaged streamfunction in the North Atlantic ocean. Evidence for Long-Range Dependence has also been found in other coupled climate models [*Østvand et al.*, 2014; *Fredriksen and Rypdal*, 2016; *Vyushin et al.*, 2004; *Zhu et al.*, 2010]. *Vyushin et al.* [2004] showed that the inclusion of volcanic eruptions improves the simulation of Long-Range Dependence in climate models.

An important question is how well climate models reproduce the true internal variability on time scales of centuries and longer. The local model surface temperature spectra seem to indicate a lower (or even zero) spectral exponent β for frequencies below 1×10^{-2} yr^{-1}, but on long time scales the finite sample size errors are so large that this cannot be concluded with high statistical confidence [*Fredriksen and Rypdal*, 2016]. This flattening of the spectra on time scales longer than a century cannot be detected in the instrumental temperatures, since the time series are too short, but they are also not detected in temperature reconstructions of Holocene climate [*Laepple and Huybers*, 2014b] even after correcting for non-climate effects on the spectra. On the contrary, some authors claim higher

exponents for local temperatures ($\beta \approx 2.2$) for these longer time scales based on composite spectra established from different proxy records [*Huybers and Curry*, 2006; *Lovejoy and Schertzer*, 2012a]. This conclusion, however, can only be drawn with confidence from proxy records spanning the last glacial period and including the last deglacation, and hence may not be valid for the Holocene climate.

Analyzing global mean temperatures that are dominated by external forcing on the other hand suggests that the scaling and variability is comparable between reconstructions and model simulations [*Crowley*, 2000; *Zhu et al.*, 2019] or climate model simulations may even overestimate global mean temperature variability.

The issue is illustrated in Fig. 12a where we have plotted time series of reconstructed annual temperatures for the Northern hemisphere and the corresponding series derived from the NorESM model with historical forcing. In Fig. 12b we have estimated their Haar fluctuation functions (see appendix C.1.4). The fluctuation level is more than twice as high for the model temperatures on time scales less than 100 yr. It turns out that this is due to the higher short-time responses to large volcanic eruptions in the models. This can be seen by elimination of these spikes from the model signal, as shown in the zoomed in signals in Fig. 12. The fluctuation function of this 'chopped' signal is very close to that of the reconstructed temperature (Fig. 12). The cloud of thin curves in Fig. 12 are fluctuation functions estimated for 20 realizations of fractional Gaussian noises with Hurst exponent $H = 0.9$ of the same length as the NorESM model run. The width of the cloud suggests that the reconstruction as well as the chopped model signal are consistent with such a fractional Gaussian noise, although the power at time scales longer than a few centuries are somewhat high. It is easy to verify that this increased power is due to the temperature difference between the Medieval Warming Anomaly and the Little Ice Age. Some authors interpret the power in this oscillation as a signature of a transition to scaling with an exponent $\beta > 1$ on time scales longer than a few centuries [*Huybers and Curry*, 2006; *Lovejoy and Schertzer*, 2012a]. On the other hand, *Nilsen et al.* [2016] argue that existing temperature reconstructions for the Holocene are generally consistent with a single scaling regime with $H \approx 0.9$ on all time scales shorter than the duration of the interglacial period.

Similar ideas are advocated by *Rypdal and Rypdal* [2016b], who demonstrate that temperatures derived from ice cores over the last 100 kyr can be described as sudden transitions between stadials and interstadials superposed on a $1/f$-noise ($H \approx 1$) background.

According to this view, mono-fractal, near Gaussian, scaling with $H \approx 1$ is a useful description of the climate noise background in Quaternary climate. Whether a scaling description is appropriate for the succession of transitions between stadials and interstadials is another story and needs more research.

3.5 Scaling behaviors in other meteorological and climatological variables

Besides the evidence of Long-Range Dependence in temperatures, there are also scaling behaviors detected in many other variables, including precipitation, river runoff, total ozone, relative humidity and sea level change. For in-situ precipitation records, small Long-Range Dependence parameters have been found by several studies [*Kantelhardt et al.*, 2006; *Jiang et al.*, 2017; *Yang and Fu*, 2019]. On average, the Detrended Fluctuation Analysis exponent H mainly ranges between 0.5 and 0.55, indicating weak Long-Range Dependence. For river runoff much stronger Long-Range Dependence has been detected. *Kantelhardt et al.* [2006] found the mean Long-Range Dependence parameter for river runoff is 0.72 based on 42 river runoff records observed from Europe, North and South America, Africa, Australia and Asia. *Wang et al.* [2008] detected Long-Range Dependence close to $1/f$ ($H \approx 1$) for the intra-annual Yangtze discharge. As for relative humidity, *Chen et al.* [2007] reported that the mean Detrended Fluctuation Analysis exponent H for in-situ relative humidity records over China is around 0.75. Recently, there were also results reported indicating that sea level changes are characterized by Long-Range Dependence. The Detrended Fluctuation Analysis exponent H has a large variation range from 0.60 to 0.95, depending on different regions [*Dangendorf et al.*, 2014]. Other variables such as wind speed, atmospheric circulation indices, and ozone anomalies, etc., have also been shown to have the scaling behavior [*Feng et al.*, 2009; *Vyushin et al.*, 2007; *Vyushin and Kushner*, 2009; *Varotsos and Kirk-Davidoff*, 2006; *Franzke et al.*, 2015b].

3.6 Evidence of multi-fractal behavior

Besides Long-Range Dependence that only needs one exponent H to describe mono-fractal behavior, there is also empirical evidence of multi-fractal behavior. For instance, in precipitation records, although the measured Long-Range Dependence is weak, pronounced multi-fractality has been found [*Kantelhardt et al.*, 2006], indicating that precipitation records of different amplitudes have different scaling behaviors. Similar properties also exist in river runoff data [*Koscielny-Bunde et al.*, 2006], where the multi-fractality

was found to be even stronger than that in precipitation records [*Kantelhardt et al.*, 2006]. For temperature related records such as the surface mean air temperature, diurnal temperature range (DTR), etc., different multi-fractal behaviors were found over different regions [*Lin and Fu*, 2008; *Yuan et al.*, 2013a]. Such as in the south of the Yangtze River, pronounced multi-fractality was found in DTR records, while in the north, the multi-fractal behavior is very weak or even non-existent [*Yuan et al.*, 2013a]. Other variables such as wind speed, relative humidity, etc., have also been shown to have the multi-fractal behavior [*Kavasseri and Nagarajan*, 2005; *Baranowski et al.*, 2015].

4 Applications of scaling in climate research

4.1 Scaling for Trend Detection

The identification of trends is one of the most frequent and prominent goals in the analysis of geophysical time series [*Chandler and Scott*, 2011; *Wu et al.*, 2007]. Although apparently an easily understandable objective, trend assessment is very challenging, starting with the lack of a precise definition of trend itself. Implicit in the intuitive notion of trend are concepts such as long-term, smoothness, or monotonicity, but it is not unambiguously defined how long is 'long-term', or how smooth needs a pattern to be in order to be a trend. Furthermore, time series characterized by scaling behavior often exhibit features that can be classified broadly as a trend, even in the absence of any genuine trend.

Unlike the notion of trend, stationarity is a well-defined statistical property. A time series (X_t) is weakly stationary if its first and second moments are time invariant (i.e., the mean and variance are constant and the covariance depends only on the time lag between the observations). The trend in a time series can be ascribed to a non-stationary generating process (at least the mean is not constant in time) and described by a trend model. Trend models can be broadly classified as either being i) deterministic or ii) stochastic. Deterministic trend models represent deterministic (non-random) non-stationary processes which are described by a function evolving in time; one example are trends which are forced by external factors such as anthropogenic greenhouse gas emissions. Stochastic trend models represent stochastic non-stationary processes described by models such as a random walk or an Autoregressive Integrated Moving Average model. Such models exhibit apparent trends without any external forcing, instead these trends are caused by the internal dynamics of the process. Unfortunately it is extremely difficult, for example by

visual judgment alone [*Percival and Rothrock*, 2005], to select a trend model or even to distinguish between deterministic and stochastic trend models, particularly in the case of short time series [*Franzke*, 2012].

Furthermore, weakly stationary processes can generate a time series with an apparent trend, particularly when a short segment of the process is observed. Such "spurious" trends can be misleadingly taken as evidence of non-stationary behavior when in fact the process is stationary. A classical example since long recognized as a potential culprit in the interpretation of climate variability [*Wunsch*, 1999] is the typical "red noise"[1] structure of climate records. A red noise (described by a simple first order autoregressive process) can produce visually appealing trends despite being a stationary process, particularly in the case of short time series.

Long-Range Dependence processes, for example described by Autoregressive Fractional Integrated Moving Average models, are another type of stationary processes that can produce apparent non-stationary behavior and local 'spurious' trends. Since Long-Range Dependence is a feature common in many geophysical time series, a crucial challenge in the identification and estimation of trends is to discriminate between non-stationary processes and stationary long-range dependent processes. The problem is, however, quite difficult since genuine trends generated by a non-stationary process and spurious trends from a Long-Range Dependence process can coexist in the same time series. Disentangling the different contributions to the observed temporal structure is not possible by visual inspection and even specific methodologies addressing the issue have to rely on substantial assumptions and simplifications, for example on the type of non-stationary behavior, or the dominance of one specific type of process. For instance, the approach proposed by *Beran and Feng* [2002] of semi-parametric fractional autoregressive (SEMIFAR) models considers a trend function modeled non-parametrically, with the remaining components of the model estimated by maximum likelihood. Despite the flexibility of SEMIFAR models, the performance is poor in the case of short time series, and the trend is estimated based on a subjective concept of smoothness. More importantly, discrimination between stochastic and deterministic trends remains difficult to achieve, given that a significant amount of

[1] Red noise sometimes means that the spectral power increases with period scale. However, in climate science red noise typically denotes the power spectrum of a first order autoregressive process which has first increasing power for increasing period but then becomes white noise; also called Lorentzian spectrum. See Fig. 3 for an example.

spurious trends associated with Long-Range Dependence behavior alone can be easily included in the non-parametric trend estimation. The statistical test of *Berkes et al.* [2006] aims to discriminate between stationary Long-Range Dependence time series and non-stationary time series with change-points in the mean, but requires previous identification of a small number of change points in the time series.

Shifting from a general trend model to a linear trend significantly constrains the problem of trend identification in the presence of Long-Range Dependence. Although such an assumption is hardly realistic and is theoretically limiting, it is nevertheless of practical relevance since the overwhelming majority of trends in geophysical records are reported as the slope from a linear regression model. Most studies are based on the up-front assumption that a time series can be described by a non-stationary linear trend with a stochastic Long-Range Dependence component (e.g. *Rybski and Bunde* [2009]; *Lennartz and Bunde* [2009]; *Franzke* [2010, 2012]; *Capparelli et al.* [2013]; *Bunde et al.* [2014]; *Ludescher et al.* [2015]; *Myrvoll-Nilsen et al.* [2019]) and focus on the assessment of the corresponding uncertainty (e.g. *Cohn and Lins* [2005], *Koutsoyiannis* [2006], and *Koutsoyiannis and Montanari* [2007]).

A complementary approach is to test the assumption of a linear deterministic trend itself. The PP test [*Phillips and Perron*, 1988] is a classical unit root test for testing non-stationarity in the form of a random walk, which is a scaling process. The KPSS test [*Kwiatkowski et al.*, 1992] is a parametric statistical test which assumes as a null hypothesis a deterministic linear trend plus a stationary stochastic noise. Although the two tests can be applied independently, their joint use is recommended for trend assessment purposes [*Kwiatkowski et al.*, 1992; *Carrion-i-Silvestre et al.*, 2001]. For a time series with a random walk stochastic trend the PP test should not reject the null hypothesis and the KPSS test should reject the linear trend null hypothesis. Conversely, for a time series with a linear deterministic trend, the PP test should reject the null hypothesis but not the KPSS test. In the case of a times series for which both tests fail to reject the null hypothesis, then the time series or the tests are not sufficiently informative to distinguish between a stochastic (random walk) trend and a deterministic trend. However, if both tests reject their respective null hypothesis, this is an indication that alternative parametrisations for long-term behavior need to be considered, such as Long-Range Dependence. Since Long-Range Dependence is a common feature of geophysical time series, this outcome of rejection of both PP and KPSS test is quite common, for example in the case of air temper-

ature [*Fatichi et al.*, 2009] or global sea surface temperature [*Barbosa*, 2011]. Unfortunately these tests require long time series (in order to meet asymptotic assumptions) and are known to have low explanatory power particularly against Long-Range Dependence alternatives (*Lee and Schmidt* [1996], *Leybourne and Newbold* [1999]).

A widely used trend test is the Mann-Kendall test [*Mann*, 1945; *Kendall*, 1948], which in its original form is only valid for independent data. The Mann-Kendall test is a non-parametric test which tests for the presence of a monotonic trend without making any assumptions about the form of the trend. This is in contrast to most other trend tests which have to assume some parametric trend form, e.g. a linear trend. The Mann-Kendall test has been extended to also account for serial correlation in time series [*Hamed and Rao*, 1998] and also for the presence of Long-Range Dependence [*Hamed*, 2008].

A further trend significance test has been developed by *Lennartz and Bunde* [2009, 2011]; *Tamazian et al.* [2015]. This method has been developed for the Detrended Fluctuation Analysis method in the presence of Long-Range Dependence in the time series [*Ludescher et al.*, 2015; *Bunde et al.*, 2014]. Based on Monte-Carlo simulations, they studied how the trend uncertainties vary with the strength of Long-Range Dependence, as well as the data length. This method has been applied to evaluate trend significances of the surface air temperature and the sea ice extent in Antarctica [*Ludescher et al.*, 2015; *Bunde et al.*, 2014; *Yuan et al.*, 2017].

Recent developments in temperature trend significance testing with long-range dependent noise show how we can also incorporate information about forced global temperature changes in the trend estimate [*Myrvoll-Nilsen et al.*, 2019]. In that way, one avoids attributing forced changes deviating from e.g. a linear trend as part of the stochastic variability [*Gil-Alana*, 2005; *Fatichi et al.*, 2009; *Franzke*, 2012, 2014]. Results show that the observed trends since 1900 are significant relative to the noise for most locations, and to a larger degree than when assuming a linear trend [*Løvsletten and Rypdal*, 2016].

4.2 Scaling for Climate Response and Sensitivity

Linear response models, which predict how the climate system will react to a change in forcing, e.g. anthropogenic greenhouse gas emissions, have shown considerable success in describing the global temperature response in climate model data, instrumental data and in multi-proxy reconstructions [*Held et al.*, 2010; *Geoffroy et al.*, 2013; *Caldeira and*

Myhrvold, 2013; *Rypdal and Rypdal*, 2014; *Østvand et al.*, 2014; *Rypdal et al.*, 2015; *Lovejoy et al.*, 2015; *Fredriksen and Rypdal*, 2016]. In particular, *Rypdal and Rypdal* [2014] demonstrated that a scaling linear response function provides a good description of the global temperature response to radiative forcing over both the historical period, and to a multi-proxy reconstruction of the temperature over the last millennium.

It has been known for several decades that Coupled Atmosphere-Ocean General Circulation Models (AOGCM) exhibit climate responses on multiple time scales [*Held et al.*, 2010; *Geoffroy et al.*, 2013; *Caldeira and Myhrvold*, 2013; *Fredriksen and Rypdal*, 2017], i.e., there is more than one time constant involved in the response. The scaling response studied in *Rypdal and Rypdal* [2014] could be considered an approximation to the multiple time-scale response, bridging the responses on time scales from years to centuries, and *Fredriksen and Rypdal* [2017] demonstrates how an energy balance box-model can provide such an approximation.

In addition to describing the response to historical radiative forcing, the same scaling response to a white noise stochastic forcing is also consistent with the observed internal variability. One way of extracting the internal variability from observed global temperature is to compute the deterministic, historically forced variability from the response model and subtract this from the observed record. The power spectrum of this estimated internal variability compares well with a power-law [*Rypdal and Rypdal*, 2014]. This tendency for a multi-box model to form a power-law spectrum is studied systematically in *Fredriksen and Rypdal* [2017], and reflects a well-known result which states that a scaling spectrum can be obtained from the aggregation over an ensemble of first order autoregressive processes [*Granger*, 1980]; see also Sec. 2.5.3. Thus, Long-Range Dependence can be caused by the constructive superposition of Short-Range Dependence processes.

The emergent scale invariance makes it possible to infer equilibrium climate sensitivity (ECS) from a scaling frequency-dependent climate sensitivity $R(f) \sim f^{\beta/2}$. This scaling response implies infinite magnitude response as $f \rightarrow 0$, and, therefore, there must exist a lower frequency limit of where the scaling response is valid, and where the response stabilizes as we go to even lower frequencies. $R(f)$ can be estimated for a given climate model by exploiting the relation between the historic radiative forcing applied to a model and the observed instrumental global temperature. *Rypdal et al.* [2018] applied this to an ensemble of Earth system models, where the inferred values of R(f) evaluated

at $f = 1/1000yr^{-1}$ correlate strongly to estimates of equilibrium climate sensitivity from idealized model runs. They could use the distribution of estimated $R(f)$ over the model ensemble to constrain the distribution of equilibrium climate sensitivity obtained from the ensemble of idealized runs. Thus, scale-invariant linear response models are useful tools for the estimation of equilibrium climate sensitivity from observation data. The advantage over multi-box energy balance models is that the scale-invariant models contain fewer free parameters and are less prone to statistical over-fitting.

4.3 Scaling for Climate Prediction

The property of Long-Range Dependence in the climate system raises the question whether models which explicitly include Long-Range Dependence can be used for skillful predictions. The first attempt for climate predictions was tried by *Baillie and Chung* [2002]. Recently, a model was developed for seasonal to decadal predictions, the Stochastic Seasonal to Inter-Annual Prediction System (StocSIPS) [*Lovejoy*, 2015b; *Lovejoy et al.*, 2015, 2018]. StocSIPS is based on the low frequency limit of a fractional differential equation, the Fractional Energy Balance model (FEBE). This model is valid for periods between 20 days through 50 years, where intermittency is relatively weak so that a quasi-Gaussian approximation can be used. StocSIPS forecast skill compares favorable with operational long-range forecasting models based on traditional climate models. One advantage of StocSIPS is that data assimilation of observations is not necessary, since it can directly be fitted to observed data. This also implies that down-scaling of forecasts is not needed.

Yuan et al. [2013b, 2014] developed a method for the extraction of the Long-Range Dependence using a fractional integrated statistical model. They proposed a new variable memory kernel which clearly shows how the states from the distant past maintain their impacts over time till the current time. Accordingly, climate variables with Long-Range Dependence can be decomposed into two parts: i) the memory part, which represents the influences accumulated from the past, and ii) the residual part, which is related to the current dynamical forcing conditions. With the memory part extracted, one can at least determine on what basis the considered time series will continue to change. By combining this with the estimated residual part, it is possible to make predictions. Therefore, they proposed a new perspective for climate prediction for climate variables with Long-Range

Dependence. Because the influence from the past can be extracted quantitatively, one only need to focus on the prediction of the residual part.

Also statistical models with non-Gaussian features have recently been developed. For instance, *Önskog et al.* [2018] show that the forecast skill of the North Atlantic Oscillation [*Feldstein and Franzke*, 2017] increases in a Short-Range Dependence statistical model when non-Gaussian noise is used. *Graves et al.* [2017b] developed a Bayesian framework for Autoregressive Fractional Integrated Moving Average models with various non-Gaussian noises and demonstrated its usefulness using the t- and the α-stable distribution.

5 Outlook and Open Questions

Here we have provided an overview of scaling methods and their relevance for understanding the climate system and its variability on time scales of days to millenia and ice ages. Scaling methods have improved our understanding of the climate system. The climate community mainly distinguishes between weather and climate, even though it is not well defined were weather ends and climate starts. Weather systems evolve over a few days, with the weather prediction limit at about 10-14 days [*Zhang et al.*, 2019], while climate starts at time scales of about 30-40 years. This leaves a large gap in between. The area between weather and climate, the weather-climate interface, consists of the active research areas of sub-seasonal-to-seasonal (S2S) up to decadal predictions.

Through scaling analysis we have now a better understanding of the climate system and that it consists of different scaling regimes distinguished by their scaling exponents. While in the weather and climate regimes the variability strongly increases with time scale, this is not the case for the regime in between where the increase is rather weak. The exact ranges of these scaling regimes and their robustness and meaning is still a matter of debate [e.g. *Nilsen et al.*, 2016; *Huybers and Curry*, 2006]. On longer time scales, such as decadal time scales, the effect of global warming might already affect the variability making the observed scaling likely not a product of internal climate processes but a response to external forcing and non-stationarity. More research is needed to clarify this point.

An important future research question is to understand these differences and to elucidate how predictable sub-seasonal-to-seasonal and decadal processes are. Forecasts on

these time scales are of societal importance and currently an important research topic. In this context it is also an important research question how the slope of the scaling relationship determines predictability? While the predictive skill in weather forecasts has significantly improved over the last few decades the skill of seasonal forecasts is rather limited and for decadal forecasts only the climate change signal and perhaps the El Nino-Southern Oscillation phenomenom are currently the predictable components.

Scaling of paleo-climate data has received a lot of attention [e.g. *Schmitt et al.*, 1995; *Huybers and Curry*, 2006; *Laepple and Huybers*, 2013; *Nilsen et al.*, 2016; *Fredriksen and Rypdal*, 2016; *Lovejoy and Schertzer*, 2013; *Lovejoy and Varotsos*, 2016; *Rypdal and Rypdal*, 2016a; *Bunde et al.*, 2014; *Zhu et al.*, 2019] and has been used in evaluating how well climate models reproduce observed long-term climate variability [*Østvand et al.*, 2014; *Fraedrich and Blender*, 2003; *Blender et al.*, 2006]. While global mean temperature variability on inter-annual to millennial time-scales seem to be consistent between climate models and climate reconstructions, the strong discrepancy of slow climate variability at regional scales calls for continued research on the temporal and spatial structure of climate variability but also on improving the interpretation and quality of paleo-climate records. How well scaling can contribute to the reconstruction of past climate needs to be assessed.

Over longer timescales there are several unanswered questions in paleoclimate where scaling approaches may help our understanding of the climate system. One of the great puzzles of Quaternary science is the transition from the 41kyr world before 1 million years ago to the current 100 kyr glacial-interglacial regime, without any external forcing changes. Potential explanations for this change have involved ice sheet dynamics [*Clark and Pollard*, 1998], the progressive cooling of Earth's temperature throughout the Quaternary [*Snyder*, 2016], the amount of dust in the atmosphere [*Chalk et al.*, 2017], or continental distribution [*Kender et al.*, 2018]. However, all studies acknowledge that the transition period between the 41kyr cycle ~1.2 Myr ago and the 100 kyr cycle since ~600 kyr ago is poorly defined and not well characterized. The recovery of the 800 kyr long EPICA ice core allowed a first look into the younger part of that transition section [*Jouzel et al.*, 2007]. The soon to be started oldest ice project aims to recover an Antarctic ice core that will reach back to the 41 kyr world, 1.2 million years ago and provide a high-resolution record throughout the Mid-Pleistocene Transition (MPT), which denotes the fundamental change in the behavior of glacial cycles around 1 million years ago. Before the MPT the glacial cycles were dominated by a 41,000 year period, after the MPT they followed

less regular cycles with an approximate period of 100,000 years. The statistical techniques described in this review could provide a robust description of the data variability upon which physical and dynamic models can be chosen to explain the observed changes.

Another interesting paleo-climatic question has been why human civilization has evolved during the Holocene and not during any of the previous interglacials [*Robinson et al.*, 2006]. Has the Holocene climate been exceptionally stable in time or in space [*Kopp et al.*, 2017]? While Greenlandic ice-core records provide evidence for a very stable Holocene [*Ditlevsen et al.*, 1996] the dependency of climate variability on the climate state seems to be much smaller in the rest of the world [*Rehfeld et al.*, 2018]. A related question is whether conditions in the fertile crescent during previous interglacials were markedly different from the Holocene. Scaling analyses could help answer these questions by providing a description of both temporal and spatial variability at different times during the Pleistocene.

A still open question is the mechanism of Long-Range Dependence in the climate system. While the evidence is strong for Long-Range Dependence, it leads to counterintuitive implications, i.e. that the distant past still influences the present. There are also studies who show that inhomogeneities on station time series increase the strength of Long-Range Dependence [*Mills*, 2007; *Rust et al.*, 2008]. These inhomogeneities take on the form of jumps or shifts due to changes in the station instruments or location. The fact that jumps lead to increased Long-Range Dependence strength would be consistent with the fact that volcanic eruptions improve the reproduction of Long-Range Dependence in climate models [*Vyushin et al.*, 2004]. *Maraun et al.* [2004] point out the difficulty in distinguishing between Long-Range Dependence and the superposition of Short-Range Dependence processes in practice. However, that Long-Range Dependence could be due to the superposition of Short-Range Dependence representing the climate system on different time scales would be physically meaningful. More work on the physical origin is needed; especially it has to be examined whether the climate system indeed has long memory, even on long time scales, or whether the observed Long-Range Dependence is the result of the superposition of short memory effects or non-linearities. Whichever of the two is the case would not only affect climate sensitivity but also the climate evolution on long time scales.

As we have shown here, scaling is an ubiquitous feature of the climate system for a multitude of time scales. Hence, it also should be included in our climate models. The parameterization problem can be seen as a model reduction problem and as discussed above the Mori-Zwanzig formulism predicts the presence of memory terms [*Gottwald et al.*, 2017; *Franzke et al.*, 2015a], however, most current weather and climate prediction models do not include memory terms [*Berner et al.*, 2017]. Recent research showed the benefit of including such memory terms [*Sakradzija et al.*, 2015; *Frederiksen et al.*, 2017; *Vissio and Lucarini*, 2018], although some difficulties in implementing the approach in simple climate models were reported [*Demaeyer and Vannitsem*, 2018]. Hence, whether memory terms in parameterization schemes are useful needs more research.

As already discussed, the presence of Long-Range Dependence hampers the detection of externally forced trends especially if the form of the trend is not a priori specified and, thus, non-parametric. Furthermore, there is also evidence of scaling breaks in temperature time series for the Holocene period [*Lovejoy and Schertzer*, 2012a] and the Central England Temperature time series [*Graves et al.*, 2015]. However, how robust these breaks are is still a matter of debate [*Nilsen et al.*, 2016] and improved statistical methods are needed. The existence of scaling breaks would also create new questions about the origin of Long-Range Dependence in climate. If Long-Range Dependence is an intrinsic property of the equations of motion then one would not expect scaling breaks; at least not without changes in external forcing or experiencing of a bifurcation (which are unlikely for the Holocene and Central England Temperature periods).

This review provided evidence for the relevance of scaling in the climate system and how it can affect the detection of trends, the estimation of climate sensitivity and the skill of long-range predictions. We also discussed various physical mechanisms which can cause scaling in the climate system.

Acknowledgments: We thank 3 anonymous reviewers for helping to improve this manuscript and Shaun Lovejoy for many discussions during the preparation of this manuscript. CF was supported through the cluster of excellence CliSAP (EXC177) and the collaborative research center 'Energy Transfers in Atmosphere and Ocean' (TRR181) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 274762653 and DFG Grant No. FR3515/3-1. We also thank PAGES and CliSAP for providing generous funding for organizing two workshops in Jouvence, Canada, and Ham-

burg, Germany, at which the idea for this review was conceived. This is a contribution to the PAGES CVAS group. Past Global Changes (PAGES) is supported by the US National Science Foundation and the Swiss Academy of Sciences. Data and code sources are provided in the text.

References

Abry, P., and D. Veitch (1998), Wavelet analysis of Long-Range Dependent traffic, *IEEE Trans. Info. Theory*, *44*(1), 2–15.

Abry, P., P. Gonçalvès, and P. Flandrin (1995), Wavelets, spectrum analysis and 1/f processes, in *Wavelets Stat.*, pp. 15–29, Springer.

Abry, P., P. Flandrin, M. Taqqu, and D. Veitch (2000), Wavelets for the analysis, estimation and synthesis of scaling data., in *Self-similar Network Traffic and Performance Evaluation*, edited by K. Park and W. Willinger, Wiley, Chichester.

Arneodo, A., B. Audit, N. Decoster, J.-F. Muzy, and C. Vaillant (2002), Wavelet based multifractal formalism: applications to DNA sequences, satellite images of the cloud structure, and stock market data, in *The Science of Disasters*, pp. 26–102, Springer.

Ashkenazy, Y., D. R. Baker, H. Gildor, and S. Havlin (2003), Nonlinearity and multifractality of climate change in the past 420,000 years, *Geophys. Res. Lett.*, *30*(22).

Baillie, R. T. (1996), Long-Memory processes and fractional integration in econometrics, *J. Econometrics*, *73*(1), 5–59.

Baillie, R. T., and S.-K. Chung (2002), Modeling and forecasting from trend-stationary Long-Memory models with applications to climatology, *I. J. Forecasting*, *18*(2), 215–226.

Bak, P. (1996), *How nature works: the science of self-organized criticality*, Springer.

Bak, P., C. Tang, and K. Wiesenfeld (1987), Self-organized criticality: An explanation of the 1/f noise, *Phys. Rev. Lett.*, *59*(4), 381.

Ball, P. (2003), The physical modelling of human social systems, *Complexus*, *1*(4), 190–206.

Baranowski, P., J. Krzyszczak, C. Slawinski, H. Hoffmann, J. Kozyra, A. Nieróbca, K. Siwek, and A. Gluza (2015), Multifractal analysis of meteorological time series to assess climate impacts, *Clim. Res.*, *65*, 39–52.

Barbosa, S. M. (2011), Testing for deterministic trends in global sea surface temperature, *J. Climate*, *24*(10), 2516–2522.

Barlow, L., J. Rogers, M. Serreze, and R. Barry (1997), Aspects of climate variability in the North Atlantic sector: Discussion and relation to the Greenland ice sheet project 2 high-resolution isotopic signal, *J. Geophys. Res.*, *102*(C12), 26,333–26,344.

Becker, M., M. Karpytchev, and S. Lennartz-Sassinek (2014), Long-term sea level trends: Natural or anthropogenic?, *Geophys. Res. Lett.*, *41*(15), 5571–5580.

Beirlant, J., Y. Goegebeur, J. Segers, and J. L. Teugels (2006), *Statistics of extremes: theory and applications*, John Wiley & Sons.

Beran, J. (1994), *Statistics for long-memory processes*, vol. 61, CRC press.

Beran, J., and Y. Feng (2002), SEMIFAR models-a semiparametric approach to modelling trends, Long-Range Dependence and nonstationarity, *Comp. Stat. Data Anal.*, *40*(2), 393–419.

Beran, J., Y. Feng, S. Ghosh, and R. Kulik (2013), *Long memory processes-probabilistic properties and statistical models*, Springer, Heidelberg.

Berkes, I., L. Horváth, P. Kokoszka, and Q.-M. Shao (2006), On discriminating between long-range dependence and changes in mean, *Annals Stat.*, pp. 1140–1165.

Berner, J., U. Achatz, L. Batte, L. Bengtsson, A. Camara, H. Christensen, M. Colangeli, D. Coleman, D. Crommelin, S. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Imkeller, H. Jarvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. Palmer, C. Penland, M. Sakradzija, J. von Storch, A. Weisheimer, M. Weniger, P. Williams, and J. Yano (2017), Stochastic parameterization: Toward a new view of weather and climate models, *Bull. Amer. Meteorol. Soc.*, *98*(3), 565–588.

Bhattacharya, R. N., V. K. Gupta, and E. Waymire (1983), The Hurst effect under trends, *J. App. Prob.*, *20*(03), 649–662.

Blender, R., and K. Fraedrich (2003), Long time memory in global warming simulations, *Geophys. Res. Lett.*, *30*(14).

Blender, R., K. Fraedrich, and B. Hunt (2006), Millennial climate variability: GCM-simulation and Greenland ice cores, *Geophys. Res. Lett.*, *33*(L04710).

Blender, R., C. Raible, and C. L. E. Franzke (2016), Vorticity and geopotential height extreme values in era-interim data during boreal winters, *Q. J. Roy. Meteorol. Soc.*

Bódai, T., and C. L. E. Franzke (2017), Predictability of extremes in heavy-tailed systems, *Phys. Rev. E*, *032120*, 96, doi:10.1103/PhysRevE.96.032120.

Bogachev, M., N. Yuan, and A. Bunde (2017), Fractals and multifractals in geophysical time series, *Fractals: Concepts and Applications in Geosciences. Ed. by Behzad Ghanbarian & Allen G. Hunt. Boca Raton: CRC Press*.

Bolster, D., R. E. Hershberger, and R. J. Donnelly (2011), Dynamic similarity, the dimensionless science, *Physics Today*, *64*(9), 42–47.

Bouchaud, J.-P., and M. Potters (2003), *Theory of financial risk and derivative pricing: from statistical physics to risk management*, Cambridge university press.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015), *Time Series Analysis: Forecasting and Control*, John Wiley & Sons.

Buckingham, E. (1914), On physically similar systems; illustrations of the use of dimensional equations, *Phys. Rev.*, *4*, 345–376, doi:10.1103/PhysRev.4.345.

Bunde, A., J. Ludescher, C. L. E. Franzke, and U. Büntgen (2014), How significant is west Antarctic warming?, *Nature Geoscience*, *7*(4), 246–247.

Burnham, K. P., and D. R. Anderson (2003), *Model selection and multimodel inference: a practical information-theoretic approach*, Springer Science & Business Media.

Caldeira, K., and N. Myhrvold (2013), Projections of the pace of warming following an abrupt increase in atmospheric carbon dioxide concentration, *Environ. Res. Lett.*, *8*(3), 034,039.

Capparelli, V., C. L. E. Franzke, A. Vecchio, M. P. Freeman, N. W. Watkins, and V. Carbone (2013), A spatiotemporal analysis of US station temperature trends over the last century, *J. Geophys. Res.*, *118*(14), 7427–7434.

Carlson, J. M., and J. Doyle (1999), Highly optimized tolerance: A mechanism for power laws in designed systems, *Physical Review E*, *60*(2), 1412.

Carlson, J. M., and J. Doyle (2000), Highly optimized tolerance: Robustness and design in complex systems, *Phys. Rev. Lett.*, *84*(11), 2529.

Carlson, J. M., and J. Doyle (2002), Complexity and robustness, *Proc. Nat. Acad. Sci.*, *99*(suppl 1), 2538–2545.

Carrion-i-Silvestre, J. L., A. Sanso-i-Rossello, and M. A. Ortuno (2001), Unit root and stationarity tests' wedding, *Econo. Lett.*, *70*(1), 1 – 8, doi:https://doi.org/10.1016/S0165-1765(00)00348-7.

Chalk, T. B., M. P. Hain, G. L. Foster, E. J. Rohling, P. F. Sexton, M. P. S. Badger, S. G. Cherry, A. P. Hasenfratz, G. H. Haug, S. L. Jaccard, A. Martínez-García, H. Pälike, R. D. Pancost, and P. A. Wilson (2017), Causes of ice age intensification across the mid-pleistocene transition, *Proc. Nat. Acad. Sci USA*, *114*(50), 13,114–13,119, doi:10.1073/pnas.1702143114.

Chandler, R., and M. Scott (2011), *Statistical methods for trend detection and analysis in the environmental sciences*, John Wiley & Sons.

Chen, X., G. Lin, and Z. Fu (2007), Long-range correlations in daily relative humidity fluctuations: A new index to characterize the climate regions over China, *Geophys. Res. Lett.*, *34*(7).

Chorin, A. J., and O. H. Hald (2013), *Stochastic tools in mathematics and science*, vol. 3, Springer.

Clark, P. U., and D. Pollard (1998), Origin of the middle pleistocene transition by ice sheet erosion of regolith, *Paleocean. Paleoclim.*, *13*(1), 1–9.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009), Power-law distributions in empirical data, *SIAM Review*, *51*(4), 661–703.

Coeurjolly, J.-F., K. Lee, and B. Vidakovic (2014), Variance estimation for fractional Brownian motions with fixed Hurst parameters, *Comm. Stat.*, *43*(8), 1845–1858.

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

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, vol. 208, Springer.

Cook, E. R., R. Seager, Y. Kushnir, K. R. Briffa, U. Büntgen, D. Frank, P. J. Krusic, W. Tegel, G. van der Schrier, L. Andreu-Hayles, et al. (2015), Old world megadroughts and pluvials during the common era, *Sci. Adv.*, *1*(10), e1500,561.

Corral, Á., A. Ossó, and J. E. Llebot (2010), Scaling of tropical-cyclone dissipation, *Nature Physics*, *6*(9), 693.

Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*(5477), 270–277, doi:10.1126/science.289.5477.270.

Crucifix, M., A. de Vernal, C. L. E. Franzke, and L. von Gunten (2017), Centennial to millennial climate variability, PAGES Magazine special issue: http://pastglobalchanges.org/products/pages-magazine/11504-25-3-centennial-millennial-clim-var.

Dahlhaus, R. (1989), Efficient parameter estimation for self-similar processes, *Annals Stat.*, *14*, 1749–1766.

Dangendorf, S., D. Rybski, C. Mudersbach, A. Müller, E. Kaufmann, E. Zorita, and J. Jensen (2014), Evidence for long-term memory in sea level, *Geophys. Res. Lett.*, *41*(15), 5530–5537.

Dansgaard, W. (1964), Stable isotopes in precipitation, *Tellus*, *16*(4), 436–468.

De Cruz, L., J. Demaeyer, and S. Vannitsem (2016), The modular arbitrary-order ocean-atmosphere model: MAOOAM v1. 0, *Geosci. Mod. Develop.*, *9*(8), 2793–2808.

Demaeyer, J., and S. Vannitsem (2018), Comparison of stochastic parameterizations in the framework of a coupled ocean-atmosphere model, *Nonlin. Proc. Geophys.*, *25*, 605–631.

Diebold, F. X., and A. Inoue (2001), Long memory and regime switching, *J. Econometrics*, *105*(1), 131–159.

Ditlevsen, P. D. (1999), Observation of alpha-stable noise induced millennial climate changes from an ice-core record, *Geophys. Res. Lett.*, *26*(10), 1441–1444.

Ditlevsen, P. D., H. Svensmark, and S. Johnsen (1996), Contrasting atmospheric and climate dynamics of the last-glacial and Holocene periods, *Nature*, *379*(6568), 810.

Doukhan, P., G. Oppenheim, and M. Taqqu (2002), *Theory and applications of long-range dependence*, Springer Science & Business Media.

Embrechts, P., and M. Maejima (2007), *Self-similar processes*, Wiley, New York.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2013), *Modelling extremal events: for insurance and finance*, vol. 33, Springer Science & Business Media.

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

Faÿ, G., E. Moulines, F. Roueff, and M. S. Taqqu (2009), Estimators of long-memory: Fourier versus wavelets, *J. Econometrics*, *151*(2), 159–177.

Feder, J. (1988), *Fractals*, Plenum Press.

Feldstein, S. B., and C. L. E. Franzke (2017), Atmospheric teleconnection patterns, in *Nonlinear and Stochastic Climate Dynamics*, edited by C. L. E. Franzke and T. O'Kane, pp. 54–104, Cambridge University Press.

Feng, T., Z. Fu, X. Deng, and J. Mao (2009), A brief description to different multi-fractal behaviors of daily wind speed records over China, *Phys. Lett. A*, *373*(45), 4134–4141.

Fisher, D. A., N. Reeh, and H. B. Clausen (1985), Stratigraphic noise in the time series derived from ice cores, *Ann. Glaciology*, *7*, 76–83.

Fox, R., and M. S. Taqqu (1986), Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals Stat.*, pp. 517–532.

Fraedrich, K., and R. Blender (2003), Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.*, *90*, 108,501, doi: 10.1103/PhysRevLett.90.108501.

Fraedrich, K., U. Luksch, and R. Blender (2004), 1/f model for long-time memory of the ocean surface temperature, *Phys. Rev. E*, *70*(3), 037,301.

Franzke, C. L. E. (2009), Multi-scale analysis of teleconnection indices: climate noise and nonlinear trend analysis, *Nonlin. Proc. Geophys.*, *16*(1), 65–76.

Franzke, C. L. E. (2010), Long-range dependence and climate noise characteristics of Antarctic temperature data, *J. Climate*, *23*(22), 6074–6081.

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

Franzke, C. L. E. (2013), Persistent regimes and extreme events of the North Atlantic atmospheric circulation, *Phil. Trans. Roy. Soc. A*, *371*(1991), 20110,471.

Franzke, C. L. E. (2014), Warming trends: Nonlinear climate change, *Nature Climate Change*, *4*(6), 423.

Franzke, C. L. E. (2017), Extremes in dynamic-stochastic systems, *Chaos*, *27*(1), 012,101.

Franzke, C. L. E., and A. J. Majda (2006), Low-order stochastic mode reduction for a prototype atmospheric GCM, *J. Atmos. Sci.*, *63*, 457–479.

Franzke, C. L. E., and T. J. O'Kane (2017), *Nonlinear and Stochastic Climate Dynamics*, Cambridge University Press.

Franzke, C. L. E., A. J. Majda, and E. Vanden-Eijnden (2005), Low-order stochastic mode reduction for a realistic barotropic model climate, *J. Atmos. Sci.*, *62*, 1722–1745.

Franzke, C. L. E., T. Woollings, and O. Martius (2011), Persistent circulation regimes and preferred regime transitions in the North Atlantic, *J. Atmos. Sci.*, *68*(12), 2809–2825.

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

Franzke, C. L. E., T. O'Kane, J. Berner, P. Williams, and V. Lucarini (2015a), Stochastic climate theory and modelling, *WIREs Climate Change*, *6*, 63–78.

Franzke, C. L. E., S. M. Osprey, P. Davini, and N. W. Watkins (2015b), A dynamical systems explanation of the Hurst effect and atmospheric low-frequency variability, *Sci. Rep.*, *5*, 9068.

Franzke, C. L. E., M. Oliver, J. Rademacher, and G. Badin (2019), Systematic multi-scale methods for geophysical flows, in *Energy transfers in atmosphere and ocean*, edited by A. Iske and C. Eden, pp. 1–51, Springer.

Frederiksen, J. S., V. Kitsios, T. J. O'Kane, and M. J. Zidikheri (2017), Stochastic subgrid modelling for geophysical and three-dimensional turbulence, in *Nonlinear and Stochastic Climate Dynamics*, edited by C. L. E. Franzke and T. O'Kane, pp. 241–275, Cambridge University Press, Cambridge.

Fredriksen, H.-B., and K. Rypdal (2016), Spectral characteristics of instrumental and climate model surface temperatures, *J. Climate*, *29*(4), 1253–1268.

Fredriksen, H.-B., and M. Rypdal (2017), Long-range persistence in global surface temperatures explained by linear multibox energy balance models, *J. Climate*, *30*(18), 7157–7168.

Frisch, U., P.-L. Sulem, and M. Nelkin (1978), A simple dynamical model of intermittent fully developed turbulence, *J. Fluid. Mech.*, *87*(4), 719–736.

Gairing, J. M., M. A. Högele, T. Kosenkova, and A. H. Monahan (2017), How close are time series to power tail Levy diffusions?, *Chaos*, *27*(7), 073,112, doi:10.1063/1.4986496.

Gan, T. Y., A. K. Gobena, and Q. Wang (2007), Precipitation of southwestern Canada: Wavelet, scaling, multifractal analysis, and teleconnection to climate anomalies, *J. Geophys. Res.*, *112*(D10).

Gao, M., and C. L. E. Franzke (2017), Quantile regression–based spatiotemporal analysis of extreme temperature change in China, *J. Climate*, *30*(24), 9897–9914.

Gardiner, C. W. (2009), *Stochastic Methods: A Handbook for the Natural and Social Sciences*, vol. 4, Springer Berlin.

Geoffroy, O., D. Saint-Martin, D. J. Olivié, A. Voldoire, G. Bellon, and S. Tytéca (2013), Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments, *J. Climate*, *26*(6), 1841–1857.

Gerlach, M., and E. G. Altmann (2019), Testing statistical laws in complex systems, *Phys. Rev. Lett.*, *122*, 168,301, doi:10.1103/PhysRevLett.122.168301.

Geweke, J., and S. Porter-Hudak (1983), The estimation and application of long memory time series models, *J. Time Ser. Anal.*, *4*(4), 221–238.

Ghil, M., and A. W. Robertson (2002), "Waves" vs."particles" in the atmosphere's phase space: A pathway to long-range forecasting?, *Proc. Nat. Acad. Sci USA*, *99*, 2493–2500.

Gil-Alana, L. (2003), An application of fractional integration to a long temperature series, *Int. J. Climatol.*, *23*(14), 1699–1710.

Gil-Alana, L. (2008), Cyclical long-range dependence and the warming effect in a long temperature time series, *Int. J. Climatol.*, *28*(11), 1435–1443.

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

Gilleland, E., and R. W. Katz (2016), extRemes 2.0: an extreme value analysis package in R, *J. Stat. Software*, *72*, 1–39, doi:10.18637/jss.v072.i08.

Giraitis, L., P. M. Robinson, and D. Surgailis (1999), Variance-type estimation of long memory, *Stoch. Proc. Appl.*, *80*(1), 1–24.

Gloter, A., M. Hoffmann, et al. (2007), Estimation of the Hurst parameter from discrete noisy data, *The Annals of Statistics*, *35*(5), 1947–1974.

Gottwald, G., D. Crommelin, and C. L. E. Franzke (2017), Stochastic climate theory, in *Nonlinear and Stochastic Climate Dynamics*, edited by C. L. E. Franzke and T. O'Kane, pp. 209–240, Cambridge University Press, Cambridge.

Granger, C. (1978), New classes of time series models, *J. Roy. Stat. Soc.*, *27*(3/4), 237–253.

Granger, C. W. (1980), Long memory relationships and the aggregation of dynamic models, *J. Econometrics*, *14*(2), 227–238.

Granger, C. W., and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing, *J. Time Series Anal.*, *1*(1), 15–29.

Graves, T., R. Gramacy, C. L. E. Franzke, and N. Watkins (2015), Efficient Bayesian inference for natural time series using ARFIMA processes, *Nonlin. Proc. Geophys.*, *22*(6), 679.

Graves, T., R. Gramacy, N. Watkins, and C. L. E. Franzke (2017a), A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951-1980, *Entropy*, *19*(9), 437, doi:10.3390/e19090437.

Graves, T., C. L. E. Franzke, N. N. Watkins, R. Gramacy, and E. Tindale (2017b), Systematic Bayesian inference of the long-range dependence and heavy-tail distribution parameters, *Physica A*, *473*, 60–71.

Guegan, D. (2005), How can we define the concept of long memory? An econometric survey, *Econ. Rev.*, *24*(2), 113–149.

Hamed, K. H. (2008), Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis, *J. Hydrology*, *349*(3), 350–363.

Hamed, K. H., and A. R. Rao (1998), A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrology*, *204*(1-4), 182–196.

Hannachi, A., D. M. Straus, C. L. E. Franzke, S. Corti, and T. Woollings (2017), Low-frequency nonlinearity and regime behavior in the Northern Hemisphere extratropical atmosphere, *Rev. Geophys.*, *55*(1), 199–234, doi:10.1002/2015RG000509, 2015RG000509.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis (2010), Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, *J. Climate*, *23*(9), 2418–2427.

Hergarten, S. (2003), Landslides, sandpiles, and self-organized criticality, *Nat. Haz. Earth Sys. Sci.*, *3*(6), 505–514.

Higuchi, T. (1990), Relationship between the fractal dimension and the power law index for a time series: a numerical investigation, *Physica D*, *46*(2), 254–264.

Hosking, J. R. M. (1981), Fractional differencing, *Biometrika*, *68*(1), 165–176, doi:10.1093/biomet/68.1.165.

Hsu, N.-J. (2006), Long-memory wavelet models, *Statistica Sinica*, pp. 1255–1271.

Huang, N. E., and Z. Wu (2008), A review on Hilbert-Huang transform: Method and its applications to geophysical studies, *Rev. Geophys.*, *46*(2).

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu (1998), The Empirical Mode Decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Roy. Soc. A*, *454*(1971), 903–995.

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

Hurst, H. E. (1957), A suggested statistical model of some time series which occur in nature, *Nature*, *180*(4584), 494.

Huybers, P., and W. Curry (2006), Links between annual, Milankovitch and continuum temperature variability, *Nature*, *441*(7091), 329–332.

Jiang, L., N. Li, and X. Zhao (2017), Scaling behaviors of precipitation over China, *Theo. Appl. Climatol.*, *128*(1-2), 63–70.

Johnsen, S. J., H. B. Clausen, K. M. Cuffey, G. Hoffmann, J. Schwander, and T. Creyts (2000), Diffusion of stable isotopes in polar firn and ice: the isotope effect in firn diffusion, in *Physics of Ice Core Records*, vol. 159, edited by T. Hondoh, pp. 121–140,

Hokkaido University Press.

Jones, P., T. Osborn, K. Briffa, C. Folland, E. Horton, L. Alexander, D. Parker, and N. Rayner (2001), Adjusting for sampling density in grid box land and ocean surface temperature time series, *J. Geophys. Res.*, *106*(D4), 3371–3380.

Jones, P. D. (1994), Hemispheric surface air temperature variations: a reanalysis and an update to 1993, *J. Climate*, *7*(11), 1794–1802.

Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Climate*, *10*(10), 2548–2568, doi:10.1175/1520-0442(1997)010<2548:ESEILS>2.0.CO;2.

Jouzel, J., V. Masson-Delmotte, O. Cattani, G. Dreyfus, S. Falourd, G. Hoffmann, B. Minster, J. Nouet, J.-M. Barnola, J. Chappellaz, et al. (2007), Orbital and millennial Antarctic climate variability over the past 800,000 years, *Science*, *317*(5839), 793–796.

Kahane, J.-P. (1985), Sur le chaos multiplicatif, *Ann. Sci. Math. Québec*, *9*(2), 105–150.

Kahane, J.-P., and J. Peyriere (1976), Sur certaines martingales de benoit mandelbrot, *Advances in mathematics*, *22*(2), 131–145.

Kantelhardt, J. W. (2009), Fractal and multifractal time series, in *Encyclopedia of Complexity and Systems Science*, pp. 3754–3779, Springer.

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

Kantelhardt, J. W., S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley (2002), Multifractal detrended fluctuation analysis of nonstationary time series, *Physica A*, *316*(1-4), 87–114.

Kantelhardt, J. W., E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin (2006), Long-term persistence and multifractality of precipitation and river runoff records, *J. Geophys. Res.*, *111*(D1).

Kavasseri, R. G., and R. Nagarajan (2005), A multifractal description of wind speed records, *Chaos, Solitons & Fractals*, *24*(1), 165–173.

Kay, S. M., and S. L. Marple (1981), Spectrum analysis-a modern perspective, *Proc. IEEE*, *69*(11), 1380–1419.

Kendall, M. G. (1948), *Rank correlation methods*, Griffin.

Kender, S., A. Ravelo, S. Worne, G. Swann, M. Leng, H. Asahi, J. Becker, H. Detlef, I. Aiello, D. Andreasen, and I. Hall (2018), Closure of the Bering Strait caused mid-

pleistocene transition cooling, *Nature Communications*, *9*, 5386.

Klein, R. (2010), Scale-dependent models for atmospheric flows, *Ann. Rev. Fluid Mech.*, *42*, 249–274.

Klemeš, V. (1974), The Hurst phenomenon: A puzzle?, *Wat. Resources Res.*, *10*(4), 675–688.

Knight, M. I., G. P. Nason, and M. A. Nunes (2016), A wavelet lifting approach to long-memory estimation, *Stat. Compu.*, pp. 1–19.

Kokoszka, P. S., and M. S. Taqqu (1994), Infinite variance stable arma processes, *J. Time Series Analys.*, *15*(2), 203–220.

Kolmogorov, A. N. (1940), Wienersche spiralen und einige andere interessante kurven in hilbertschen raum, *Acad. Sci. URSS (NS)*, *26*, 115–118.

Kolmogorov, A. N. (1962), A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number, *J. Fluid Mech.*, *13*(1), 82–85.

Kolmogorov, A. N. (1991a), The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers, *Proc. Roy. Soc. A*, *434*(1890), 9–13, doi:10.1098/rspa.1991.0075.

Kolmogorov, A. N. (1991b), Dissipation of energy in the locally isotropic turbulence, *Proc. Roy. Soc. A*, *434*(1890), 15–17, doi:10.1098/rspa.1991.0076.

Kopp, R., A. Dutton, and A. Carlson (2017), Centennial to millennial-scale sea-level change during the Holocene and last interglacial periods, *PAGES Magazine*, *25*, 148–149, doi:10.22498/pages.25.3.148.

Koscielny-Bunde, E., A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H.-J. Schellnhuber (1998), Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.*, *81*(3), 729.

Koscielny-Bunde, E., J. W. Kantelhardt, P. Braun, A. Bunde, and S. Havlin (2006), Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies, *J. Hydrology*, *322*(1-4), 120–137.

Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, *J. Hydrology*, *324*(1), 239–254.

Koutsoyiannis, D., and A. Montanari (2007), Statistical analysis of hydroclimatic time series: Uncertainty and insights, *Wat. Resources Res.*, *43*(5).

Kupferman, R. (2004), Fractional kinetics in Kac–Zwanzig heat bath models, *J. Stat. Phys.*, *114*(1-2), 291–326.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *J. Econometrics*, *54*(1-3), 159–178.

Laepple, T., and P. Huybers (2013), Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability, *Earth Plan. Sci. Lett.*, *375*, 418–429.

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

Laepple, T., and P. Huybers (2014b), Ocean surface temperature variability: Large model-data differences at decadal and longer periods, *Proc. Nat. Acad. Sci. USA*, *111*(47), 16,682–16,687, doi:10.1073/pnas.1412077111.

Laepple, T., T. Münch, M. Casado, M. Hoerhold, A. Landais, and S. Kipfstuhl (2017), On the similarity and apparent cycles of isotopic variations in east Antarctic snow-pits, *Cryosphere Discuss.*, *2017*, 1–29, doi:10.5194/tc-2017-199.

Lambert, F., B. Delmonte, J.-R. Petit, M. Bigler, P. R. Kaufmann, M. A. Hutterli, T. F. Stocker, U. Ruth, J. P. Steffensen, and V. Maggi (2008), Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core, *Nature*, *452*(7187), 616–619.

Lamperti, J. (1962), Semi-stable stochastic processes, *Trans. Amer. Math. Soc.*, *104*(1), 62–78.

Lee, D., and P. Schmidt (1996), On the power of the KPSS test of stationarity against fractionally-integrated alternatives, *J. Econometrics*, *73*(1), 285–302.

Lennartz, S., and A. Bunde (2009), Trend evaluation in records with long-term memory: Application to global warming, *Geophys. Res. Lett.*, *36*(16).

Lennartz, S., and A. Bunde (2011), Distribution of natural trends in long-term correlated records: A scaling approach, *Phys. Rev. E*, *84*(2), 021,129.

Lévy, P. (1953), *Random functions: general theory with special reference to Laplacian random functions*, 12, University of California Press.

Leybourne, S. J., and P. Newbold (1999), The behaviour of Dickey–Fuller and Phillips–Perron tests under the alternative hypothesis, *Econ. Journal*, *2*(1), 92–100.

Lin, G., and Z. Fu (2008), A universal model to characterize different multi-fractal behaviors of daily temperature records over China, *Physica A*, *387*(2-3), 573–579.

Lobato, I., and P. M. Robinson (1996), Averaged periodogram estimation of long memory, *J. Econometrics*, *73*(1), 303–324.

Longair, M. S. (2003), *Theoretical concepts in physics: an alternative view of theoretical reasoning in physics*, Cambridge University Press.

Lorenz, E. N. (1968), Climatic determinism, in *Causes of climatic change*, pp. 1–3, Springer.

Lorenz, E. N. (1976), Nondeterministic theories of climatic change, *Quat. Res.*, *6*(4), 495–506.

Lovejoy, S. (2014), Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming, *Clim. Dyn.*, *42*(9-10), 2339–2351.

Lovejoy, S. (2015a), A voyage through scales, a missing quadrillion and why the climate is not what you expect, *Clim. Dyn.*, *44*(11-12), 3187–3210.

Lovejoy, S. (2015b), Using scaling for macroweather forecasting including the pause, *Geophys. Res. Lett.*, *42*(17), 7148–7155, doi:10.1002/2015GL065665, 2015GL065665.

Lovejoy, S. (2018), Spectra, intermittency, and extremes of weather, macroweather and climate, *Sci. Rep.*, *8*(1), 12,697.

Lovejoy, S., and D. Schertzer (2012a), Low-frequency weather and the emergence of the climate, *Extreme Events and Natural Hazards: The Complexity Perspective*, pp. 231–254.

Lovejoy, S., and D. Schertzer (2012b), Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, *Nonlin. Proc. Geophys.*, *19*(5), 513–527.

Lovejoy, S., and D. Schertzer (2013), *The weather and climate: emergent laws and multifractal cascades*, Cambridge University Press.

Lovejoy, S., and C. Varotsos (2016), Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings, *Earth Sys. Dyn.*, *7*(1), 133–150.

Lovejoy, S., A. Tuck, S. Hovde, and D. Schertzer (2007), Is isotropic turbulence relevant in the atmosphere?, *Geophys. Res. Lett.*, *34*(15).

Lovejoy, S., A. Tuck, D. Schertzer, and S. Hovde (2009), Reinterpreting aircraft measurements in anisotropic scaling turbulence, *Atmos. Chem. Phys.*, *9*(14), 5007–5025.

Lovejoy, S., L. d. Rio Amador, and R. Hébert (2015), The ScaLIng Macroweather Model (SLIMM): using scaling to forecast global-scale macroweather from months to decades, *Earth Sys. Dyn.*, *6*(2), 637–658.

Lovejoy, S., L. D. R. Amador, and R. Hébert (2018), Harnessing butterflies: theory and practice of the stochastic seasonal to interannual prediction system (StocSIPS), in *Adv.*

Nonlin. Geosci., pp. 305–355, Springer.

Løvsletten, O. (2017), Consistency of detrended fluctuation analysis, *Phys. Rev. E*, *96*(1), 012,141.

Løvsletten, O., and M. Rypdal (2016), Statistics of regional surface temperatures after 1900: Long-range versus short-range dependence and significance of warming trends, *J. Climate*, *29*(11), 4057–4068.

Lucarini, V., D. Faranda, A. C. Freitas, J. M. Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, and S. Vaienti (2016), *Extremes and Recurrence in Dynamical Systems*, John Wiley & Sons.

Ludescher, J., A. Bunde, C. L. E. Franzke, and H. J. Schellnhuber (2015), Long-term persistence enhances uncertainty about anthropogenic warming of Antarctica, *Clim. Dyn.*, *46*, 263–271.

Majda, A., and X. Wang (2006), *Nonlinear dynamics and statistical theories for basic geophysical flows*, Cambridge University Press.

Majda, A., C. L. E. Franzke, and D. Crommelin (2009), Normal forms for reduced stochastic climate models, *Proc. Natl. Acad. Sci. USA*, *106*, 3649–3653, doi:doi:10.1073/pnas.0900173106.

Majda, A. J., I. Timofeyev, and E. V. Eijnden (1999), Models for stochastic climate prediction, *Proc. Nat. Acad. Sci. USA*, *96*(26), 14,687–14,691.

Majda, A. J., I. Timofeyev, and E. Vanden Eijnden (2001), A mathematical framework for stochastic climate models, *Comm. Pure Appl. Math.*, *54*(8), 891–974.

Majda, A. J., C. L. E. Franzke, and B. Khouider (2008), An applied mathematics perspective on stochastic modelling for climate, *Phil. Trans. R. Soc. A*, *366*, 2429–2455, doi:10.1098/rsta.2008.0012.

Mandelbrot, B. (1963), The variation of certain speculative prices, *J. Business*, *36*(4), 394–419.

Mandelbrot, B. B. (1965), Une classe de processus stochastiques homothetiques a soi-application a la loi climatologique de he Hurst, *Comptes Rendus Hebdomadaires des seances de l'academie des sciences*, *260*(12), 3274–3277.

Mandelbrot, B. B. (1974), Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, *J. Fluid. Mech.*, *62*(2), 331–358.

Mandelbrot, B. B. (1982), *The fractal geometry of nature*, WH freeman New York.

Mandelbrot, B. B., and M. S. Taqqu (1979), *Robust R/S analysis of long run serial correlation*, vol. 48, 69–104 pp., IBM Thomas J. Watson Research Division.

Mandelbrot, B. B., and J. W. Van Ness (1968), Fractional Brownian motions, fractional noises and applications, *SIAM Review*, *10*(4), 422–437.

Mandelbrot, B. B., and J. R. Wallis (1968), Noah, Joseph, and operational hydrology, *Wat. Resources Res.*, *4*(5), 909–918.

Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, pp. 245–259.

Mann, M. E. (2011), On long range dependence in global surface temperature series, *Clim. Change*, *107*(3), 267–276.

Mantegna, R. N., and H. E. Stanley (1999), *Introduction to econophysics: correlations and complexity in finance*, Cambridge University Press.

Mantua, N. J., and S. R. Hare (2002), The Pacific Decadal Oscillation, *J. Oceano.*, *58*(1), 35–44.

Maraun, D., H. Rust, and J. Timmer (2004), Tempting long-memory-on the interpretation of DFA results, *Nonlin. Proc. Geophys.*, *11*(4), 495–503.

Maslov, L. A. (2014), Self-organization of the earth's climate system versus Milankovitch-Berger astronomical cycles, *J. Adv. Mol. Earth Sys.*, *6*(3), 650–657.

McNeil, A. J., R. Frey, and P. Embrechts (2015), *Quantitative risk management: Concepts, techniques and tools*, Princeton University Press.

Meinsma, G. (2019), Dimensional and scaling analysis, *SIAM Review*, *61*(1), 159–184.

Mesa, O., V. Gupta, and P. O'Connell (2012), Dynamical system exploration of the Hurst phenomenon in simple climate models, *Extreme Events and Natural Hazards: The Complexity Perspective*, pp. 209–230.

Mills, T. C. (2007), Time series modelling of two millennia of northern Hemisphere temperatures: long memory or shifting trends?, *J. Roy. Stat. Soc. A*, *170*(1), 83–94.

Mitchell, J. M. (1976), An overview of climatic variability and its causal mechanisms, *Quat. Res.*, *6*(4), 481–493.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data, *Nature*, *433*(7026), 613.

Mori, H. (1965), Transport, collective motion, and Brownian motion, *Progress Theor. Phys.*, *33*(3), 423–455.

Moritz, M. A., M. E. Morais, L. A. Summerell, J. Carlson, and J. Doyle (2005), Wildfires, complexity, and highly optimized tolerance, *Proceedings of the National Academy of Sciences*, *102*(50), 17,912–17,917.

Mudelsee, M. (2013), *Climate time series analysis*, Springer.

Münch, T., and T. Laepple (2018), What climate signal is contained in decadal-to centennial-scale isotope variations from Antarctic ice cores?, *Climate of the Past*, *14*(12), 2053–2070.

Munk, W., and C. Wunsch (1998), Abyssal recipes ii: Energetics of tidal and wind mixing, *Deep-sea Res.*, *45*(12), 1977–2010.

Muzy, J.-F., E. Bacry, and A. Arneodo (1991), Wavelets and multifractal formalism for singular signals: Application to turbulence data, *Phys. Rev. Lett.*, *67*(25), 3515.

Myrvoll-Nilsen, E., H.-B. Fredriksen, S. H. Sørbye, and M. Rypdal (2019), Warming trends and long-range dependent climate variability since year 1900: a Bayesian approach, *Front. Earth Sci.*, *7*, 214.

Nastrom, G., and K. S. Gage (1985), A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft, *J. Atmos. Sci.*, *42*(9), 950–960.

Newman, M., G. P. Compo, and M. A. Alexander (2003), Enso-forced variability of the pacific decadal oscillation, *J. Climate*, *16*(23), 3853–3857.

Newman, M., M. A. Alexander, T. R. Ault, K. M. Cobb, C. Deser, E. D. Lorenzo, N. J. Mantua, A. J. Miller, S. Minobe, H. Nakamura, N. Schneider, D. J. Vimont, A. S. Phillips, J. D. Scott, and C. A. Smith (2016), The pacific decadal oscillation, revisited, *J. Climate*, *29*(12), 4399–4427, doi:10.1175/JCLI-D-15-0508.1.

Newman, M. E. (2005), Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, *46*(5), 323–351.

Nicolis, C. (1990), Chaotic dynamics, Markov processes and climate predictability, *Tellus*, *42*(4), 401–412.

Nicolis, C., and G. Nicolis (1995), Chaos in dissipative systems: Understanding atmospheric physics, *Adv. Chem. Phys.*, *91*, 511–570.

Nicolis, C., W. Ebeling, and C. Baraldi (1997), Markov processes, dynamic entropies and the statistical prediction of mesoscale weather regimes, *Tellus*, *49*(1), 108–118.

Nicolis, G., and C. Nicolis (1988), Master-equation approach to deterministic chaos, *Phys. Rev. A*, *38*(1), 427.

Nilsen, T., K. Rypdal, and H.-B. Fredriksen (2016), Are there multiple scaling regimes in Holocene temperature records?, *Earth Sys. Dyn.*, *7*(2), 419–439.

Novikov, E. A., and R. Stewart (1964), Intermittency of turbulence and spectrum of fluctuations in energy-disspation, *Izv. Akad. Nauk. SSSR. Ser. Geofiz*, *3*, 408–412.

Obukhov, A. (1962), Fluctuations of the energy dissipation in turbulence, *J. Geophys. Res*, *67*, 3011.

Önskog, T., C. L. E. Franzke, and A. Hannachi (2018), Predictability and non-Gaussian characteristics of the North Atlantic oscillation, *J. Climate*, *31*(2), 537–554, doi: 10.1175/JCLI-D-17-0101.1.

Önskog, T., C. L. E. Franzke, and A. Hannachi (2019), Nonlinear time series models for the North Atlantic Oscillation, *Adv. Stat. Clim. Atm. Ocean.*, p. submitted.

Østvand, L., T. Nilsen, K. Rypdal, D. Divine, and M. Rypdal (2014), Long-range memory in internal and forced dynamics of millennium-long climate model simulations, *Earth Sys. Dyn.*, *5*(2), 295.

Paillard, D. (2001), Glacial cycles: toward a new paradigm, *Reviews of Geophysics*, *39*(3), 325–346.

Palma, W. (2007), *Long-memory time series: theory and methods*, vol. 662, John Wiley & Sons.

Parker, D., C. Folland, and M. Jackson (1995), Marine surface temperature: observed variations and data requirements, *Clim. Change*, *31*(2), 559–600.

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

Peavoy, D., and C. L. E. Franzke (2010), Bayesian analysis of rapid climate change during the last glacial using Greenland δ18O data, *Clim. Past*, *6*(6), 787–794.

Pelletier, J. D. (1997), Analysis and modeling of the natural variability of climate, *J. Climate*, *10*(6), 1331–1342.

Pelletier, J. D. (1998), The power spectral density of atmospheric temperature from time scales of 10^{-2} to 10^{6} yr, *Earth Plan. Sci. Lett.*, *158*(3), 157–164.

Peng, C.-K., S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger (1994), Mosaic organization of DNA nucleotides, *Phys. Rev. E*, *49*(2), 1685.

Penland, C., and P. D. Sardeshmukh (2012), Alternative interpretations of power-law distributions found in nature, *Chaos*, *22*(2), 023,119.

Percival, D. B., and P. Guttorp (1994), Long-memory processes, the Allan variance and wavelets, *Wavelets Geophys.*, *4*, 325–344.

Percival, D. B., and D. A. Rothrock (2005), 'Eyeballing' trends in climate time series: A cautionary note, *J. Climate*, *18*(6), 886–891, doi:10.1175/JCLI-3300.1.

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

Peters, D. P., R. A. Pielke, B. T. Bestelmeyer, C. D. Allen, S. Munson-McGee, and K. M. Havstad (2004), Cross-scale interactions, nonlinearities, and forecasting catastrophic events, *Proc. Nat. Acad. Sci.*, *101*(42), 15,130–15,135.

Peters, O., and K. Christensen (2006), Rain viewed as relaxational events, *J. Hydrology*, *328*(1-2), 46–55.

Peters, O., and J. D. Neelin (2006), Critical phenomena in atmospheric precipitation, *Nature Physics*, *2*(6), 393.

Peters, O., C. Hertlein, and K. Christensen (2001), A complexity view of rainfall, *Phys. Rev. Lett.*, *88*(1), 018,701.

Peters, O., A. Deluca, A. Corral, J. D. Neelin, and C. E. Holloway (2010), Universality of rain event size distributions, *J. Stat. Mech.*, *2010*(11), P11,030.

Phillips, P. C., and P. Perron (1988), Testing for a unit root in time series regression, *Biometrika*, pp. 335–346.

Qiu, B., N. Schneider, and S. Chen (2007), Coupled decadal variability in the north pacific: An observationally constrained idealized model, *J. Climate*, *20*(14), 3602–3620.

Rea, W., L. Oxley, M. Reale, and J. Brown (2009), Estimators for long range dependence: an empirical study, *arXiv preprint arXiv:0901.0762*.

Rehfeld, K., T. Münch, S. L. Ho, and T. Laepple (2018), Global patterns of declining temperature variability from the last glacial maximum to the Holocene, *Nature*, *554*(7692), 356.

Rial, J. A., R. A. Pielke, M. Beniston, M. Claussen, J. Canadell, P. Cox, H. Held, N. De Noblet-Ducoudré, R. Prinn, J. F. Reynolds, et al. (2004), Nonlinearities, feedbacks and critical thresholds within the earth's climate system, *Climatic Change*, *65*(1-2), 11–38.

Riedi, R. H., M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk (1999), A multifractal wavelet model with application to network traffic, *IEEE Trans. Info. Theo.*, *45*(3), 992–1018.

Robinson, S. A., S. Black, B. W. Sellwood, and P. J. Valdes (2006), A review of palaeoclimates and palaeoenvironments in the Levant and Eastern Mediterranean from 25,000 to 5000 years BP: setting the environmental background for the evolution of human civilisation, *Quat. Sci. Rev.*, *25*(13-14), 1517–1541.

Royer, J.-F., A. Biaou, F. Chauvin, D. Schertzer, and S. Lovejoy (2008), Multifractal analysis of the evolution of simulated precipitation over france in a climate scenario, *Comptes Rendus Geoscience*, *340*(7), 431–440.

Rust, H., O. Mestre, and V. Venema (2008), Fewer jumps, less memory: Homogenized temperature records and long memory, *J. Geophys. Res.*, *113*(D19).

Rybski, D., and A. Bunde (2009), On the detection of trends in long-term correlated records, *Physica A*, *388*(8), 1687–1695.

Rypdal, K., and M. Rypdal (2016a), Comment on" scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcing" by S. Lovejoy and C. Varotsos (2016), *Earth Sys. Dyn.*, *7*(3), 597–609.

Rypdal, K., M. Rypdal, and H.-B. Fredriksen (2015), Spatiotemporal long-range persistence in earth's temperature field: Analysis of stochastic–diffusive energy balance models, *J. Climate*, *28*(21), 8379–8395.

Rypdal, M., and K. Rypdal (2014), Long-memory effects in linear response models of earth's temperature and implications for future global warming, *J. Climate*, *27*(14), 5240–5258.

Rypdal, M., and K. Rypdal (2016b), Late quaternary temperature variability described as abrupt transitions on a 1/f noise background, *Earth Sys. Dyn.*, *7*(1), 281–293.

Rypdal, M., H.-B. Fredriksen, E. Myrvoll-Nilsen, K. Rypdal, and S. H. Sorbye (2018), Emergent scale invariance and climate sensitivity, *Climate*, *6*(4), doi:10.3390/cli6040093.

Saichev, A. I., Y. Malevergne, and D. Sornette (2009), *Theory of Zipf's law and beyond*, vol. 632, Springer Science & Business Media.

Sakradzija, M., A. Seifert, and T. Heus (2015), Fluctuations in a quasi-stationary shallow cumulus cloud ensemble, *Nonlin. Proc. Geophys.*, *22*(1), 65–85.

Samorodnitsky, G. (2007), Long range dependence, *Found. Trends Stochas. Sys.*, *1*(3), 163–257.

Samorodnitsky, G. (2016), *Stochastic Processes and Long Range Dependence*, Springer.

Sardeshmukh, P. D., and C. Penland (2015), Understanding the distinctively skewed and heavy tailed character of atmospheric and oceanic probability distributions, *Chaos*, *25*(3), 036,410.

Sardeshmukh, P. D., and P. Sura (2009), Reconciling non-Gaussian climate statistics with linear dynamics, *J. Climate*, *22*(5), 1193–1207.

Schertzer, D., and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, *J. Geophys. Res.*, *92*(D8), 9693–9714.

Schmitt, F., S. Lovejoy, and D. Schertzer (1995), Multifractal analysis of the Greenland ice-core project climate data, *Geophys. Res. Lett.*, *22*(13), 1689–1692.

Schmitt, F. G., and Y. Huang (2016), *Stochastic analysis of scaling time series: from turbulence theory to applications*, Cambridge University Press.

Schmittbuhl, J., J.-P. Vilotte, and S. Roux (1995), Reliability of self-affine measurements, *Phys. Rev. E*, *51*(1), 131.

Schneider, N., and B. D. Cornuelle (2005), The forcing of the pacific decadal oscillation, *J. Climate*, *18*(21), 4355–4373.

Snyder, C. W. (2016), Evolution of global temperature over the past two million years, *Nature*, *538*, 226–228.

Sornette, D. (2006), *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*, second ed., Springer Science & Business Media.

Stoev, S., and M. S. Taqqu (2005), Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations, *J. Time Ser. Anal.*, *26*(2), 211–249.

Stoev, S., M. S. Taqqu, C. Park, G. Michailidis, and J. S. Marron (2006), LASS: a tool for the local analysis of self-similarity, *Comp. Stat. Data Anal.*, *50*(9), 2447–2471.

Straus, D. M., and P. Ditlevsen (1999), Two-dimensional turbulence properties of the ECMWF reanalyses, *Tellus*, *51*(5), 749–772.

Sura, P. (2011), A general perspective of extreme events in weather and climate, *Atmos. Res.*, *101*(1), 1–21.

Sura, P. (2013), Stochastic models of climate extremes: Theory and observations, in *Extremes in a Changing Climate*, pp. 181–222, Springer.

Sura, P., and A. Hannachi (2015), Perspectives of non-Gaussianity in atmospheric synoptic and low-frequency variability, *J. Climate*, *28*(13), 5091–5114.

Sura, P., and P. D. Sardeshmukh (2008), A global view of non-gaussian sst variability, *J. Phys. Oceano.*, *38*(3), 639–647.

Talkner, P., and R. O. Weber (2000), Power spectrum and detrended fluctuation analysis: Application to daily temperatures, *Phys. Rev. E*, *62*(1), 150.

Tamazian, A., J. Ludescher, and A. Bunde (2015), Significance of trends in long-term correlated records, *Phys. Rev. E*, *91*(3), 032,806.

Taqqu, M. S. (2013), Benoît mandelbrot and fractional brownian motion, *Statistical Science*, *28*(1), 131–134.

Taqqu, M. S., V. Teverovsky, and W. Willinger (1995), Estimators for long-range dependence: an empirical study, *Fractals*, *3*(04), 785–798.

Vallis, G. K. (2017), *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*, Cambridge University Press.

Vannitsem, S. (2001), Toward a phase-space cartography of the short-and medium-range predictability of weather regimes, *Tellus*, *53*(1), 56–73.

Vannitsem, S., J. Demaeyer, L. De Cruz, and M. Ghil (2015), Low-frequency variability and heat transport in a low-order nonlinear coupled ocean–atmosphere model, *Physica D*, *309*, 71–85.

Varotsos, C., and D. Kirk-Davidoff (2006), Long-memory processes in ozone and temperature variations at the region 60S–60N, *Atmos. Chem. Phys.*, *6*(12), 4093–4100.

Veenstra, J. Q. (2012), Persistence and anti-persistence: Theory and software, Ph.D. thesis, Western University.

Vimont, D. J. (2005), The contribution of the interannual enso cycle to the spatial pattern of decadal enso-like variability, *J. Climate*, *18*(12), 2080–2092.

Vissio, G., and V. Lucarini (2018), A proof of concept for scale-adaptive parametrizations: the case of the Lorenz'96 model, *Q. J. Roy. Meteorol. Soc.*, *144*(710), 63–75.

von Storch, H., and F. W. Zwiers (2003), *Statistical analysis in climate research*, Cambridge University Press.

Vyushin, D., I. Zhidkov, S. Havlin, A. Bunde, and S. Brenner (2004), Volcanic forcing improves atmosphere-ocean coupled general circulation model scaling performance, *Geophys. Res. Lett.*, *31*(10), L10,206, doi:10.1029/2004GL019499, l10206.

Vyushin, D. I., and P. J. Kushner (2009), Power-law and long-memory characteristics of the atmospheric general circulation, *J. Climate*, *22*(11), 2890–2904.

Vyushin, D. I., V. E. Fioletov, and T. G. Shepherd (2007), Impact of long-range correlations on trend detection in total ozone, *J. Geophys. Res.*, *112*(D14).

Wang, G., T. Jiang, R. Blender, and K. Fraedrich (2008), Yangtze 1/f discharge variability and the interacting river–lake system, *J. Hydrology*, *351*(1-2), 230–237.

Watkins, N. W., G. Pruessner, S. C. Chapman, N. B. Crosby, and H. J. Jensen (2016), 25 years of self-organized criticality: concepts and controversies, *Space Science Reviews*, *198*(1-4), 3–44.

Wilks, D. S. (2011), *Statistical methods in the atmospheric sciences*, vol. 100, Academic press.

Williams, P. D., M. J. Alexander, E. A. Barnes, A. H. Butler, H. C. Davies, C. I. Garfinkel, Y. Kushnir, T. P. Lane, J. K. Lundquist, O. Martius, R. N. Maue, W. R. Peltier, K. Sato, A. A. Scaife, and C. Zhang (2017), A census of atmospheric variability from seconds to decades, *Geophys. Res. Lett.*, *44*(21), 11,201–11,211, doi: 10.1002/2017GL075483, 2017GL075483.

Willinger, W., D. Alderson, J. C. Doyle, and L. Li (2004), More" normal" than normal: scaling distributions and complex systems, in *Simulation Conference, 2004. Proceedings of the 2004 Winter*, vol. 1, IEEE.

Witt, A., and B. D. Malamud (2013), Quantification of long-range persistence in geophysical time series: conventional and benchmark-based improvement techniques, *Surveys in Geophysics*, *34*(5), 541–651.

Wu, Z., N. E. Huang, S. R. Long, and C.-K. Peng (2007), On the trend, detrending, and variability of nonlinear and nonstationary time series, *Proc. Nat. Acad. Sci USA*, *104*(38), 14,889–14,894.

Wunsch, C. (1999), The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations, *Bull. Am. Meteorol. Soc.*, *80*(2), 245–255.

Wunsch, C. (2003), The spectral description of climate change including the 100 ky energy, *Clim. Dyn.*, *20*(4), 353–363.

Wunsch, C., and R. Ferrari (2004), Vertical mixing, energy, and the general circulation of the oceans, *Annu. Rev. Fluid Mech.*, *36*, 281–314.

Yang, L., and Z. Fu (2019), Process-dependent persistence in precipitation records, *Physica A*, *527*, 121,459.

Yang, L., C. L. E. Franzke, and Z. Fu (2019), Power-law behavior of hourly precipitation in intensity and dry spell durations over the US, *Int. J. Climatol.*, *8*, 1–16, doi:10.1002/

joc.6343.

Yuan, N., Z. Fu, and J. Mao (2013a), Different multi-fractal behaviors of diurnal temperature range over the north and the south of China, *Theo. Appl. Climatol.*, *112*(3-4), 673–682.

Yuan, N., Z. Fu, and S. Liu (2013b), Long-term memory in climate variability: A new look based on fractional integral techniques, *J. Geophys. Res.*, *118*(23).

Yuan, N., Z. Fu, and S. Liu (2014), Extracting climate memory using fractional integrated statistical model: A new perspective on climate prediction, *Sci. Rep.*, *4*.

Yuan, N., M. Ding, J. Ludescher, and A. Bunde (2017), Increase of the Antarctic sea ice extent is highly significant only in the ross sea, *Sci. Rep.*, *7*, 41,096.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel (2019), What is the predictability limit of midlatitude weather?, *J. Atmos. Sci.*, *76*(4), 1077–1091.

Zhu, F., J. Emile-Geay, N. P. McKay, G. J. Hakim, D. Khider, T. R. Ault, E. J. Steig, S. Dee, and J. W. Kirchner (2019), Climate models can correctly simulate the continuum of global-average temperature variability, *Proc. Nat. Acad. Sci. USA*, *116*(18), 8728–8733, doi:10.1073/pnas.1809959116.

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

Zwanzig, R. (1973), Nonlinear generalized langevin equations, *J. Stat. Phys.*, *9*(3), 215–220.

Zwanzig, R. (2001), *Nonequilibrium statistical mechanics*, Oxford University Press, USA.

Glossary

Brownian Motion Brownian motion is a continuous-time stochastic process, also called Wiener process. The increments of Brownian motion are Gaussian distributed independent random variables.

Fractal A fractal is a self-similar object, i.e. when shrinking or enlarging a fractal pattern, its appearance remains statistically unchanged. A good introduction is given by *Feder* [1988].

Heavy-tailed Distribution Heavy-tailed distributions are distributions whose tail decays slower than exponential. In particular, its tail is heavier than for a corresponding Gaussian distribution. A good introduction is given by *Sornette* [2006].

Leptokurtic A leptokurtic distribution has a kurtosis which is larger than 3. The Gaussian distribution has a kurtosis of 3. Hence, a leptokurtic distribution has fatter tails than the corresponding Gaussian distribution.

Long Memory Synonym for Long-Range Dependence

Long-Term Persistence (LTP) Synonym for Long-Range Dependence

Long-Range Dependence Long-Range Dependence is the property of the autocorrelation function of a time series to decay according to a power-law. Consequently, the power spectrum of such a time series has increasing power for lower frequencies and a singularity at zero frequency.

Mono-fractal Mono-fractals are fractals described by a single scaling exponent

Multi-fractal Multi-fractals are fractals described by multiple scaling relationships and whose exponents are functions of scale

Power Law A power law describes a functional relationship between two variables where a change in one variable results in a proportional relative change in the other variable. Mathematically it is of the following form: $f(x) = ax^{-k}$ where k is the power law exponent.

Random Walk Also known as Drunkards walk, has scaling power spectrum with slope -2, variance increases as $\sqrt{t}$. A good introduction is given by *Sornette* [2006].

Red Noise Red noise means that the spectral power increases on longer time scales. However, in climate science red noise typically denotes the power spectrum of a first order autoregressive process which has first increasing power for increasing period but then becomes white noise; also called Lorentzian spectrum

Scaling In mathematical form we can express scaling as follows:

$$F(at) \stackrel{d}{=} a^{\gamma} F(t), \tag{14}$$

where F is the fluctuation function, γ the scaling exponent, a is a rescaling factor of time t and $\stackrel{d}{=}$ denotes equality in distribution. a can be seen as the factor with which one is zooming in or out and the scaling property now means that the statistical properties of the data stay the same [*Mandelbrot*, 1982; *Feder*, 1988; *Franzke et al.*, 2012] and this is the same property as fractals have.

Short-Range Dependence Short-Range Dependence is the property of the autocorrelation function of a time series to decay exponentially. Consequently, the power spectrum of such a time series has almost constant power at lower frequencies.

Self-Similarity A self-similar object is exactly or approximately similar to a part of itself. When zooming in or out one sees similar structures. Self-similarity is a property of Fractals.

Unit Root A Unit Root is a characteristic of stochastic processes. In particular, it denotes that a stochastic process is non-stationary without necessarily having a trend. A good introduction is given by *Box et al.* [2015].

Volatility clustering Volatility clustering refers to the observation that in many time series large changes are followed by large changes of either sign, while small changes are followed by small changes of either sign.

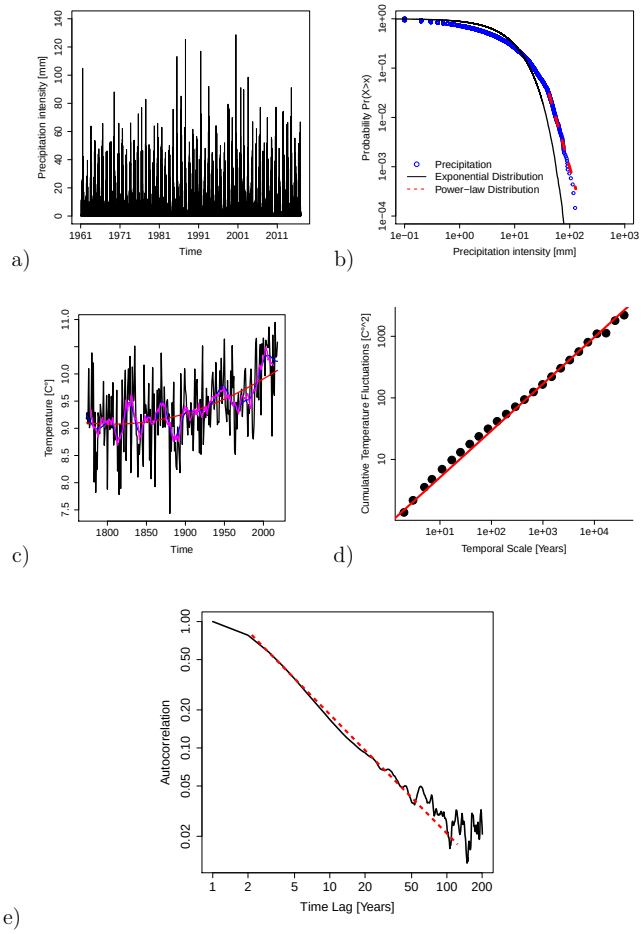
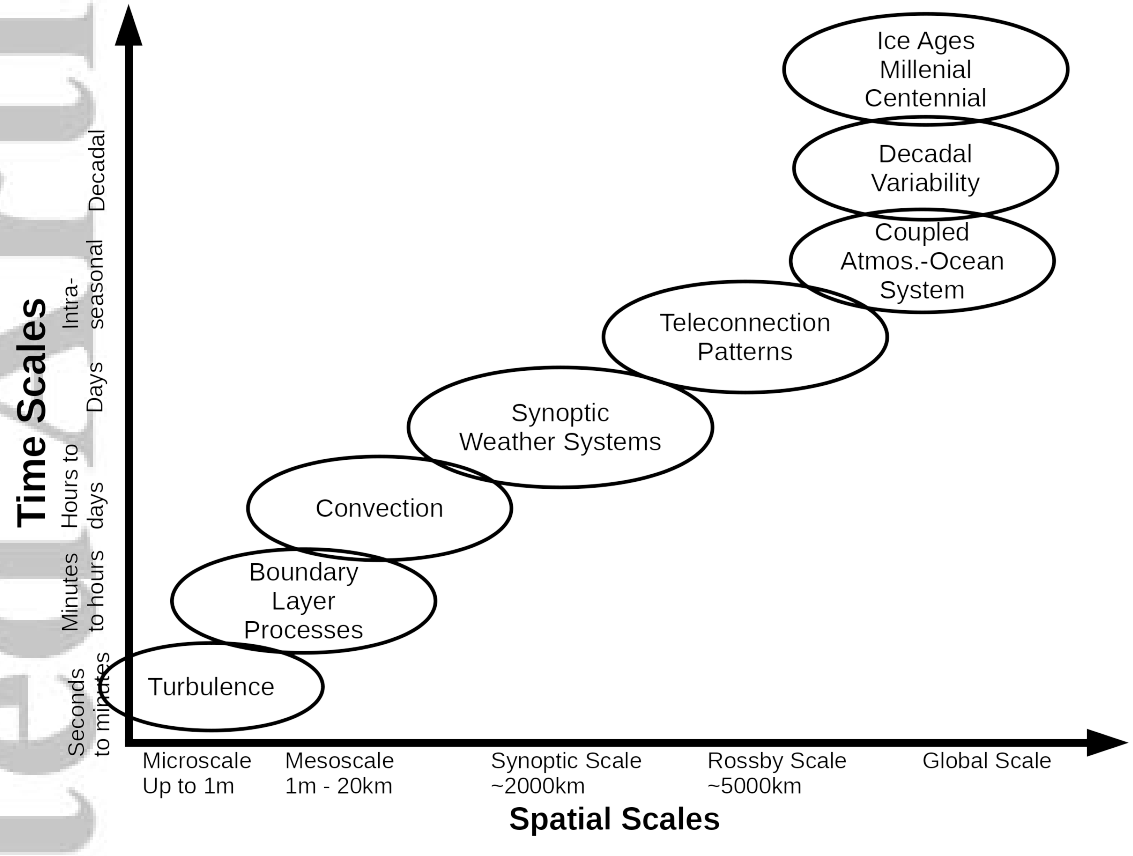


Figure 1. a) Daily precipitation at Xichang, China, b) Probability Density Function of precipitation (Red dashed line: corresponding power-law fit with exponent 4.97; black dashed line: corresponding exponential Probability Density Function with parameter 8.21); c) Annual mean Central England Temperature (CET). Red line: Non-Linear trend, Magenta line: 11-year running mean and blue line: Decadal scale fluctuations as derived from an Empirical Mode Decomposition (EMD), and d) Detrended Fluctuation Analysis (DFA) plot with d=0.25: Circles: fluctuation function and red line: straight line with slope 0.25. e) Autocorrelation function of CET (Black line) and the red dashed line indicates a power-law decay.



2076 **Figure 2.** Schematic diagram of important spatial and temporal scales in the climate system. The solid
2077 lines denotes an estimate of the relative variance of climate variability. The dashed lines denotes the vari-
2078 ance contribution to the total variance from climatic processes with characteristic spatial scales smaller than
2079 those indicated on the x-axis. The periodic climate components are denoted by spikes of arbitrary width. See
2080 *Mitchell* [1976] for more details. Figure source *Mitchell* [1976].

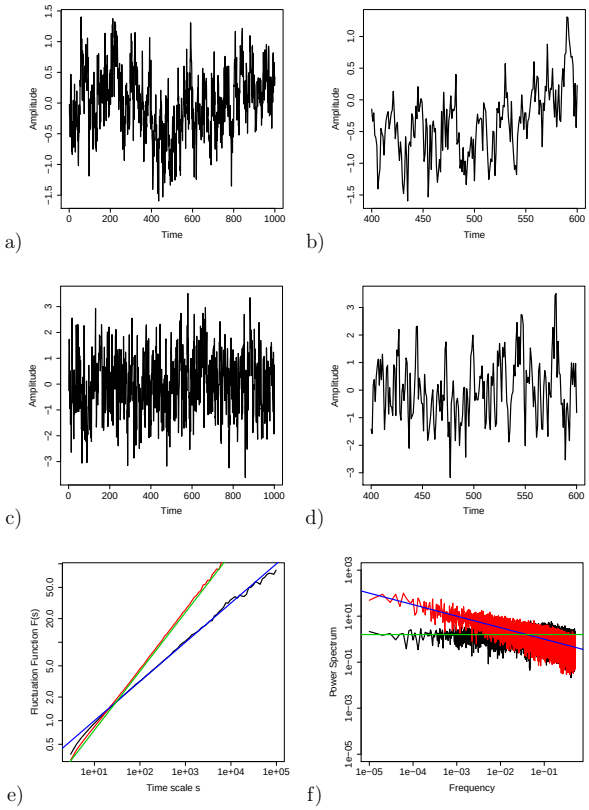
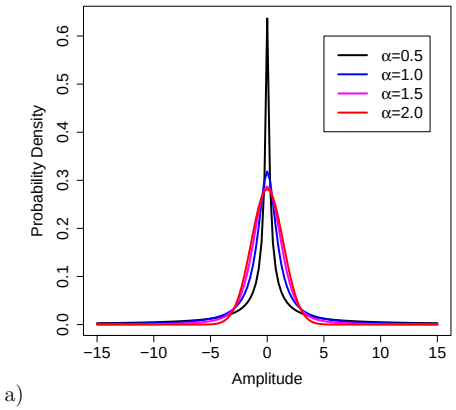
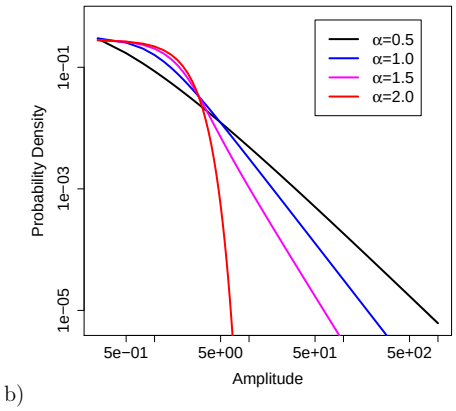


Figure 3. Time series with scaling and non-scaling behavior. a) A time series with scaling behavior (Long-term persistence parameter d=0.495) and b) zooms in the time period between 400 and 600 time units of a). After zooming in, the time series in b) shows a similar pattern as the time series in a). c) A time series without scaling behavior (First order autoregressive process $x_{t+1} = 0.5x_t + \zeta_t$) and d) zooms in the time period between 400 and 600 time units of c). e) Fluctuation Functions for a Short-Term Dependent process (first order autoregressive process) (Black line) and scaling model in form of a Long-Term Dependent process (Red line) with regression lines with slopes of 0.5, which corresponds to d=0.0 (blue line), and slope of 0.75, which corresponds to d=0.25 (green line). f) Power spectrum of the Short-Term Dependent process (black), and the Long-Term Dependent process (red) plotted in a) and c). The blue line is the theoretical slope line of a Long-Term Dependent process with slope β = -0.5 (d=0.25), and the red green line is the theoretical slope line of the Short-Term Dependent process with slope β = 0.0 (d=0.0). The relationship between slopes of the power spectrum β and the DFA is as follows: $\beta = 2(d + 0.5) - 1$.

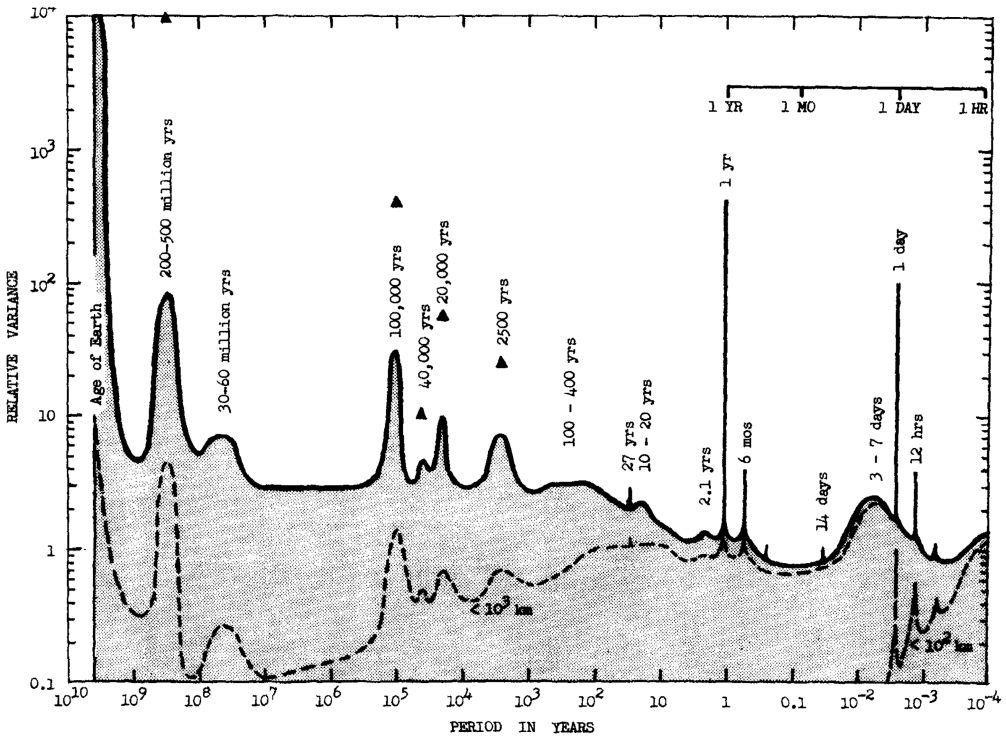


a)



b)

Figure 4. Time series with scaling and non-scaling behavior. a) Probability distribution function of an α-stable distribution with linear axis scaling and b) with logarithmic axis scaling. The case $\alpha = 2$ corresponds to the exponential Gaussian distribution while α values less than 2 correspond to power-laws.



2096 **Figure 5.** Estimates of relative variance of climate over all periods of variation in the climate system.

2097 Source: *Mitchell* [1976]

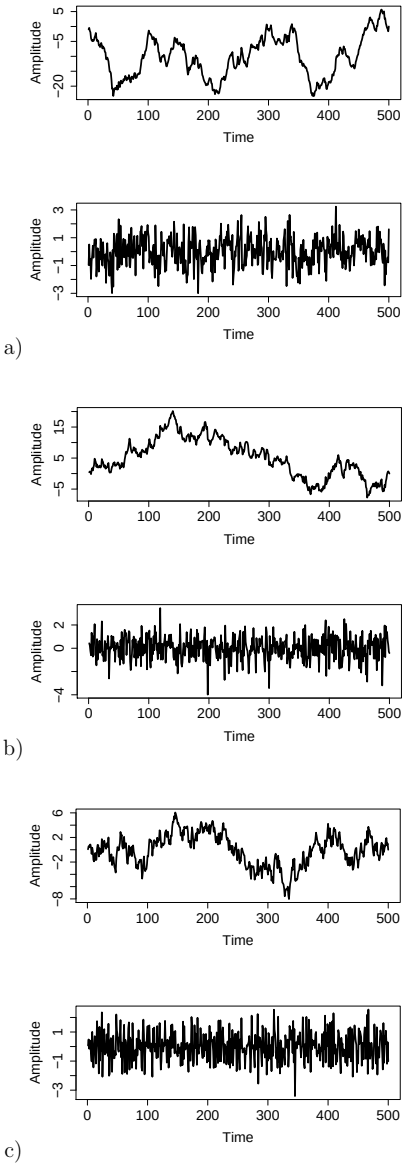


Figure 6. Example time series for fractional Brownian motion (fBm) and the corresponding fractional Gaussian noise (fGn; lower panel) for a) H=0.7 (fGn is persistent), b) H=0.5 (fGn is uncorrelated white noise) and c) H=0.3 (fGn is anti-persistent). The fractional Brownian motion has self-similarity exponent H, and if H is greater than 0.5 is long range dependent, as in the H=0.7 case above.

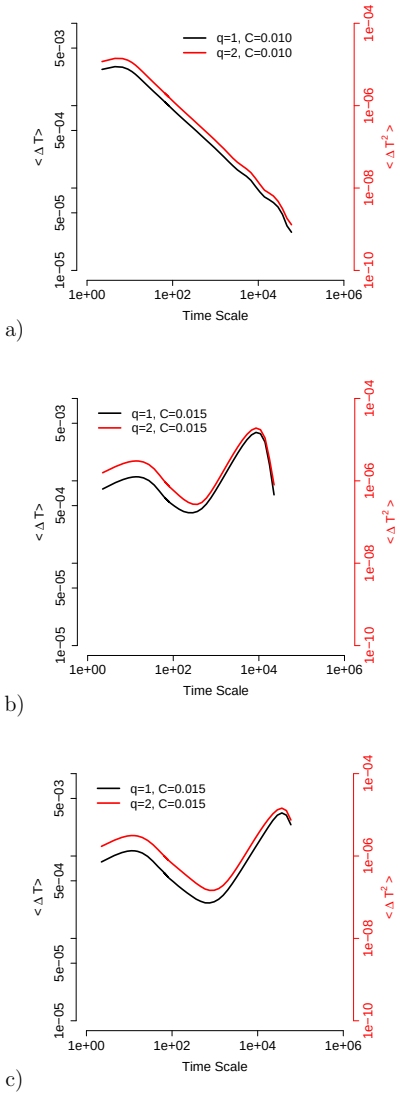


Figure 7. (a) First and second moments, $q = 1, 2$, of the first mode of the streamfunction field as a function of time scale for a wind stress drag coefficient $C = 0.010$ kg m^{-2} s^{-1} and ocean layer depths $h = 164.8$ m; (b) as in (a) for $C = 0.015$ kg m^{-2} s^{-1} and $h = 164.8$ m; (c) as in (b) for $C = 0.015$ kg m^{-2} s^{-1} and $h = 41.2$ m.

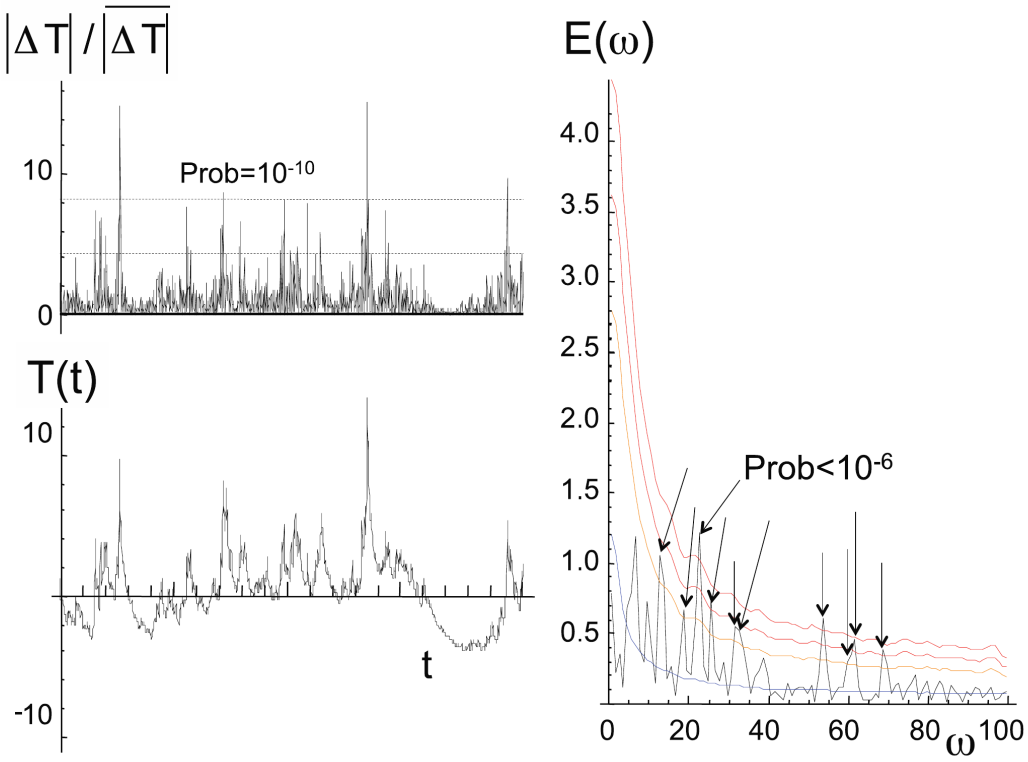
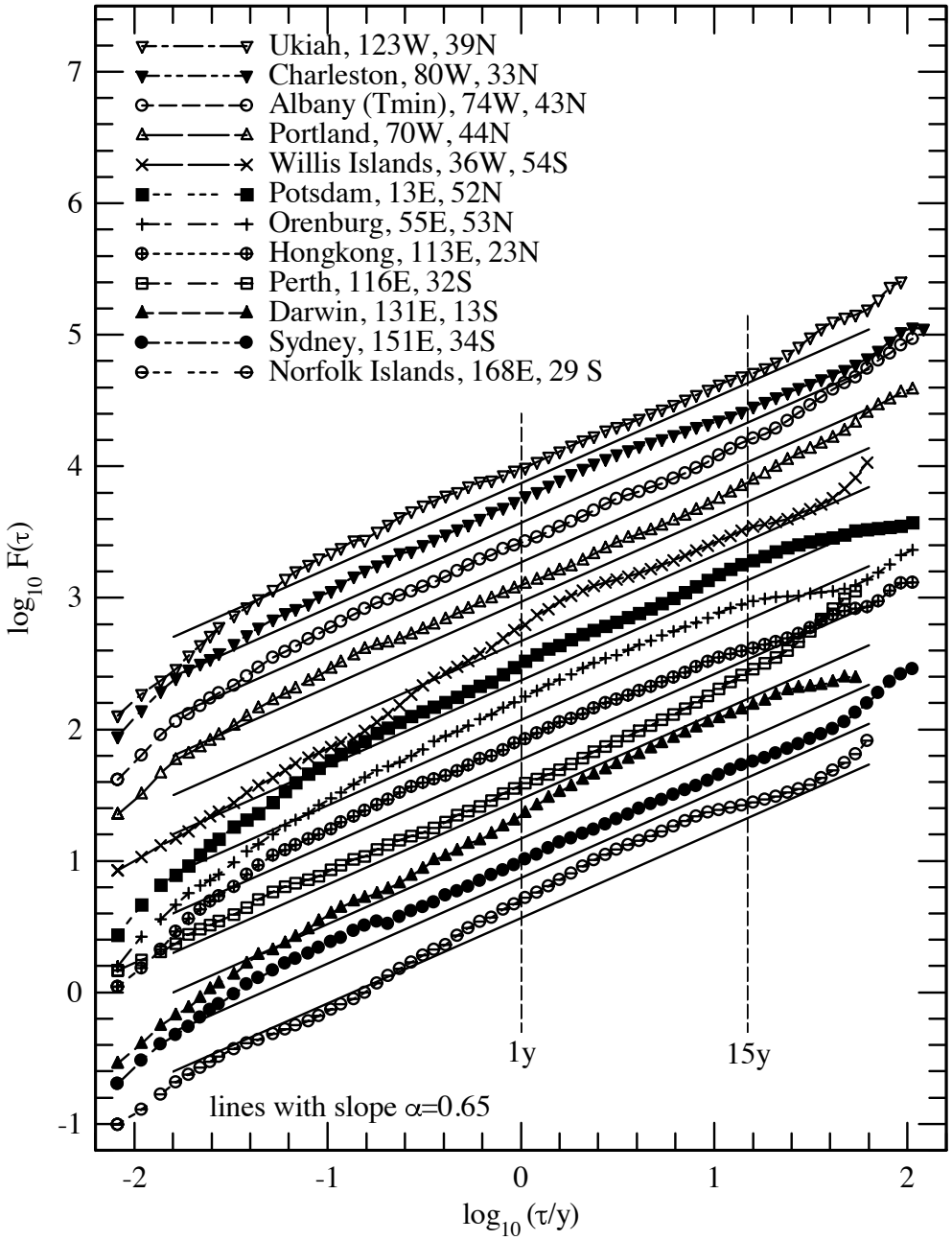
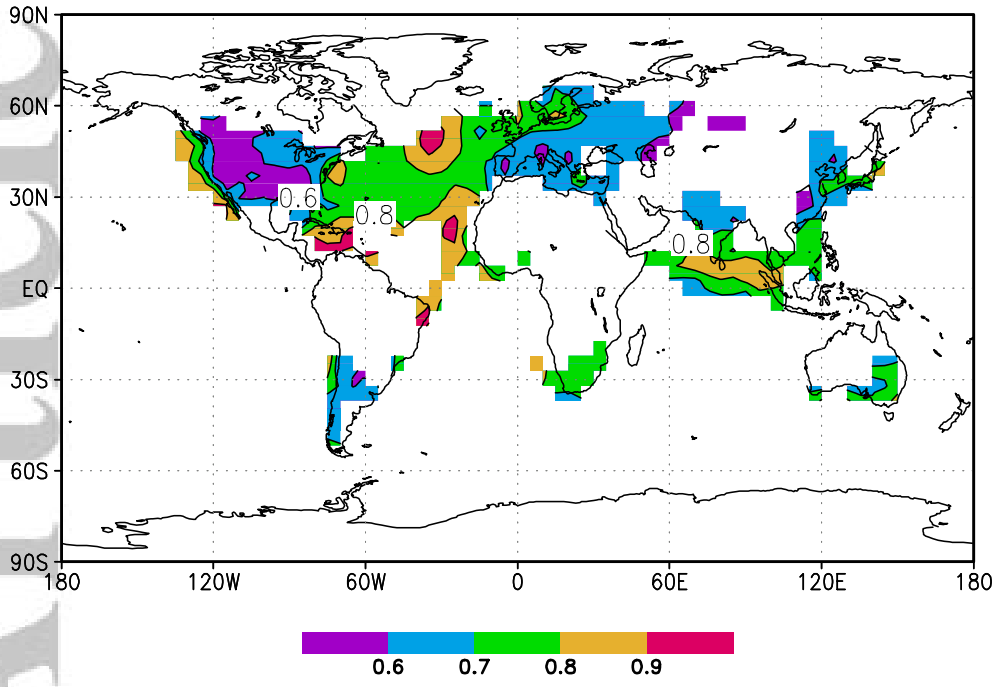


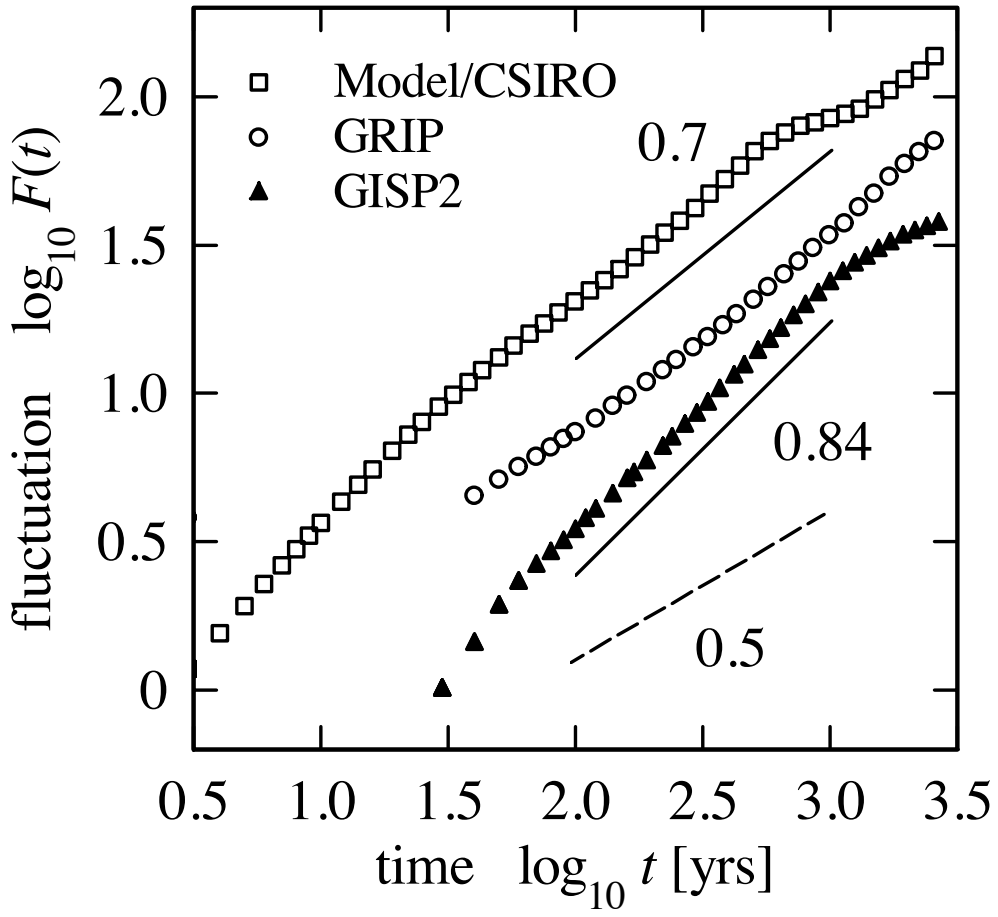
Figure 8. Time series (lower left) and spike plot (upper left) of a 2048 point multi-fractal time series simulation. Its corresponding power spectrum is display on the right (black line). In the spike plots the horizontal dashed lines correspond to the 10^{-5} and the 10^{-10} probability levels. The blue curve in the power spectrum plot is the averaged spectrum over 5000 multi-fractal simulations. Above the blue curve, is an orange 2 standard deviation curve, and (red), 3, 4 standard deviation curves (probabilities 0.1%, 0.003% respectively). The arrows indicate spikes with Gaussian probability p<0.05. See *Lovejoy* [2018] for more details of the used multi-fractal model. Figure is from *Lovejoy* [2018]



2112 **Figure 9.** Detrended Fluctuation Analysis fluctuation functions $F(\tau)$ for daily temperature data at the indi-
2113 cated stations. The lines show the exponents (slopes) $H = 0.65$. Dashed vertical lines at 1-year and 15-years
2114 indicate the time range denoted as decadal here.



2115 **Figure 10.** Fluctuation exponent H in observed sea surface and near surface air temperatures from Had-
2116 CRUT2 data (Climatic Research Unit, University of East Anglia, Norwich) estimated by Detrended Fluctua-
2117 tion Analysis with quadratic trend for the decadal scale (see the slopes in 9).



2118 **Figure 11.** Detrended Fluctuation Analysis fluctuation functions for the Greenland ice cores GRIP, GISP2,
2119 and simulated sea surface temperature (model CSIRO) close to 30W, 65N. The slopes indicate the exponents
2120 $H = 0.5$ (no memory) and $0.7, 0.84$.

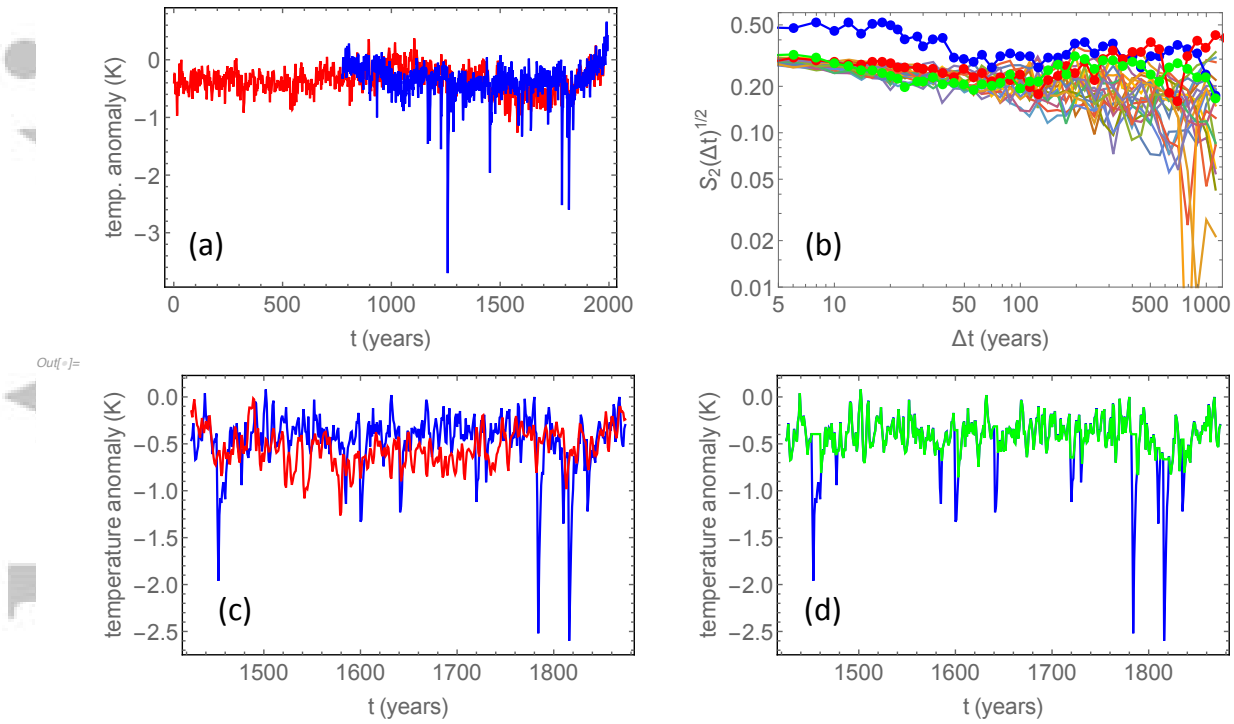


Figure 12. (a) Red: the Moberg Northern hemisphere (NH) temperature reconstruction [*Moberg et al.*, 2005]. Blue: NH surface temperature in a NorESM simulation with historical forcing. (b) (b) Haar fluctuation functions for various signals. Red bullets: the NH reconstruction shown in panel (a). Blue bullets: the NorESM simulation shown in panel (a). Green bullets: The same NorESM simulation with the spikes due to volcanic eruptions removed, as shown by the green curve in panel (d). The full curves are Haar fluctuation functions for 20 realizations of an fGn with H=0.9. (c) A close-up on panel (a) to illustrate that the fast responses du to volcanic eruptions are almost absent in the reconstruction. (d) The green curve illustrates how we have chopped off the volcanic responses from the NorESM signal.

Table 1. Table of scaling exponents. d is used in the statistics community in Autoregressive Fractional Integrated Moving Average models. These models are asymptotically self-similar. H is used in the physical and climatological communities and can be a measure of Long-Range Dependence or self-similarity in systems with Gaussian fluctuations. Here, we use H only as a measure of Long-Range Dependence.

Exponent	Name	Relationship to other exponents
γ	general power-law exponent	
γ_{SS}	self-similarity exponent	
H	Hurst exponent	$H := \frac{\beta+1}{2}$ where H measures Long-Range Dependence
α	stability exponent	
β	power spectrum exponent from a stationary process	$\beta := 2H - 1$ where H measures Long-Range Dependence
d	Long-Range Dependence parameter	$d := H - \frac{1}{2}$ for Gaussian processes
$\tau(q)$	Multi-fractal exponent/Renyi scaling exponent	

A: Fractionally Integrated Processes

Integration, or the inverse procedure differentiation, is a standard procedure in time series analysis to deal with non-stationary time series [e.g. *Box et al.*, 2015]. For instance, a linear trend can be removed from a time series by examining the time series increments instead; higher order trends can consequently be removed by repeating differentiation multiple times and examining the resulting increment time series. Hence, repeated application of differentiation can make every time series stationary. The resulting increment time series can be modeled with an Autoregressive Moving Averaging time series model. In order to represent the original time series the modeled time series would be subsequently cumulatively summed up as many times as differences have been taken before. This results in an Autoregressive Integrated Moving Average model. A fractional Brownian motion (see appendix B:) is an example of a non-stationary time series; its variance goes to infinity with increasing time. Its increments, fractional Gaussian noise (see appendix B:), on the other hand, are stationary.

In standard time series analysis only integer order integration or differentiation is used. However, the integration and differentiation processes can be generalized to also use non-integer integration orders, so-called fractional integration and differentiation [*Samorodnitsky*, 2016]. This allows to mathematically model Long-Range Dependent time series.

The fractional integration parameter d is introduced as follows: let (X_n) be a fractionally differenced process with

$$(1-B)^d X_n = Z_n,\ d \in \mathbb{R},$$

where (Z_n) is white noise with zero-mean and unit variance, and

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} B^j,$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx, \Re(z) > 0$, is the Gamma function. Hence,

$$X_n = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} Z_{n-j}.$$

Observe that

$$\frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \sim \Gamma(d)^{-1} j^{d-1},\ j \to \infty$$

and for $d \in (0, 0.5)$

$$\gamma(f) \sim K_d f^{2d-1},\ f \to \infty, \tag{A.1}$$

where $K_d := \pi^{-1}\Gamma(1-2d)\sin(\pi d)$. Note that the auto-covariance Eq. (A.1) has the same (asymptotic) power-law decay as the auto-covariance Eq. (B.2).

B: Fractional Brownian Motion

Brownian motion is an important stochastic process [e.g. *Embrechts and Maejima*, 2007]. Brownian motion (also called the Wiener process) is the limit of the symmetric Random Walk. While the Random Walk is a discrete-time process, Brownian motion has continuous sampling paths and is a continuous-time process, while at the same time it is nowhere differentiable.

Brownian motion has independent increments. In contrast, fractional Brownian motion has dependent increments. These increments X_n are called fractional Gaussian noise and the strength of the dependence is measured by the parameter H:

$$X_n = B_H(n+1) - B_H(n),\ n = 1, 2, \ldots, \tag{B.1}$$

where H is often called the Hurst exponent and can take values in $(0, 1]$. Fractional Gaussian noise is a discrete-time increment process of fractional Brownian motion. For fractional Gaussian noise even those values that are far apart in time are still serially correlated. Hence, even the distant past affects the current values. If $H = \frac{1}{2}$ then the process is standard Brownian motion, if $H > \frac{1}{2}$ then the increments are positively correlated, while for $H < \frac{1}{2}$ they are negatively correlated and anti-persistent, which is the opposite of Long-Range dependence because the process will wildly fluctuate.

Note that the stationarity of the increments of fractional Brownian motion implies that this is a stationary zero-mean Gaussian process whose autocorrelation function, $\mathrm{acf}(h) := E(X_n X_{n+h})$, satisfies [e.g. *Beran et al.*, 2013]

$$\mathrm{acf}(h) \sim H(2H-1)h^{-2(1-H)},\ h \to \infty, \tag{B.2}$$

provided that $H \in [0, 1]$. If $H \in [0.5, 1]$ then the correlations are not summable, thus, they go to infinity, and we say that X_n exhibits Long-Range Dependence and H measures its intensity. If, on the other hand, $H \in [0, 0.5]$ we say that X_n is anti-persistent.

Fractional Brownian motion is self–similar. By considering probability distributions it can be shown that

$$B_H(at) \overset{d}{=} |a|^H B_H(t) \tag{B.3}$$

The Hurst exponent describes the raggedness of the time series, with a higher H leading to smoother time series. Examples for persistent, white noise (serial uncorrelated time series) and anti-persistent time-series are displayed in Fig. 6. Fractional Brownian motion was introduced by *Kolmogorov* [1940]. More rigorous treatment of fractional Brownian motion can be found in the books by *Embrechts and Maejima* [2007] and *Beran et al.* [2013].

C: Details of Long-Range Dependence parameter estimators

C.1 Time-domain methods

C.1.1 R/S estimator

The R/S method was the first Long-Range Dependence estimator. For a time series $X_1, \ldots, X_N$, the R/S statistic [*Hurst*, 1951; *Beran*, 1994] is given by

$$\frac{R_n}{S_n} = \frac{\max_{0\le i\le n}\left(Y_i - \frac{i}{n}Y_n\right) - \min_{0\le i\le n}\left(Y_i - \frac{i}{n}Y_n\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}Y_n)^2}} =: \frac{I - II}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}Y_n)^2}},$$

where $Y_i = \sum_{j=1}^{i} X_j$. I measures how far the partial sums, Y_i, exceed the straight line they would follow if all observations were equal (to the sample mean). $I - II$ is the difference between the highest and lowest positions of the partial sums with respect to the straight line of uniform growth. For either fractional Gaussian noise or the Autoregressive Fractionally Integrated Moving Average model

$$E(R_n/S_n) \sim K_H \cdot n^H,\ n \to \infty,$$

here K_H is a positive, finite constant which depends on H. $H > 0.5$ for data with Long-Range Dependence. Following *Taqqu et al.* [1995], the methodology for estimating H comprises the following steps: subdivide the time series $X_1, \ldots, X_N$, into K blocks of size $r := N/K$. For each lag n, compute $R_{r_i,n}/S_{r_i,n}$, starting at points $r_i = iN/K + 1$, for $i = 1, 2, \ldots$, such that $r_i \le N - n$. Plot $(\log R_{r_i,n}/S_{r_i,n})$ versus $\log(n)$ by fitting a straight line. The slope of the line gives H. However, this R/S approach does not result in reliable estimates and its use is no longer recommended [*Rea et al.*, 2009; *Franzke et al.*, 2012].

C.1.2 Variance-type estimator

As a more robust alternative, *Taqqu et al.* [1995] proposed the aggregated variance method to estimate H. Variance-type estimators are a popular method to estimate the

Long-Range Dependence parameter. The variance-type estimator of *Taqqu et al.* [1995] takes the form

$$\hat{H} = -\frac{S_m^2}{\log(m)}, \tag{C.1}$$

where

$$S_m^2 = [m/N] \sum_{k=1}^{[N/m]} \left(X_k^{(m)} - [m/N] \sum_{j=1}^{[N/m]} X_j^{(m)} \right)^2$$

with $[\cdot]$ denoting the integer part, and $X_k^{(m)}$ is the aggregated series of order m

$$X_k^{(m)} = \frac{1}{m} \sum_{i=1}^{m} X_{i+(k-1)m}, \; k = 1, 2, \ldots .$$

A major drawback of this variance-type estimator is that its bias is of order no less than $1/\log(N)$ so that only when dealing with very long time series such an estimator can provide reliable point estimates for H. Thus, *Giraitis et al.* [1999] introduced the following refined estimator of (C.1)

$$\hat{H} = -\frac{\sum_{j=m_0}^{m_1} a_j \log(S_j^2)}{\sum_{j=m_0}^{m_1} a_j^2}, \tag{C.2}$$

where

$$a_j := \log(j) - \frac{1}{m_1 - m_0} \sum_{i=m_0}^{m_1} \log(i),$$

for $m_0 < m_1$, such that $m_0 \to \infty$ as $N \to \infty$ and $N/m_1 \to \infty$. *Giraitis et al.* [1999] proved that the estimator in (C.2) is less biased than (C.1). Specifically, this method plots the logarithm of the variance of an aggregated (averaged) process against the logarithm of the aggregation level. A least-squares line is then fitted to the data, the slope of which provides an estimate of H.

C.1.3 Detrended Fluctuation Analysis estimator

The Detrended Fluctuation Analysis (DFA) is a variant of the above method [*Peng et al.*, 1994; *Koscielny-Bunde et al.*, 1998; *Kantelhardt et al.*, 2001; *Lennartz and Bunde*, 2009; *Rybski and Bunde*, 2009; *Lennartz and Bunde*, 2011; *Bunde et al.*, 2014; *Ludescher et al.*, 2015] and estimates the variability of a time series, X_t, on different time scales. First, a profile is computed by $Y(i) = \sum_{t=1}^{i} X_t$. The profile is then split into Ns non-overlapping segments of equal length s and then the local trend is subtracted for each segment v by a polynomial least-squares fit. Linear (DFA1), quadratic (DFA2) or higher-order

polynomials can be used for detrending. In the nth-order Detrended Fluctuation Analysis, trends of order n in the profile, and of order $n-1$ in the original time series, are eliminated. Next, the variance for each of the Ns segments is calculated by averaging over all data points i in the vth segment:

$$F_s^2(v) := < Y_s^2(i) > = \frac{1}{s}\sum_{i=1}^{s} Y_s^2[(v-1)s+i]. \tag{C.3}$$

By computing the average over all segments and taking the square root we obtain the fluctuation function:

$$F(s) = \sqrt{\frac{1}{N_s}\sum_{v=1}^{N_s} F_s^2(v)}. \tag{C.4}$$

For time series with Long-Range Dependence, F(s) will increase with s as a power-law,

$$F(s) \sim s^H \tag{C.5}$$

with the exponent H>0.5. Unlike most algorithms, the DFA algorithm developed by *Løvsletten* [2017] is capable of dealing with missing data. The R package nonlinearTseries provides code for Detrended Fluctuation Analysis. The Detrended Fluctuation Analysis method is biased for $H < 0.5$ [*Franzke et al.*, 2012].

C.1.4 Wavelet-based estimators

The Long-Range Dependence parameter can also be estimated using wavelets. A Wavelet (WL) ψ is a localized wave function with zero average and is normalized to one. A family of Wavelets is generated by scaling the wavelet ψ by a factor s and translating it by u $(\psi_{u,s}(t)) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})$. The Wavelet transform allows one to construct a time-frequency representation of a signal, the Wavelet spectrum. One can then infer the self-similarity parameter from the Wavelet spectrum via ordinary least squares at large Wavelet scales [*Stoev and Taqqu*, 2005; *Abry and Veitch*, 1998]. A widely used wavelet for scaling analysis is the Haar wavelet [*Lovejoy and Schertzer*, 2012b,a; *Lovejoy*, 2014]. The Haar wavelet mother function is given by:

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \le t < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \le t < 1 \\ 0 & \text{if otherwise} \end{cases} \tag{C.6}$$

In the Haar wavelet technique, one usually considers the original time series $X_1, \ldots, X_N$ and divides the time series into N_s segments of length s. For each segment v, one first

determines the mean value $\bar{x}_v$ of the data, then considers the quantity $G_v^2(s) = (\bar{x}_v)^2$ for a zeroth order wavelet (WT0), $G_v^2(s) = (\bar{x}_v - \bar{x}_{v-1})^2$ for first order wavelet (WT1), and $G_v^2(s) = (\bar{x}_v - 2\bar{x}_{v-1} + \bar{x}_{v-2})^2$ for second order wavelet (WT2). By averaging $G_v^2(s)$ over all segments and taking the square root, the wavelet fluctuation function can be obtained as [*Bogachev et al.*, 2017],

$$G_2(s) = \sqrt{\frac{1}{N_s}\sum_{v=1}^{N_s} G_v^2(s)} \tag{C.7}$$

For time series with Long-Range Dependence, the parameter can be estimated according to the relationship

$$G_2(s) = G_2(1)s^{H-1} \tag{C.8}$$

Similar to the orders of detrended fluctuation analysis, the different orders of wavelet methods also indicates trend elimination. For example, in WT2, effects of the linear external trends are eliminated.

C.2 Frequency-domain methods

Spectral methods are also widely used for estimating the Long-Range Dependence parameter.

C.2.1 Geweke-Porter-Hudak estimator

A widely used method is the Geweke-Porter-Hudak (GPH) estimator [*Geweke and Porter-Hudak*, 1983]. Spectral methods find d by estimating the spectral slope. The periodogram, an estimate of the spectral density of a finite-length time series, is given by

$$f(\lambda_j) = \frac{1}{N}|\sum_{t=1}^{N} x_t e^{-i2\pi t\lambda_j}|^2, j = 1, \ldots, [N/2], \tag{C.9}$$

where $\lambda_j = j/N$ is the frequency and the square brackets denote rounding towards zero. A series with Long-Range Dependence has a spectral density proportional to $|\lambda|^{-2d}$ close to the origin. Since $f(\lambda)$ is an estimator of the spectral density, d is estimated by a regression of the logarithm of the periodogram versus the logarithm of the frequency λ. Thus having calculated the spectral density estimate $f(\lambda)$, semi-parametric estimators fit a power law of the form $f(\lambda, b, d) = b|\lambda|^{-d}$, where b is the scaling factor. The R package fracdiff provides code for GPH.

C.2.2 Whittle estimator

The Whittle estimator is based on the periodogram. Specifically, it involves the function

$$G(\theta) := \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda,\theta)} d\lambda,$$

where $I(\cdot)$ represents the periodogram, $f(\cdot,\cdot)$ is the spectral density at frequency λ, and θ denotes the vector of unknown parameters. The Whittle estimator corresponds to the value of θ which minimizes the function $G(\cdot)$. In the case of fractional Gaussian noise or fractional Autoregressive Integrated Moving Average model, $\theta = \{H\}$. The R package longmemo provides code for the Whittle estimator.

D: Hurst Exponent and Long-Range Dependence

Examining water levels of the Nile river, Harold E. Hurst discovered that if the variance is computed for windows of different sizes and then plotted against the window size he obtained a power-law behavior [*Hurst*, 1951, 1957]. This has been named the Hurst phenomenon and the exponent of this power-law is the Hurst exponent.

The Hurst exponent, as we defined it here, is related to Long-Range Dependence [*Talkner and Weber*, 2000]. The tail exponent, which measures the power-law decay of Probability Density Functions, does not affect Long-Range Dependence but does affect the self-similarity exponent [*Franzke et al.*, 2012].

The Long-Range Dependence parameter, d, can be related to H in mono-fractal Gaussian systems as $H = d + \frac{1}{2}$. However, it is typically used with Autoregressive Fractionally Integrated Moving Average models that are only asymptotically self-similar.

E: Estimation of multi-fractality

For some climatic time series, it may be not sufficient to characterize the scaling behavior using only one constant exponent. This is the so-called multi-fractality. To quantify this property, a traditional method is the partition function,

$$Z_q(s) = \sum_{\nu=1}^{Ns} |Y(\nu s) - Y((\nu-1)s)|^q \sim s^{\tau(q)}, \tag{E.1}$$

where $\tau(q)$ is the Renyi scaling exponent and $Y(i) = \sum_{t=1}^{i} x_t$ is the profile of the time series x_t as for Detrended Fluctuation Analysis. When $\tau(q)$ is linear in q, the time series

is considered mono-fractal, otherwise it is multi-fractal. In recent years, the multi-fractal DFA (MF-DFA) has gained increasing popularity [*Kantelhardt et al.*, 2002].

MF-DFA is a generalized version of Detrended Fluctuation Analysis (DFA), as shown below,

$$F_q(s) = [\frac{1}{Ns}\sum_{\nu=1}^{Ns}[F_\nu^2(s)]^{q/2}]^{1/q}. \tag{E.2}$$

For $q = 2$, the mono-fractal Detrended Fluctuation Analysis (DFA) is retrieved. Analogous to Eq. (25), for each q, the generalized fluctuation exponent $h(q)$ can be defined as

$$F_q(s) \sim s^{h(q)} \tag{E.3}$$

Since it is easy to verify that $Z_q(s)$ is related to $F_q(s)$ by $F_q(s) = [(1/N_s)Z_q(s)]^{1/q}$, the Renyi scaling exponent $\tau(q)$ can be connected with $h(q)$ as [*Bogachev et al.*, 2017],

$$h(q) = [\tau(q) + 1]/q. \tag{E.4}$$

Another way to characterize the multi-fractality is the singularity strength k (or Holder exponent) and the singularity spectrum $f(k)$ [*Koscielny-Bunde et al.*, 2006]. Based on a Legendre transform, the singularity spectrum $f(k)$ can be derived as

$$f(k) = qk - \tau(q) \tag{E.5}$$

where k is given by

$$k = \frac{d\tau(q)}{dq}. \tag{E.6}$$

Using Eq. (30), k and $f(k)$ can be related to $h(q)$ as,

$$k = h(q) + q\frac{dh(q)}{dq}, \tag{E.7}$$

and

$$f(k) = q[k - h(q)] + 1. \tag{E.8}$$

Accordingly, the strength of the multi-fractality can be estimated from MF-DFA, by calculating the width of the singularity spectrum (the differences between the maximum and the minimum k.

While MF-DFA is equivalent to the wavelet transform modulus maxima (WTMM) method, it is much easier to implement on a computer [*Muzy et al.*, 1991; *Arneodo et al.*, 2002].

F: Power Spectrum $\mathcal{P}(f)$ and $1/f$ noise

Power spectra are important to understand temporal variability [*Kay and Marple*, 1981]. Power spectra are especially useful for detecting (quasi)-periodic signals like the diurnal and annual cycles which constitute an important aspect of climate variability. However, power spectra can also reveal the background variability of the climate system [*Huybers and Curry*, 2006].

A power spectrum displays the fraction of squared amplitudes at different frequency ranges after Fourier transformation of a time series [*von Storch and Zwiers*, 2003; *Wilks*, 2011]. The most common ways of computing a power spectrum are via the Fourier transform or the maximum entropy method [*von Storch and Zwiers*, 2003].

$1/f$ noise has a power-law form of f^{-1} in which the squared amplitudes increase with decreasing frequencies; hence, longer time scales exhibit a stronger variability. $1/f$ is a generic term which also applies to $1/f^{\beta}$ where the power-law has a different exponent β.