Statistical Approximation of Natural Climate Variability

by

Dmitry I. Vyushin

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Abstract

Statistical Approximation of Natural Climate Variability

Dmitry I. Vyushin

Doctor of Philosophy

Graduate Department of Physics

University of Toronto

2009

One of the main problems in statistical climatology is to construct a parsimonious model of natural climate variability. Such a model serves for instance as a null hypothesis for detection of human induced climate changes and of periodic climate signals. Fitting this model to various climatic time series also helps to infer the origins of underlying temporal variability and to cross validate it between different data sets. We consider the use of a spectral power-law model in this role for the surface temperature, for the free atmospheric air temperature of the troposphere and stratosphere, and for the total ozone. First, we lay down a methodological foundation for our work. We compare two variants of five different power-law fitting methods by means of Monte-Carlo simulations and their application to observed air temperature. Then using the best two methods we fit the power-law model to several observational products and climate model simulations. We make use of specialized atmospheric general circulation model simulations and of the simulations of the Coupled Model Intercomparison Project 3 (CMIP3). The specialized simulations allow us to explain the power-law exponent spatial distribution and to account for discrepancies in scaling behaviour between different observational products. We find that most of the pre-industrial control and 20th century model simulations capture many aspects of the observed horizontal and vertical distribution of the power-law exponents. At the surface, regions with robust power-law exponents — the North Atlantic, the North Pacific, and the Southern Ocean — coincide with regions with strong inter-decadal variability. In the free atmosphere, the large power-law exponents are detected on annual to decadal time scales in the tropical and subtropical troposphere and stratosphere. The spectral steepness in the former is explained by its strong coupling to the surface and in the latter by its sensitivity to volcanic aerosols. However power-law behaviour in the tropics and in the free atmosphere saturates on multi-decadal timescales. We propose a novel diagnostic to evaluate the relative goodness-of-fit of the autoregressive model of the first order (AR1) and the power-law model. The collective behaviour of CMIP3 simulations appears to fall between the two statistical models. Our results suggest that the power-law model should serve as an upper bound and the AR1 model should serve as a lower bound for climate persistence on monthly to decadal time scales. On the applied side we find that the presence of power-law like natural variability increases the uncertainty on the long-term total ozone trend in the Northern Hemisphere high latitudes attributable to anthropogenic chlorine by about a factor of 1.5, and lengthens the expected time to detect ozone recovery by a similar amount.

Dedication

To my wife, Oxana, and my parents, Igor and Eugenia

Acknowledgements

Firstly, I want to thank my supervisor, Paul J. Kushner. I am deeply indebted to Paul Kushner for his guidance and support during the years of my PhD study. Professor Kushner significantly improved my ability to focus on important research questions and to identify the most important features of a phenomenon under study. He was almost always available for scientific discussions and full of stimulating ideas, but also provided me a lot of freedom. I would also like to acknowledge Professor Kushner's career advice.

A part of my thesis work has been done in collaboration with Vitali Fioletov from Environment Canada, Theodore G. Shepherd, and Josh Mayer. I have benefited from Vitali Fioletov's expertise in ozone research and statistical modeling. I truly appreciate the fact that Professor Shepherd introduced me to the Atmospheric Physics Group in the Department of Physics. His sharp critical thinking and a strong common sense were very helpful in my scientific development. Josh Mayer spent two summers as a summer student in Prof. Kushner's group and helped a lot with the analysis of CMIP3 models and coauthoring an R-package containing functions used in the thesis.

I would like to thank my committee members, Dylan B. A. Jones and Francis W. Zwiers, for their support, encouragement, and fruitful advice. I acknowledge the insights of my external reviewer, Professor Carl Wunsch, which lead to an improvement of the thesis. Special thanks go to Michael Sigmond, Chris Fletcher, Slava Kharin, Steven Hardiman, Edwin Gerber, Henning Rust, Alexander Korobov, David Stephenson, Nickolas Watkins, and Wladislav Tchernov for useful feedback and helpful criticism. I am thankful to the Atmospheric Physics former and current graduate students and postdocs, Sorin Codoban, Tiffany Shaw, Lei Wang, Thomas Birner, Constantine Nenkov, Tobias Kerzenmacher, Ellie Farahani, Andreas Jonsson, Mark Parrington, Robert Field, Jane Liu, Annemarie Fraser, and Jeffrey Taylor, for their help and suggestions as well as for creating a friendly and stimulating atmosphere in the group.

I am very much obliged to my wife, Oxana, for her support of my academic endeavours.

I am grateful to my former supervisor, Professor Vadim Strygin, who brought me into science and shared with me the joy of doing it.

I acknowledge the support of the Natural Sciences and Engineering Research Council, Canadian Meteorological and Oceanographic Society, Environment Canada, and the Centre for Global Change Science, which prevented my life during my PhD study from being miserable.

Contents

The list of acronyms

20c3m	CMIP3 simulations of the 20th century
ACF	AutoCorrelation Function
AR1	AutoRegressive model of the first order
ARFIMA	AutoRegressive Fractionally Integrated Moving Average model
ARMA	AutoRegressive Moving Average model
CCMVal	Chemistry-Climate Model Validation Activity
CET	Central England Temperature
CMIP3	The third phase of the Coupled Model Intercomparison Project
CRU	Climatic Research Unit
DFA3	Detrended Fluctuation Analysis of the third order
ECMWF	European Centre for Medium-Range Weather Forecasts
EESC	Equivalent Effective Stratospheric Chlorine
ENSO	El Niño-Southern Oscillation
ERA40	ECMWF reanalysis (09.1957-08.2002)
FAAT	Free Atmosphere Air Temperature
FAR	Fractional AutoRegressive model
GCM	General Circulation Model
GFDL	Geophysical Fluid Dynamics Laboratory
GFDL AM	GFDL Atmospheric Model
GISS	NASA Goddard Institute for Space Studies
GPHE	Geweke-Porter-Hudak Estimator
GSPE	Gaussian SemiParametric Estimator
HadCM	Hadley Centre Climate Model
HDD	Heating Degree Days
IPCC	Intergovernmental Panel on Climate Change

The list of acronyms (continued)

JRA-25	Japanese ReAnalysis (1979 – till present)
LRC	Long-Range Correlations
LTR	linear trend
MTM	multitaper method
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NP	North Pacific
picntrl	pre-industrial control CMIP3 simulations
PC	Principal Component (from Principal Component Analysis)
PCM	Parallel Climate Model
PWLT	PieceWise-Linear Trend
QBO	Quasi-Biennial Oscillation
SAGE	Stratospheric Aerosol and Gas Experiment (satellite instrument)
SAT	Surface Air Temperature
SBUV	Solar Backscatter Ultraviolet Radiometer (satellite instrument)
SLP	Sea Level Pressure
SPARC	Stratospheric Processes And their Role in Climate (project)
SST	Sea Surface Temperature
TIROS	Television Infrared Observation Satellites
TOMS	Total Ozone Mapping Spectrometer (satellite instrument)
TOVS	TIROS Operational Vertical Sounder (satellite instrument)
VTPR	Vertical Temperature Profile Radiometer (satellite instrument)
WMO	World Meteorological Organization

Chapter 1

Introduction

All models are wrong, but some are useful.

George E. P. Box

1.1 Characteristics of natural climate variability

Climate variability on interannual to multi-decadal time scales involves a mix of anthropogenically and naturally generated variability (Wigley and Raper, 1990)[1]. There are two main approaches to model climate variability in modern climate science: physical and statistical. The first one is based on the fundamental laws of physics, chemistry, geology, and biology, such as the laws of hydrodynamics, radiative heat transfer, carbon cycle, etc. These laws are written in a form of partial or ordinary differential equations, supplied with initial and boundary conditions, and solved using various numerical schemes on supercomputers. Roughly speaking, these equations combined with a computer code for their solution are called climate models. The physical approach provides a huge amount of information about climate variability, but it is very complicated, because it involves numerous interacting physical processes and their parameters, some of which are poorly constrained by observations or scientific understanding.

[1]Text that appears in non-black fonts is hyperlinked, either to a cross-reference in the thesis or a URL, in the electronic version of the thesis.

An alternative and complementary approach is to use statistical models for description of climate variability. This approach is phenomenological, i.e. it does not have to be based on the fundamental laws of nature. Statistical models involving relatively few parameters, fitted to observed data or to climate model output, provide a compact ("parsimonious") description of large data arrays. In climate science statistical models are mainly used for assessment, interpretation, and comparison of various observational products and climate model simulations. They are also extensively used for forecasting, but as physically based models are getting more mature they achieve a similar skill to statistical models in this kind of application.

The main focus of this thesis are two statistical models describing natural climate variability. By means of these two models we will attempt to get new insights into climate dynamics and provide new ways for its assessment. The development of such conceptual, but at the same time very practically important, statistical models is a necessary step for construction of a theory of climate variability. During our study of the statistical models of climate variability we will employ multiple simulations of physical climate models as nice tools for testing various hypotheses about the origins of climate variability. Therefore we will demonstrate that both approaches, physical and statistical, are mutually beneficial.

Climate variability can be decomposed into three parts: internal climate variability, naturally forced, and anthropogenically forced variability. The internal climate variability is generated by climatic processes at various time scales, e.g. atmospheric convection and breezes on hourly scales, baroclinic life cycles on daily scales, annular modes on weekly scales, monsoons and midlatitude air-mixed ocean layer interactions on monthly scales, quasi-biennial oscillation (QBO) and El Niño Southern Oscillation (ENSO) on annual scales, thermohaline circulation variability on decadal scales, interactions with biosphere on centennial scales, glacial dynamics on millennial scales, etc. The natural forcings are solar and volcanic forcings. The latter can be considered as a stochastic forcing due to the stochastic nature of the location, size and frequency of climate affecting volcanic eruptions. The effect of the solar forcing on the atmospheric air temperature in the 20th century is weak and highly debated (e.g. Benestad and Schmidt, 2009). Here for simplicity we regard the solar forcing effect on interannual to multidecadal time scales as stochastic, although some of its components might be quasi-deterministic, e.g. the so called 11 year solar cycle. A combination of internal and naturally forced climate variability is called natural climate variability. The anthropogenically forced climate variability is deterministic and is not of primary interest in this thesis.

Natural climate variability can be characterized by the temporal spectral density of either observed/reconstructed climatic time series after filtering anthropogenically induced changes if necessary or of climate model simulations forced by external natural forcings. Very often it is useful to represent the spectral density with a simple approximate statistical model. This kind of model, which is statistically parsimonious in the sense that it uses a small number of parameters, provides input to studies of trend and periodicity detection, climate predictability, extreme value statistics, etc. An advantage of a parsimonious statistical model is that it is relatively easy to compare a few parameters, that this model depends on, across different data sets and, as we do in this thesis, to look at the geographic distribution of such parameters. A disadvantage of a complex statistical model, i.e. a model with many parameters, is that in case some of its parameters turn out to be useless their only effect is to increase the probability of error generation. In addition, if a model becomes too detailed, i.e. the data are overfit, then there might be no ability to generalize it.

A parsimonious statistical model should be distinguished from a parsimonious physical model. The principle of parsimony or Occam's razor principle in physics and in science in general is often a subjective matter that depends on the problem and the user's prior knowledge and way of thinking. For example, it is commonly accepted that linear models are simpler than nonlinear models. Thus given a linear and a nonlinear model, which depend on the same number of parameters and provide equally good explanation for data, the linear model would most typically be chosen. Although, one can imagine an individual who dealt only with nonlinear models in his/her entire life. Because this individual has a different than commonly accepted Occam's razor, he/she would choose the nonlinear model as the most simple theory for given data. In this thesis we will mainly rely on the principle of statistical parsimony, in particular because of its more rigorous formulation. However statistically parsimonious models might be insightful for construction of parsimonious (low-order) physical climate models for explanation and prediction of climate variability and change.

The best known statistically parsimonious approximation for discrete climatic time series is the autoregressive model of the first order (AR1), which was theoretically justified by Hasselmann (1976). It is based on an idea of temporal scale separation between oceanic and atmospheric dynamics and on their linear interaction. In Hasselmann's (1976) model a fast stochastic (weather-noise) atmospheric variability drives slow damped components of the climate system such as the ocean.

Autoregressive models belong to a class of Markov models. In a continuous time framework the AR1 model might be written as a linear stochastic differential equation of the first order. In discrete time it has a very simple form

$$M_t = \phi M_{t-1} + \varepsilon_t, \quad |\phi| < 1, \tag{1.1}$$

where ϕ is a lag-one autocorrelation and ε_t are white noise innovations. The AR1 process has an exponentially decaying autocorrelation function (ACF)

$$C_{AR1}(t) = \phi^{|t|}, \quad |\phi| < 1, \quad -\infty < t < +\infty. \tag{1.2}$$

The shape of the AR1 model spectral density is

$$S_{AR1}(\lambda) = \frac{\sigma_\varepsilon^2}{1 - 2\phi\cos(2\pi\lambda) + \phi^2}, \quad |\lambda| \le 1/2, \tag{1.3}$$

where λ is the frequency, with $\lambda = 1/2$ corresponding to the Nyquist frequency, and σ_ε^2 is proportional to the time series variance (see e.g. Brockwell and Davis, 1998). In typical applications σ_ε^2 and ϕ are estimated from time series or from power spectrum (spectral density estimate) and used to test for the presence of significant periodic or externally forced signals (e.g. Ghil et al., 2002). In other applications (e.g. Bretherton and Battisti, 2000) the model is taken as a simplified physical model to analyze climate variability.

There are multiple ways to generalize the AR1 model, e.g. an ARK model (e.g. Wunsch, 2004), a multivariate AR1 model (e.g. Barsugli and Battisti, 1998; Caballero et al., 2002), a sum of several independent AR1 models (e.g. Granger, 1980; Maraun et al., 2004). Let us consider the most straightforward and therefore the most popular of these generalizations, i.e. the ARK model (see e.g. Brockwell and Davis, 1998; von Storch and Zwiers, 1999). The autoregressive model of the K-th order for monthly mean temperature, M_t, might be written as follows:

$$M_t = \phi_1 M_{t-1} + \phi_2 M_{t-2} + \ldots + \phi_K M_{t-K} + \varepsilon_t, \tag{1.4}$$

where ϕ_k are autoregressive coefficients such that M_t is a stationary process (see e.g. Brockwell and Davis, 1998; von Storch and Zwiers, 1999) and ε_t are white noise innovations with variance σ_ε^2. The spectral density of the ARK process is

$$S_{ARK}(\lambda) = \frac{\sigma_\varepsilon^2}{\left|1 - \sum_{k=1}^{K} \phi_k \exp(-2\pi i k \lambda)\right|^2}, \quad |\lambda| \leq 1/2. \tag{1.5}$$

It is easy to see that $S_{ARK}(\lambda) \to const$ as $\lambda \to 0$.

ARK models are used much more rarely in climate science than the AR1 model. We will consider fitting ARK models to a particular climatic time series a few pages below. Another generalization of the AR1 model, a sum of several multiscale AR1 models, is considered in Section 6.6.

The spectral density $S_{AR1}(\lambda)$ scales as λ^{-2} at high-frequencies and then, as well as the spectral density of an ARK model, saturates to a constant at low-frequencies. But this behaviour is not always observed. Recent research has pointed out potential limitations of the AR1 model (e.g. Hall and Manabe, 1997; Schneider and Fan, 2007). Also many studies in the past two decades (e.g. Bloomfield, 1992; Pelletier, 1997; Tsonis et al., 1999; Eichner et al., 2003; Fraedrich and Blender, 2003; Vyushin et al., 2004; Huybers and Curry, 2006) have reported that the power spectra of various climatic time series do not seem to saturate but keep growing at low-frequencies, although with slope shallower than -2.

A recently well developed mathematical theory of long-range correlated (LRC) processes (also known as long-range dependent or long-memory processes) provides a nice framework for modelling or at least constraining the temporal variability of such time series. This theory describes stochastic processes for which the ACF decays algebraically:

$$C_{PL}(t) = a|t|^{2H-2}, \quad 1/2 < H < 1, \quad |t| \geq t_h, \tag{1.6}$$

where H is the "Hurst exponent", named after the British hydrologist, who first observed this phenomenon while studying the Nile river (Hurst, 1951). It can be shown that the spectral density of such processes increases by the power-law with decreasing frequency (see e.g. Taqqu, 2002)

$$S_{PL}(\lambda) = b|\lambda|^{1-2H}, \quad 0 < |\lambda| \leq \lambda_{\text{high}} \leq 1/2, \tag{1.7}$$

where b represents the overall spectral power and λ_{high} is a high-frequency cutoff. $H = 1/2$ corresponds to a white-noise or short-memory spectrum and $H = 1$ corresponds to a "$1/f$" noise spectrum. Power-law variability represents temporal scaling behaviour without a characteristic timescale. In this formulation for $\lambda_{\text{high}} < 1/2$ the power-law model is less parsimonious than the AR1 model, because it has one extra parameter, λ_{high}. Statistically speaking, the power-law model is a semi-parametric model, because it describes time series only partially (in our case its low-frequency variability), whereas the AR1 model is formally a full parametric model. Although implicitly the AR1 model is also semi-parametric, because, as we will show below, it depends on time series aggregation time scale, i.e. whether monthly, annual, decadal or whatever means are considered, which is somewhat similar to the high-frequency cutoff parameter. However in some applications, such as trend detection (see e.g. Smith, 1993), only the low-frequency variability is important, which motivates the use of a semi-parametric model, because for instance a full parametric model might be seriously affected by the high-frequency variability. It has been shown by Granger (1980); Caballero et al. (2002); Maraun et al. (2004) that various generalizations of the AR1 model can approximate the power-law model for frequency ranges which exclude the zero frequency.

One important point, which some researchers miss, is that long-range correlations necessarily lead to temporal power-law behaviour, but not the other way around. Temporal power-law behaviour could be caused by either the Joseph effect, which comes from the Old Testament story about Joseph, where Egypt would experience seven years of feast followed by seven years of famine, and represents long-range correlations, or the Noah effect, which refers to another Old Testament story when the God caused it to rain upon the Earth for forty days and forty nights, and represents fat (power-law) tails of the underlying probability density function (Mandelbrot and Wallis, 1968). A simple test to distinguish between these two effects is to compare the Hurst exponent estimates for the original and for a shuffled version of the time series. Because shuffling destroys serial correlations and preserves the distribution, it removes the Joseph effect, but leaves the Noah effect. It has been shown in several studies (e.g. von Storch and Zwiers, 1999) that surface temperature and total ozone (with monthly and coarser temporal aggregation), which are two of the three variables we focus on here, are approximately normally distributed in time. Thus, their potential power-law behaviour should be attributed to long-range correlations — the Joseph effect. Consistently, we have found that when we randomize in time the time series of surface temperature, free atmosphere air temperature, and total ozone, the resulting estimates of H are not distinguishable from 1/2, to within the confidence of our H estimation techniques (not shown). This points to the absence of the Noah effect. The next question is if an observed Joseph effect is physical or an artefact of data inhomogeneities, which, as we will discuss, are known to lead to power-law behaviour (see Berton (2004); Rust et al. (2008) and Chapter 5 of the thesis).

Power-law behaviour has been reported in globally and hemispherically averaged surface air temperature (Bloomfield, 1992; Gil-Alana, 2005), station surface air temperature (Pelletier, 1997), geopotential height at 500hPa (Tsonis et al., 1999), temperature paleo climate proxies (Pelletier, 1997; Huybers and Curry, 2006) and many other studies (see Table A.1 for a more complete historical list of relevant studies). However only a few of these studies have performed quantitative tests to determine if the power-law model is superior to the AR1 model. Those who have (Stephenson et al., 2000; Percival et al., 2001) find that both models demonstrate similar scores for the goodness-of-fit tests employed. Both Stephenson et al. (2000) and Percival et al. (2001) conclude that century long time series are not long enough to clearly demonstrate the superiority of one model over the other. In contrast, in most of the empirical power-law model studies in all fields including climate science the only criteria used to test the validity of the power-law model is the fact that the estimated Hurst exponent is significantly greater than 1/2 (e.g. Willinger et al., 1999). We will start with this somewhat naive assumption in Chapter 3 and then progress to more sophisticated approaches to distinguish between time series models in Chapter 6.

With the discussion of parsimony in mind, we base much of our analysis in this thesis on the AR1 and the power-law model for the following reasons: (a) they are the most parsimonious red-noise models (a process with a power-law spectrum sometimes is also called a pink noise); (b) they give a lower and an upper bound on climate persistence (see Chapter 6); (c) at the moment the AR1 is the most commonly used climate noise model, for instance it is extensively used in the two most influential recent climate assessments (Intergovernmental Panel on Climate Change, 2007; World Meteorological Organization, 2007); (d) the power-law model is probably the second most cited statistical model of natural climate variability (see Table A.1). Unlike some of the literature cited in Table A.1, nowhere in the thesis do we claim that power-law behaviour is universal on all time scales. Instead, we use $S_{PL}(\lambda)$ to provide a sense of how quickly power builds towards lower frequencies on annual to multidecadal scales. Regions where $\hat{H} = 0.5$ (the flat spectrum limit) might be well described by either model, while regions where $\hat{H}$ is closer to 1 (the $1/f$ limit) are candidates for true power-law behaviour.

In this thesis we will mainly use temperature variability to represent climate variability. We employ temperature because it is probably the best observationally constrained and physically understood climate variable. To set the stage for our analysis of the modern temperature record, we would like to start with Fig. 1.1, which is Fig. 2 in Huybers and Curry (2006). It shows a compilation of power spectra of various temperature records, predominately derived from paleo-proxies. The time scales extend from two days to several hundred thousand years. The extratropical records have larger variance than the tropical ones. It can be argued that the power spectra shown in Fig. 1.1 have at least three scaling (power-law like) regimes: subannual, from one year to about hundred years, and longer than hundred years. In this thesis we mainly deal with the second regime.

Huybers and Curry (2006) suggested that the scaling regime on scales longer than hundred years could be related to the Milankovitch cycles. However one can imagine a simpler world without the Milankovitch cycles. What would the power spectrum of temperature be in such a world? Would it saturate at low frequencies or would it maintain the slope (in log-log coordinates) inferred for the annual-to-centennial band? At this stage of the climate science development we can not definitely answer these questions. It is hard to separate the effect of the Milankovitch cycles from the internal climate variability in existing paleo records (e.g. Wunsch, 2004) and there is not yet enough trust in climate model paleo simulations. Therefore we should consider both of the above mentioned possibilities for the unobserved low frequency part of the spectrum, which are the limits stationary time series might tend to.

Answers to the above questions are essential for trend detection. The assumption that climatic power spectra saturate after a certain low frequency is the current standard practice (e.g. Intergovernmental Panel on Climate Change, 2007; World Meteorological Organization, 2007). However many studies, which reported power-law like increase of temperature power spectra, in particular Fig. 1.1, shed doubt on this assumption. Alternatively, if one wants to stay on a conservative side, i.e. to assume a strong natural climate variability, the assumption that temperature power spectra increase by a power-law as frequency tends to zero is not unreasonable. In this case the power-law fit obtained using observed variability on interannual to multidecadal time scales can be extrapolated to zero frequency by keeping the Hurst exponent constant. Apart from trend detection a power-law fit to a specific frequency range is useful as a parsimonious tool for intercomparison of temporal variability on specific time scales between different observational products and climate model simulations (see Chapters 5-6).

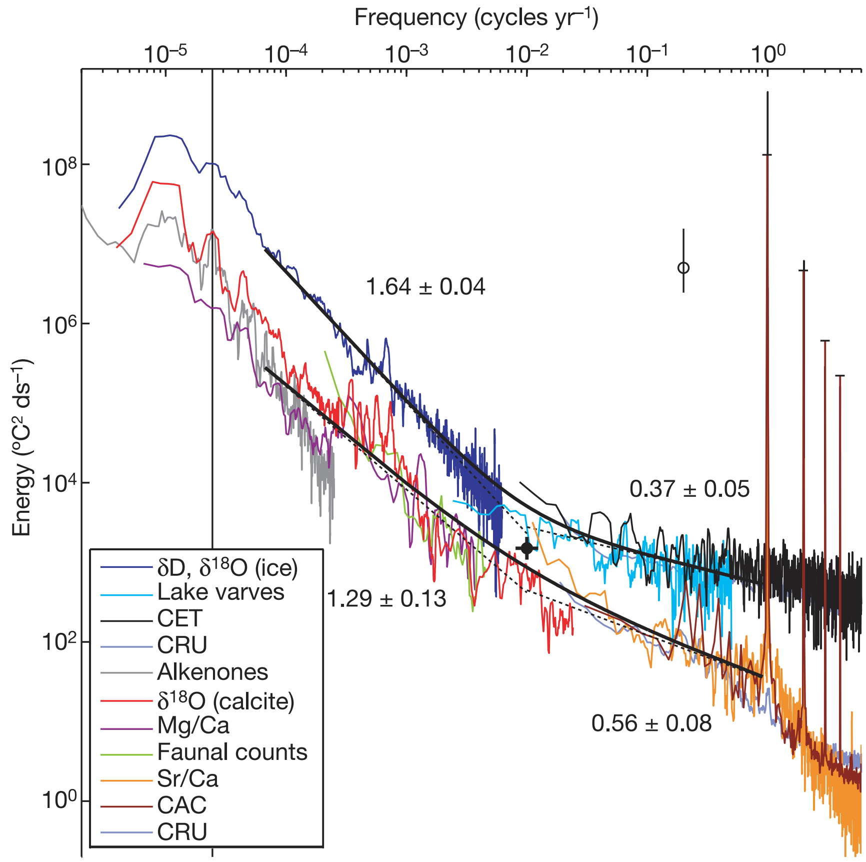

Figure 1.1: (Adopted from Huybers and Curry (2006)) A combination of spectral estimates obtained using instrumental and proxy records of surface temperature variability, and insolation at 65°N. The more-energetic spectral estimate is from high-latitude continental records and the less-energetic estimate from tropical sea surface temperatures. Highlatitude spectra are estimated from Byrd, Taylor and GISP2 ice-core δ^{18}O; Vostok and Dome C ice-core δD; Donard lake varve thickness; Central England Temperature; and a Climate Research Unit's (CRU) instrumental compilation. From low latitudes ODP846 marine sediment-core alkenones; W167-79, OCE205-103, EW9209-1, ODP677 and ODP927 calcite δ^{18}O; PL07-39 and TR163-19 Mg/Ca; ODP658 foram assemblages; Rarotonga coral Sr/Ca; and the Climate Analysis Center and CRU instrumental compilations are used. Temperature spectral estimates from records of the same data type are averaged together. Power-law estimates between 1.1-100 and 100-15,000 year periods are listed along with standard errors and indicated by the dashed lines. The sum of the power-laws fitted to the long- and short-period continuum are indicated by the black curve. The vertical line-segment indicates the approximate 95% confidence interval, where the circle indicates the background level. The mark at 1/(100 years) indicates the region mid-way between the annual and Milankovitch periods.

One of the statistical characteristics of a time series is its decorrelation time scale, which might be defined as follows

$$\tau_D = \int_{-\infty}^{\infty} C(t)dt, \quad (1.8)$$

where $C(t)$ is a process autocorrelation function. Eq. 1.8 is a continuous version of Eq. 17.5 in von Storch and Zwiers (1999). Stochastic processes can be classified as short-memory and long-memory processes. A short-memory process, for instance an ARK process, has a summable ACF and therefore a finite decorrelation time scale. In contrast, the integral of a long-memory process ACF diverges, e.g. in the case $C(t) \sim t^{2H-2}$, $1/2 < H < 1$ for large t, and therefore its decorrelation time scale is undefined. Thus, in the case an ARK model is fitted to a particular time series its decorrelation time scale can be estimated using the ARK model autocorrelation function, i.e.

$$\hat{\tau}_D = \int_{-\infty}^{\infty} C_{ARK}(t)dt. \quad (1.9)$$

This approach for decorrelation time scale estimation is suggested, for instance, in Bretherton et al. (1999); von Storch and Zwiers (1999). The ARK model decorrelation time scale is unique, i.e. if the model was fitted to a time series of *monthly* means and the decorrelation time scale is estimated using the model autocorrelation function to be equal, for example, to 24 months, then the decorrelation time scale for the *annual* means of this time series should be equal to 2 years according to the ARK model. We will apply the concept of decorrelation time scale to a particular climatic time series below.

Let us illustrate an application of the ARK model together with the two limiting cases, the AR1 and the power-law, by fitting the three models to the monthly mean Central England Temperature (CET) anomalies time series (1659-1958). Here we consider only the first 300 years of the CET record in order to avoid the effect of anthropogenic components, because our focus is on natural climate variability.

In Fig. 1.2a we plot a multitaper spectrum estimator of the CET monthly mean anomalies (black curve) together with the three fits to the spectrum. (This figure and much of the analysis

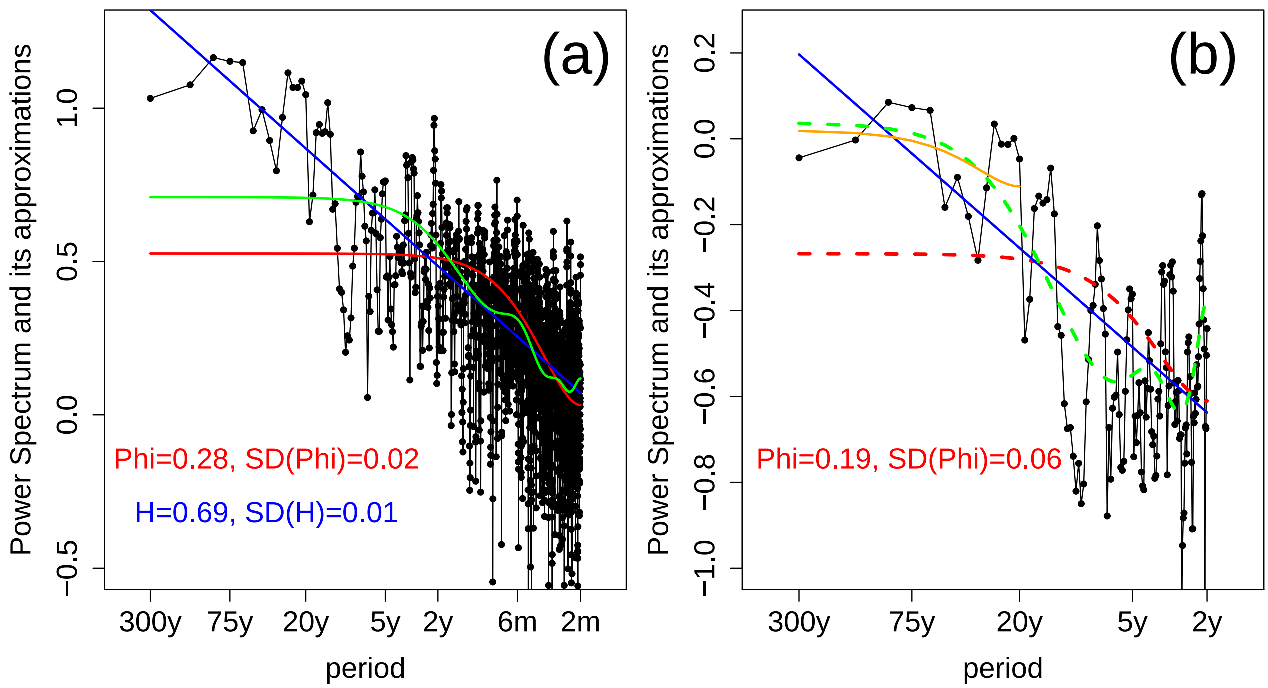

Figure 1.2: Central England Temperature (CET, 1659-1958) (a) monthly means and (b) annual means power spectrum estimators and their approximations. The black curve is a 5 sine tapers multitaper power spectrum estimator (Section 2.2.1). The red solid (dashed) curve is a spectral density of the AR1 model fitted, using a maximum likelihood algorithm, to the CET monthly (annual) means time series. The green solid (dashed) curve is a spectral density of the best fitted, according to the Akaike information criteria, autoregressive model (AR6 for monthly and AR4 for annual means) to the CET monthly (annual) means time series. The power-law fit, estimated using the Geweke-Porter-Hudak Estimator (see Section 2.2.1) applied to the monthly means, is shown by the blue line. The spectral density of the AR1 model fitted to the decadal means is shown by the orange curve in panel (b). The blue curve is the same in both plots. Also included are numerical estimates for the AR1 parameter ϕ and its standard deviation $\mathrm{sd}(\phi)$ (see Section 9.8 of Percival and Walden, 1993) and the Hurst exponent estimate $\hat{H}$ and its standard deviation $\mathrm{sd}(\hat{\mathrm{H}})$ (see McCoy et al., 1998). The spectral and power-law estimators are described in Section 2.2.

in this thesis have been produced using an open-source package, PowerSpectrum, that I have developed using the R statistical language during my Ph.D. study with the help of undergraduate student Josh Mayer. A manual for this package is included in the thesis as Appendix B.) The methods employed in Fig. 1.2 are specified in the caption and the power-law fit and multitaper methods will be described in Chapter 2. Fig. 1.2a demonstrates that the AR1 spectral density overestimates the CET power spectrum at high-frequencies and underestimates it at low-frequencies, whereas the power-law, which also depends only on two parameters as does the AR1, does a much better job.

According to the Akaike information criteria the best fit to the CET monthly mean anomalies among autoregressive models is given by the AR6, which spectral density is shown by the green solid curve in Fig. 1.2a. It better approximates the CET spectrum than the AR1, but it still underestimates it on time scales longer than 20 years. The decorrelation time scale for the CET monthly mean anomalies estimated according to the definition given by the Eq. 1.9 is equal to 3 months (the estimated AR6 autocorrelation function is nonnegative everywhere).

Let us now see what happens to the three models during the transition from the monthly to annual means. The power spectrum of the CET annual means is plotted in Fig. 1.2b by the solid black curve. According to the AR1 model fitted the CET monthly mean anomalies the year-to-year autocorrelation should be equal to 0.026 ± 0.003 (see Section 6.3.1 for details). Instead, the estimated year-to-year autocorrelation is 0.19 ± 0.06, i.e. about seven times larger. To obtain a reasonable fit for the annual means using the AR1 model one has to fit it again. The spectral density of the newly fitted AR1 model is shown in Fig. 1.2b by the dashed red curve. (We plot the spectral densities of the AR1 and the best fit ARK model for the CET annual means by the dashed curves to underline the fact that these fits are obtained by direct fitting to the annual means). As for the monthly means the AR1 spectral density overestimates a high-frequency and underestimates a low frequency part of the spectrum.

Akaike information criteria chooses the AR4 as the best fit autoregressive model for the CET annual means. Its spectral density is shown by the dashed green curve in Fig. 1.2b. It is probably the best fit for the CET annual means power spectrum among the three models. The decorrelation time scale estimated using the AR4 model fitted to the CET annual means is equal to 3 years (the estimated AR4 autocorrelation function is also nonnegative everywhere) in contrast to the 3 months estimated using the monthly means. This fact hints that presumably different physics work on monthly and annual time scales over Central England and that the ARK model cannot capture the dependence of the decorrelation time scale on the temporal aggregation scale (the AR6 model fitted to the monthly means predicted that the annual means should be uncorrelated).

The spectral density of the power-law model, shown by the blue curve in Fig. 1.2b, is obtained just by truncating time scales shorter than 2 years from the corresponding spectral density shown in Fig. 1.2a. Thus an advantage of this model is that in contrast to the AR1 and the best fit autoregressive model it does not have to be refitted each time the aggregation time scale is increased and therefore it better captures the overall power spectrum shape.

Bloomfield and Nychka (1992) compared the estimated linear trend confidence intervals based on white noise, AR1, AR2, AR8, power-law, and two versions of Wigley and Raper (1990) analytically solvable energy balance model for globally averaged annual mean surface air temperature anomalies time series. They found that the estimates based on the AR8, power-law model, and a multibox energy balance model are close to each other and about four times greater than the white noise based confidence interval and about 70% greater than the AR1 and AR2 estimates. However the estimated trend was statistically significant relatively to any of the above mentioned confidence intervals. We will develop the ideas of Bloomfield and Nychka (1992) and apply them to a problem of ozone recovery detection in Chapter 3. Estimates for a globally averaged surface air temperature will be updated and compared to those reported in Intergovernmental Panel on Climate Change (2007) in Section 6.5.

We have also fitted the AR1 model to the CET decadal means. Its spectral density is shown by the orange curve in Fig. 1.2b. It closely follows the spectral density of the AR4 model fitted to the annual means. Therefore for the CET record the AR1 provides a good fit to the spectrum only for the decadal means and it could be used, under the assumption of the CET spectrum saturation near zero frequency, as a noise model for detection of a trend in the decadal means.

Given this background, we will make the working assumption through much of the thesis that the atmospheric general circulation might be well characterized by power-law behaviour on interannual and longer time scales. We will revisit this assumption in Chapter 6, in which we will return to the question of the goodness of fit of each statistical model for observed and simulated climate variability.

Unlike the Hasselmann model, the power-law model, which indicates temporal scaling behaviour rather than dependence on any particular timescale, has no simple established physical interpretation. The possible origins of the power-law spectral behaviour (also called $1/f^\beta$ noise), which might be relevant to climate on annual to centennial time scales, are: aggregation of multiple scales (Granger, 1980; Caballero et al., 2002), in particular in self-organized criticality type models (Rios and Zhang, 1999; Maslov et al., 1999), stochastically forced diffusion equations (Pelletier, 2002; Fraedrich et al., 2004; Dommenget and Latif, 2008), nonlinear stochastic differential equations (Naidenov and Kozhevnikova, 2000; Kaulakys and Alaburda, 2009), a sum of slowly decaying intermittent shocks (Cox, 1984; Parke, 1999; Mandelbrot, 2003), point processes (Davidsen and Schuster, 2002; Kaulakys et al., 2005), chaotic Hamiltonian dynamics (Geisel et al., 1987; Zaslavsky, 2002), intermittent nonlinear maps (Barenco and Arrowsmith, 2004; Miyaguchi and Aizawa, 2007), etc. By the aggregation of multiple scales we mean that various physical processes, which have a wide range of characteristic time scales and which comprise an internal climate variability, can generate a power-law like spectrum for a wide range of frequencies. We have demonstrated an example of such mechanism above for the CET record.

One way to introduce multiple time scales into the climate system is through a vertical diffusion of the ocean temperature. Dommenget and Latif (2008) coupled a comprehensive atmospheric GCM to a simple ocean model mainly represented by the vertical diffusion equation. Thus their global climate model (ECHAM5-OZ) can be roughly approximated by the following equation

$$c\frac{dT}{dt} = -\gamma_{surf}T + k_z\Delta_z T + \varepsilon_{surf}, \tag{1.10}$$

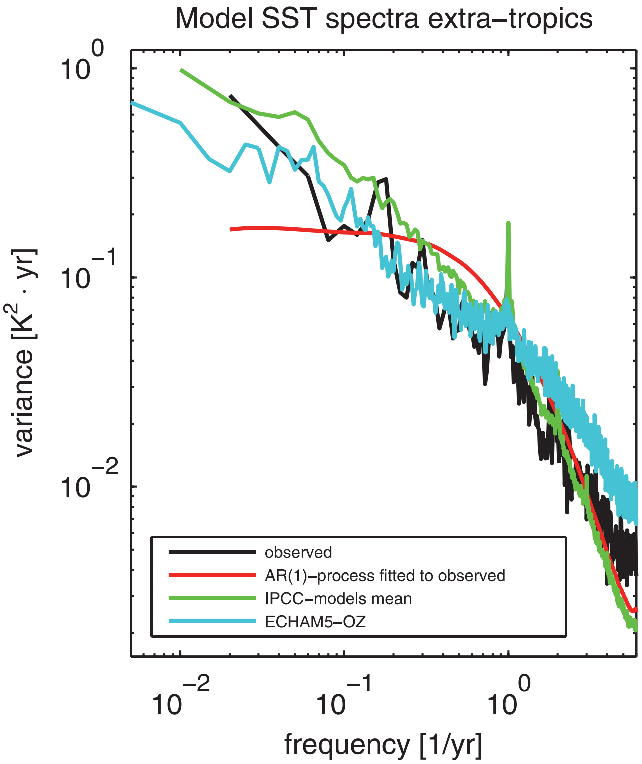

Figure 1.3: (Courtesy of Dietmar Dommenget) The mean spectrum of observed (black curve) and simulated (green and cyan curves) midlatitudinal SSTs. The spectrum is averaged over all grid points in the North Pacific and North Atlantic Oceans between 30°N and 55°N. The red curve shows the spectral density of a fitted to the observations AR1 process.

where T is an upper ocean layer temperature, c is the heat capacity of the ocean layer, γ_{surf} is the damping coefficient, k_z is the exponentially decreasing with depth diffusion coefficient, and ε_{surf} is the atmospheric white noise forcing. The spectrum of a midlatitude sea surface temperature (SST) from an ECHAM5-OZ 800 year long simulation is shown in Fig. 1.3 together with the spectrum of observations and its AR1 fit and the spectrum the IPCC models mean. (We will discuss the analysis of the IPCC simulations in Chapter 6.) All the spectra but the AR1 grow with the decreasing frequency at all time scales. In contrast, the AR1 spectrum saturates to a constant after several years, which is consistent with the behaviour of mixed layer ocean models coupled to an atmosphere, but inconsistent with the behaviour of dynamical ocean models

(Dommenget and Latif, 2002).

While the previous discussion briefly touches on some of the dynamical factors that might underlie power-law like behaviour in climate, the purpose of this thesis is not to develop a detailed dynamical theory of such behaviour. Instead, our main purpose is to test the robustness of spectral slope estimation techniques and to use the most robust techniques to estimate spectral slopes in observations and simulations.

Here we provide a brief overview of the thesis structure. The following two sections of the Introduction provide a literature and applications overview. In Chapter 2 we lay down theoretical and methodological foundations for our work. In particular we compare two variants of five different Hurst exponent estimators by means of Monte-Carlo simulations. Chapter 2 also describes analysis related to trend detection. Chapter 3 deals with understanding and quantification of low-frequency variability in total ozone. It describes a multilinear regression model for the total ozone variability and measures the spectral steepness of the model residuals using Hurst exponent estimates obtained by two methods. Then these estimates are employed to calculate the confidence intervals for the observed trend and the number of years required to detect this trend, which represents an important application of the power-law analysis. The Hurst exponent based confidence intervals are compared to the standard for climate literature AR1 based confidence intervals and shown to be more conservative (larger). In Chapter 4, which begins the main focus of the thesis, we switch from total ozone to reanalysis free atmosphere air temperature (FAAT) and compare the five Hurst exponent estimators described in Chapter 2. We find that the estimators agree provided equal frequency ranges are chosen and known high-frequency climate signals, such as the quasi-biennial oscillation, are filtered out. In Chapter 4 we also compare the Hurst exponents for the zonally averaged air temperature with the zonally averaged Hurst exponents estimated for each grid point time series. Chapter 5 focuses on physical mechanisms giving rise to the FAAT spectral power buildup at low-frequencies. It compares the results for reanalyses with several specialized simulations of an atmospheric GCM and shows that the high values of $\hat{H}$ on annual to decadal time scales are caused by atmosphere-ocean interaction in the tropical troposphere and by volcanic aerosols in the tropical and subtropical stratosphere. The influence of data inhomogeneities, for instance in the tropical upper troposphere and the southern midlatitudes, on FAAT spectrum steepness and the benefits of power-law analysis for their detection and for general low-frequency variability cross-validation are also presented in Chapter 5. Surface air temperature derived from observational products and 17 climate models from the Coupled Model Intercomparison Project 3 archive is analysed in Chapter 6. We compare the Hurst exponent estimates for the observed and simulated temperature for various climate scenarios, temporal scales, and geographical regions. As mentioned above, in Chapter 6 we also evaluate the relative goodness-of-fit of the AR1 and power-law models. The results demonstrate that the real world seems to fall between the AR1 and power-law statistical models. Chapter 7 summarizes and discusses main findings and describes future work. Appendix A lists previous studies related to the power-analysis of temporal climate variability. Appendix B is a manual of the PowerSpectrum package.

Most of the results described in my thesis have been already published. Thus the results of Chapter 3 have been published in (Vyushin et al., 2007), of Chapter 4 in (Vyushin and Kushner, 2009), and of Chapter 5 in (Vyushin et al., 2009). Appendix B of Chapter 3 has been submitted for publication in the *Journal of Geophysical Research: Atmospheres*. Chapter 6 represents a manuscript in preparation.

1.2 Literature Review

In the past half a century there were more than a thousand papers in mathematics, physics, statistics, Earth and life sciences, social sciences, engineering, etc. dealing with the phenomenon of temporal power-law behaviour. The website "A Bibliography on 1/f Noise" attempts to collect all of them. In this Section, we review the statistical climatology literature on this topic, which involves a range of methods that often provide inconsistent results. For reference, we provide in Table A.1 a list of several studies in which temporal power-law behaviour has been quantified. The columns of the table specify analysed variables, estimation methods, the time scale range for which the Hurst exponent was estimated, a range of estimated H, and a reference.

It can be noticed from Table A.1 that the analysis originally started from studies of individual time series and progressed to studies of hundreds of station time series and of gridded data sets of individual observational products and climate models and then to intercomparison of multiple observational products with climate model ensembles. However only a few studies used more than one estimation method and varied frequency ranges. Detrended Fluctuation Analysis (DFA) seems to be the most popular estimation method, especially in the past 15 years. The majority of the papers is devoted to surface air temperature (SAT) followed by precipitation, humidity and sea level pressure. Connections with the previous studies will be made throughout the thesis.

The first three articles (Bloomfield, 1992; Bloomfield and Nychka, 1992; Smith, 1993), which studied temporal power-law spectral behaviour in SAT and its impact on trend detection, did not attract much attention in the climate community. They found that, although a confidence interval of a linear trend in globally averaged SAT is wider under a power-law assumption for the residuals, the observed 20th century trend is still significant under this assumption. These articles had relatively little impact, perhaps because their findings merely reinforced previous results.

Several subsequent papers (e.g. Pelletier, 1997; Koscielny-Bunde et al., 1998) focused on the power-law behaviour of SAT from station records and Antarctic ice cores. Koscielny-Bunde et al. (1998) made a serious claim, a so-called “universality hypothesis”, that all SAT time series have the same Hurst exponent equal to 0.65. This claim was based on the analysis of just 14 stations, most of which were located in coastal areas in midlatitudes. Govindan et al. (2002), who questioned the fidelity of general circulation models on the basis of their inability to reproduce the power-law behaviour in 6 station SAT time series, stimulated some controversy. The climate models showed an absence of long-range correlations, i.e.

$H = 0.5$, while all the stations had $H \approx 0.65$. Global warming contrarians attracted attention to this article (e.g. http://www.heartland.org, Richard Lindzen's talks) by using it in their critique of climate models, all of which attributed the surface warming of the second half of the 20th century to anthropogenic emissions.

Using NCEP/NCAR reanalysis SAT and simulations of two climate models Fraedrich and Blender (2003) and Blender and Fraedrich (2003) showed that the "universality hypothesis" is not valid. The Hurst exponent was found to be greater over ocean than over land and therefore the comparison between individual stations and nearest grid points of coarsely resolved climate models is not fair, because the latter does not necessarily capture local climate conditions, especially in coastal areas. Fraedrich and Blender (2003) and Blender and Fraedrich (2003) also demonstrated that the large scale spatial Hurst exponent patterns are similar between the reanalysis and the models. However they also oversimplified the situation by stating that $H = 0.5$ for inner continental sites, $H = 0.65$ for coastal stations, and $H = 1$ over the ocean. Vyushin et al. (2004) examined 20th century simulations of NCAR PCM with 10 different combinations of anthropogenic and natural forcings. They concluded that the simulations with a historical volcanic forcing provide model SAT H close to the observed ones over land and also increase H over ocean, whereas the simulations without the volcanic forcing underestimate H everywhere.

Table A.1 also refers to several studies that show that the Hurst exponent estimates for observed SAT over land are likely affected by local conditions, such as regional land surface types, and by possible inhomogeneities present in the data. Eichner et al. (2003) analysed around a hundred stations and found that most of the SAT H values fall between 0.6 and 0.7. Kurnaz (2004) studied 384 stations in the western US and found H estimates mainly between 0.55 and 0.65. Kiraly et al. (2006) estimated H for more than 9000 stations around the globe and found most of the values between 0.6 and 0.9. Kiraly et al. (2006) used shorter time scales than Eichner et al. (2003) and Kurnaz (2004), which could lead to the higher estimates of H. None of the above mentioned station studies found a dependence of the H estimates on station distance to the nearest coast or on altitude. Kurnaz (2004) has noticed that at least a part of the spatial variability of H over the western US might be explained by local land surface types. Rust et al. (2008) analysed SAT from 24 European stations before and after removing detected data inhomogeneities caused, for instance, by a relocation of the measurement station or by installing a new type of shelter. They found that homogenization typically leads to a reduction of H estimates by 0.04-0.06. The sensitivity of the Hurst exponent estimates for observed SAT to local conditions and to possible data inhomogeneities makes it difficult for the current generation of coarsely resolved global climate models to reproduce the precise spatial distribution of H estimated for observed SAT.

Because of this controversy in the previous work in the field, we will focus in this thesis on the *large-scale* pattern of the H distribution within observational products and climate model simulations. A well characterized H distribution is required before any physical theory for the H distribution can be developed. At the time we began this work, we encountered several open questions:

- How method-independent and robust[2] are the observed and simulated spatial patterns of H for SAT? This general question might be broken into more specific questions, such as what is the sensitivity to:
  - estimation methods, including the choice of a frequency range for which the estimation is performed?
  - the choice of an observational product or a climate model?
  - the presence of radiative forcings and internal climate modes, such as ENSO?
  - the presence of data inhomogeneities?
- What is the statistical significance of differences between results?
- What are the spatial patterns of the power-law exponents for variables other than SAT?

[2] In this thesis the word "robust" is used as a synonym of the word "stable" (insensitive to small perturbations) and also as a synonym for "having a small variance" when it is applied to a statistical estimator.

- What is the ability of climate models to capture the H distribution for various variables?
- What are the underlying physical mechanisms leading to a growth of spectral power at low-frequencies?
- Is there a true scaling in the SAT or any other climate variable's temporal variability?
- What are the implications of the fact that climate variables' power spectra can be well approximated by a power-law?

In this thesis we will try to answer some of these questions.

1.3 Applications

Temporal power-law behaviour characterization has already been successfully applied to space-time modeling of winds in Ireland for wind energy studies (Haslett and Raftery, 1989), to weather derivatives pricing (Caballero et al., 2002), and to trend confidence interval estimation (Smith (1993) and Chapter 3 of this thesis). Recently it has been shown numerically that long-range correlations qualitatively affect extreme value statistics, e.g. the distribution and serial correlations of extreme events return intervals (see e.g. Bunde et al., 2005). The first applications of these results to climate have been reported already (e.g. Zorita et al., 2008), but many more are expected to appear. Such results and their applications are of significant practical importance for national and regional policy, public health, agriculture, industry, infrastructure, insurance, etc., because it can be argued that accurate quantification of extreme value statistics saves lives and money. Another field, where power-law spectral approximation might be useful, is the studies of potential climate predictability (e.g. Boer, 2004). Here we will summarize some of the above mentioned applications.

Haslett and Raftery (1989) used long-range temporal correlations to model wind velocity at 12 sites in Ireland, which allowed them to accurately estimate the confidence intervals for the annual mean generated wind energy. The basic idea is that in the case that the time series

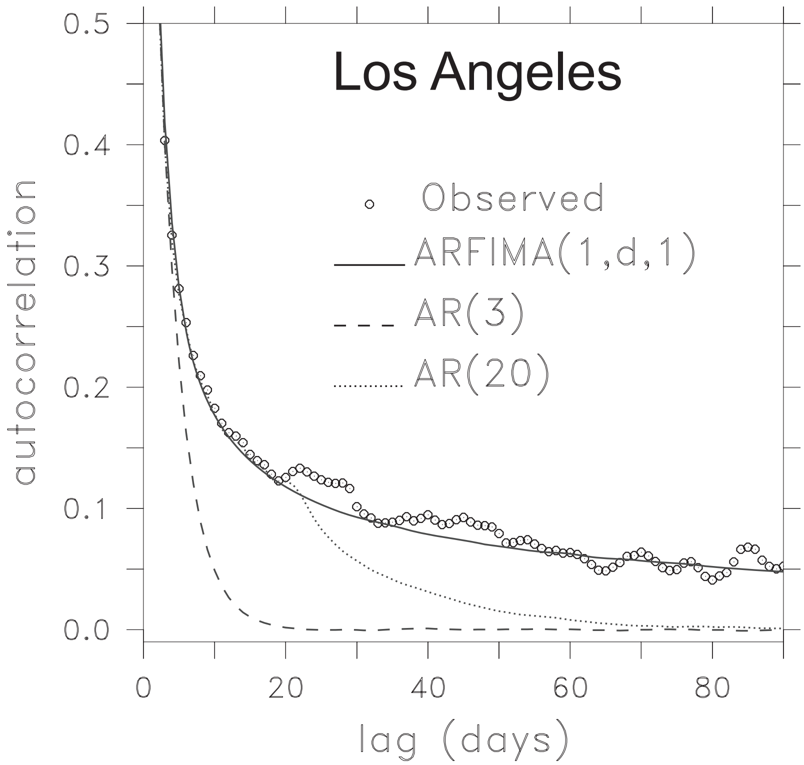

Figure 1.4: (Courtesy of Rodrigo Caballero) Autocorrelation function for Los Angeles SAT anomaly time series with the observed data (circles) and fits for the ARFIMA(1,d,1) (solid curve), AR3 (dashed curve) and AR20 (dotted curve) models.

autocorrelation function scales as s^{2H-2} for large time lags s then the standard error of the time series sample mean scales as σ/N^{1-H}, where σ is the time series standard deviation and N is the time series length. For the conventional case of short-memory processes, e.g. white noise or AR1, $H = 1/2$ and we get a conventional dependence on the inverse of the square root of N. However, when $H > 1/2$ the standard error of the sample mean decays slower with N than in the conventional case. Consequently it can be shown that the standard error of a linear trend superimposed on long-range correlated time series scales as σ/N^{2-H} (see e.g. Smith (1993) and Section 2.4 of this thesis).

A weather derivative is a form of insurance against adverse weather; a relatively cold winter is an example of adverse weather for natural gas consumers, whereas a relatively warm winter is adverse weather for natural gas suppliers. Trading of weather related securities, including weather derivatives, began around 1997 and according to Weather Risk Management Association reached a volume of around US$50 billion in 2006. Consider a weather derivative based on heating degree days (HDD), which are defined as $HDD_i = \max(T^* - T_i, 0)$, where T_i is the averaged temperature on day i and T^* is a threshold temperature, usually 18^{o}C (Caballero et al., 2002). The heating degree days index is defined as a sum of heating degree days over a certain period of length N, i.e.

$$I = \sum_{i=1}^{N} HDD_i.$$

The most important term in weather derivative price is

$$S = \int_0^{\infty} Q(I)P(I)dI,$$

where $Q(I)$ is a given payout function and $P(I)$ is the heating degree days index probability density function. Assuming that I is second order stationary and that $P(I)$ is Gaussian we have to estimate the mean and the standard deviation of I to estimate S. The mean of I is obtained from the temperature climatology, whereas the equation for the standard deviation of I is more elaborate:

$$\sigma_I^2 = \sigma_T^2 \Big(N + 2 \sum_{k=1}^{N} (N-k)\rho_k \Big),$$

where σ_T^2 is the standard deviation and ρ_k is the autocorrelation function of the temperature anomalies. Thus accurate estimation of the autocorrelations is very important for accurate pricing of weather derivatives. Underestimation of the autocorrelations leads to underestimation of a weather derivative and to a potential loss for its issuer. Caballero et al. (2002) found that a statistical model with autocorrelation function decaying asymptotically by a power-law, namely an autoregressive fractionally integrated moving average (ARFIMA(1,d,1), Beran, 1994; Taqqu, 2002) better captures a slow decrease of the observed daily temperature anomalies autocorrelations than autoregressive models even of the 20th order and therefore provides a more accurate estimate for a weather derivative price. An example of such slow autocorrelations decrease is demonstrated in Fig. 1.4. It plots an estimate of the autocorrelation function for Los Angeles SAT daily mean anomalies and its three approximations. AR3 and AR20 interpolate the first three and twenty autocorrelations respectively and then unsurprisingly quickly decay to zero. In contrast, the ARFIMA(1,d,1) model, which is the ARMA(1,1) model forced instead of white noise innovations by power-law innovations with the Hurst exponent equal to $d + 1/2$, closely approximates the autocorrelation function on time scales up to 90 days and possibly longer.

The idea of potential predictability was introduced to climate research by Madden (1976) and later developed by Zwiers (1987) and Zwiers and Kharin (1998). Boer (2004) considered several definitions of potential climate predictability, the basic idea of which is the ratio of a measure of the low-frequency variability to that of the total climate time series variability, i.e. $p = \sigma_L^2/\sigma^2$. We recall that time series variance is equal to the integral of its spectral density. Therefore using a power-law approximation for the spectral density we can rewrite the equation for p in the following way

$$p = (\int_{-\lambda_L}^{\lambda_L} b\lambda^{1-2H} d\lambda)/(\int_{-1/2}^{1/2} b\lambda^{1-2H} d\lambda),$$

where $0 < \lambda_L \leq 1/2$ is a threshold frequency between the low- and high-frequency variability. Thus we get

$$p = (2\lambda_L)^{2-2H}.$$

In two limiting cases $H = 1/2$ and $H = 1$ we have $p = 2\lambda_L$ and $p = 1$ respectively and p monotonically increases with H between these values. This explains the similarity between the spatial distribution of the p estimates for surface air temperature shown in Fig. 5 from Boer (2004) and of the H estimates shown in Chapter 6. More detailed connections between potential climate predictability studies and power-law temporal behaviour are outside the scope of the thesis and might be a subject of future research.

Chapter 2

Methodological Basis

2.1 Introduction to long-range correlated processes

The theory of stochastic processes with long-range correlated increments was originated by Kolmogorov in two short notes (Kolmogorov, 1940a,b) during his studies of turbulence. The seminal paper of Mandelbrot and Ness (1968) developed many of their properties and introduced the term "self-similar" to describe these processes.

There are at least two definitions of a self-similar process. The first one states that a real-valued stochastic process $Y = \{Y(t)\}_{t\in\mathbb{R}}$ is self-similar with index $H > 0$ if, for any $a > 0$,

$$\{Y(at)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{a^H Y(t)\}_{t\in\mathbb{R}}, \tag{2.1}$$

where $\stackrel{d}{=}$ denotes the equality of the finite-dimensional distributions (Taqqu, 2002). The Hurst exponent, H, comes in as a fundamental parameter governing the scaling properties of a self-similar stochastic process. In this thesis we use increments of a self-similar process for modeling low-frequency natural climate variability. The increments are defined as follows

$$X_i = Y_i - Y_{i-1}, \quad i \in \mathbb{Z}. \tag{2.2}$$

The autocovariance of X_i

$$\gamma(k) = \mathrm{cov}(X_i, X_{i+k}) = \frac{\sigma_X^2}{2}\left[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right], \quad k \in \mathbb{Z}, \tag{2.3}$$

where σ_X^2 is the variance of X_i, asymptotically decays by a power law (e.g. Beran, 1994)

$$\gamma(k) \sim \sigma_X^2 H(2H-1)|k|^{2H-2}, \quad \text{as} \quad k \to \infty. \tag{2.4}$$

The increments of a self-similar stochastic process for $1/2 < H < 1$ have long-range correlated behavior, since $\gamma(k)$ decays to 0 so slowly that $\sum_{k=-\infty}^{\infty} \gamma(k)$ diverges. In the case $\{Y(t)\}_{t\in\mathbb{R}}$ is a Gaussian process and satisfies (2.1) it is called fractional Brownian motion and the corresponding sequence $\{X_i\}_{i\in\mathbb{Z}}$ is called fractional Gaussian noise.

The second definition of a self-similar process states that a second order stationary sequence $\{X_i\}_{i\in\mathbb{Z}}$ with zero mean and finite variance is called second order self-similar if its autocovariance function is equal to that of fractional Gaussian noise (see Eq. (2.3)). One useful property of a second order self-similar process is that its autocorrelation function is invariant to temporal aggregation (Cox, 1984), e.g. its day to day autocorrelation is equal to month to month autocorrelation, year to year autocorrelation, and so on. We will use this property in Chapter 6 to compare the performance of two competing statistical models.

Physicists prefer to work with a spectral domain analog of the autocovariance function, namely the spectral density, due to the superior statistical properties of its estimates compared to autocovariance estimates. The spectral density of the increment sequence, $\{X_i\}_{i\in\mathbb{Z}}$, of a self-similar process scales by a power law in the vicinity of the origin

$$S_X(\lambda) \sim b|\lambda|^{1-2H}, \quad \text{as} \quad \lambda \to 0. \tag{2.5}$$

A stochastic process with such spectral density for all frequencies is called a "pink" noise or $1/f$ noise (more correctly $1/f^\beta$ noise) when $1/2 < H < 2$. Note that in the case $H \geq 1$ the variance of a process becomes infinite. In the case $H = 1/2$ and relation (2.5) holds for all frequencies we get a flat spectral density, which corresponds to a "white" noise process. For $H < 1/2$ a $1/f^\beta$ noise turns into a "blue" noise. The blue noise has $\beta < 0$. Stochastic processes with $H < 1/2$ are also called antipersistent.

Climatic time series usually have power spectrum (a spectral density estimate) with more complicated structure than that described by a single power-law function. However numerous studies in the past two decades have shown that on interannual to centennial time scales climatic spectra often have a single scaling regime with $1/2 \leq H < 1$, i.e. either a flat spectrum or a spectrum that corresponds to a class of long-range correlated processes. The processes which spectral density grows at high-frequencies and then saturates are called short-memory processes. Thus at low-frequencies short-memory processes' $H = 1/2$. Typically such processes can be well modeled by conventional autoregressive moving average (ARMA) models.

The next section provides an overview of statistical methods for estimating the parameters b and H for a given time series. In Section 2.3 we describe and use Monte-Carlo benchmarking to compare a suite of power-law estimation methods. The implications of LRC behavior for estimation of trend uncertainties and the number of years to detect a linear trend are mathematically described in Section 2.4. We provide a summary of this chapter in Section 2.5. The material in this chapter has been published in the Journal of Geophysical Research (Vyushin et al., 2007) and in the Journal of Climate (Vyushin and Kushner, 2009).

2.2 Description and Tests of Power-law Estimators

Many methods for estimating the Hurst exponent H are documented in the literature and a significant challenge in our analysis has been to reconcile the non-robust aspects of these methods. In this and the following section we describe several of the documented methods, develop some variants of our own, and characterize them using Monte-Carlo benchmarking. In Chapters 4-6, we will apply the methods to observed and simulated temperature data. The methods are summarized in Table 2.1. They include time domain methods, and periodogram and multitaper (spectral domain) methods. The Monte-Carlo benchmarking will show that all the methods agree reasonably well for simulated pure power-law stochastic processes. But when we apply the methods to observed data in Chapter 4, we will find that the methods are sensitive in various ways to the range of frequencies chosen and the filtering applied to the time series.

Table 2.1: The Hurst exponent estimation methods considered in the thesis. HF stands for high frequency and LF for low frequency.

Method	HF cutoff	LF cutoff	Remark
DFA(t)	s_{short}=18m	s_{long}=11y	Kantelhardt et al. (2001)
DFA(a)	s_{short}=18m	s_{long}=45y	Vyushin and Kushner (2009)
GPHE(t)	λ_{high}=1/18m	λ_{low}=1/15y ($l = 2$)	Robinson (1995b)
GPHE(a)	λ_{high}=1/18m	λ_{low}=1/45y ($l = 0$)	Hurvich et al. (1998)
MTM GPHE(t)	λ_{high}=1/18m	λ_{low}=1/15y ($l = 2$)	McCoy et al. (1998)
MTM GPHE(a)	λ_{high}=1/18m	λ_{low}=1/45y ($l = 0$)	Vyushin and Kushner (2009)
GSPE(t)	λ_{high}=1/18m	λ_{low}=1/15y ($l = 2$)	Vyushin and Kushner (2009)
GSPE(a)	λ_{high}=1/18m	λ_{low}=1/45y ($l = 0$)	Robinson (1995a)
MTM GSPE(t)	λ_{high}=1/18m	λ_{low}=1/15y ($l = 2$)	Vyushin and Kushner (2009)
MTM GSPE(a)	λ_{high}=1/18m	λ_{low}=1/45y ($l = 0$)	Vyushin and Kushner (2009)

2.2.1 Spectral Methods

The spectral methods find H by estimating the spectral slope. These methods first calculate an estimate $\hat{S}(\lambda)$ from a finite-length time series of the true spectrum $S(\lambda)$ and then find the best power-law fit to $\hat{S}(\lambda)$. We consider two choices of spectral estimators $\hat{S}(\lambda)$: the periodogram estimator (corresponding to the raw discrete spectrum) and the multitaper estimator (Percival and Walden, 1993). For a time series $X(t)$, $t = 1, \ldots, N$, the periodogram estimator is simply the square amplitude of the discrete Fourier transform divided by the time series length:

$$\hat{S}^{(p)}(\lambda_j) = \frac{1}{N}\left|\sum_{t=1}^{N} X(t)e^{-i2\pi t\lambda_j}\right|^2, \quad j = 1, \ldots, [N/2], \tag{2.6}$$

where $\lambda_j = j/N$ and the square brackets denote rounding towards zero. The periodogram is an asymptotically unbiased but inconsistent[3] spectrum estimator, since its variance is not a decreasing function of N: periodograms, as illustrated by the gray curve in Fig. 1.2, tend to appear noisy in spectral plots.

Multitaper spectral estimation (Thomson, 1982) provides an estimated spectrum with relatively reduced variance compared to the periodogram. It employs a set of K orthogonal "tapers" $h_k(t)$, $k = 1, \ldots, K$, that is applied to the time series $X(t)$. The multitaper spectral estimate is given by

$$\hat{S}^{(mt)}(\lambda_j) = \frac{1}{K}\sum_{k=1}^{K} S_k^{(d)}(\lambda_j), \quad j = 1, \ldots, [N/2], \tag{2.7}$$

where

$$\hat{S}_k^{(d)}(\lambda_j) = \left|\sum_{t=1}^{N} h_k(t) X(t) e^{-i2\pi t\lambda_j}\right|^2, \quad j = 1, \ldots, [N/2], \tag{2.8}$$

is the k-th direct spectral estimator. In this thesis we use sine tapers (Riedel and Sidorenko, 1995)

$$h_k(t) = \sqrt{\frac{2}{N+1}} \sin\left[\frac{k\pi t}{N+1}\right], \quad t = 1, \ldots, N. \tag{2.9}$$

The number of tapers, K, used in geophysical applications usually ranges between 3 and 5 (e.g., Ghil et al., 2002; Huybers and Curry, 2006). We choose $K = 3$ because of the large number of time series analyzed.

It can be shown that the variance of $\hat{S}^{(mt)}$ is a factor K smaller than the variance of $\hat{S}^{(p)}$ for large N. Thus multitaper spectra appear smoother in spectral plots; the smoothing effect is evident in the multitaper spectral estimate shown by the black curve in Fig. 1.2.

Given the spectral density estimator $\hat{S}(\lambda)$, we find a power law fit to $\hat{S}(\lambda)$ of the form $f(\lambda; b, H) = b|\lambda|^{1-2H}$ over a frequency range $\lambda_{\text{low}} \leq \lambda \leq \lambda_{\text{high}}$, where H is the Hurst exponent, b is a scaling factor, and λ_{low} and λ_{high} are low and high cutoff frequencies. For a review of these methods, known as spectral semiparametric estimation methods, see e.g.

[3] This is the only place in the thesis where the word "inconsistent" is used in a statistical sense. Everywhere else it has its regular meaning.

Moulines and Soulier (2002). We estimate b and H by minimizing

$$K(b,H) = \frac{1}{m-l} \sum_{j=l+1}^{m} k(\hat{S}(\lambda_j), f(\lambda_j; b, H)), \tag{2.10}$$

where $k(u, v)$ is the so called contrast function, which can be thought of as a distance between functions u and v. In the summation, l and m are indices related to the low and high-frequency cutoffs: $\lambda_{\text{low}} = \lambda_{l+1}$ and $\lambda_{\text{high}} = \lambda_m$.

The semiparametric power-law fits differ in their choice of contrast function $k(u, v)$. We here consider the Geweke-Porter-Hudak estimator (GPHE, Geweke and Porter-Hudak, 1983) with $k(u, v) = [\log(u) - \log(v)]^2$, which corresponds to log-linear regression, and the Gaussian semiparametric estimator (GSPE, Fox and Taqqu, 1988) with $k(u, v) = \log(u) + u/v$, which corresponds to a maximum likelihood estimator. GPHE is the best known and simplest of the two methods; the optimal b and H can be found in closed form along with confidence intervals. We use GPHE to obtain the CET power-law fit for the multitaper spectrum estimator in Fig. 1.2; the confidence intervals for GPHE with multitapering are found in McCoy et al. (1998).

GSPE is relatively more sophisticated and Robinson (1995a,b) has shown it to be superior to GPHE in various ways. Its optimization is not in closed form but reduces to a standard one dimensional numerical optimization procedure. Robinson (1995b) and Hurvich et al. (1998) show that GSPE has a factor of $\pi^2/6 \approx 1.7$ smaller asymptotic variance than GPHE $[1/(4(m-l))$ vs. $\pi^2/(24(m-l))$ for large $N]$. This property leads to practical advantages: in an analysis of power-law behavior in stratospheric ozone (see Chapter 3) it is found that GSPE yielded H estimates that are spatially smoother and more robust (have smaller variance) than GPHE.

2.2.2 Time domain method: Detrended fluctuation analysis (DFA)

The DFA (Peng et al., 1993; Kantelhardt et al., 2001) time domain estimator of H is, along with GPHE, the best known power-law fitting technique and has been applied widely in the life sciences, the earth sciences and physics. DFA works as follows. First, a cumulative sum time series is generated from the original time series $X(t)$. The cumulative sum time series is then split into segments of size s. Each of these segments is approximated by a least squares fit to a polynomial of order P, with P typically chosen to be between 1 and 5. The standard deviation of the residual of each least-squares fit is then calculated for each segment and then averaged over all the segments. This quantity is denoted $F(s)$ and is calculated for segment sizes s, where $P + 2 \leq s \leq s_{\mathrm{max}}$. The standard method is to take $s_{\mathrm{max}} = [N/4]$ (Kantelhardt et al., 2001) but we will test a variant with $s_{\mathrm{max}} = N$. The so-called "fluctuation function" $F(s)$ characterizes the noise at each time scale s; if the spectral density $S(\lambda) \sim \lambda^{1-2H}$ for small λ, the fluctuation function $F(s) \sim s^H$ for large s (Taqqu et al., 1995; Heneghan and McDarby, 2000).

Given this, we determine H by least-squares linear regression of $\log F$ against $\log s$ in the range $s_{\mathrm{short}} \leq s \leq s_{\mathrm{long}}$ (Peng et al., 1993; Kantelhardt et al., 2001), where s_{short} and s_{long} are short and long timescale cutoffs that correspond to the high and low frequency cutoffs for the spectral methods. For DFA3 (i.e. DFA with $P = 3$) we use a lower (high frequency) cutoff scale of s_{short}=18 months because it is only for longer timescales $F(s)$ for DFA3 might be well represented by a power-law function (Kantelhardt et al., 2001). The choice of s_{long} will be discussed in Section 2.3.

DFA is relatively straightforward to implement and can be used to infer information about the order of a trend of the time series. For example, a quadratic trend is effectively filtered out by DFA with $P > 2$. But unlike the GPHE and GSPE spectral methods DFA is numerically based and so lacks rigorous expressions for bias and confidence intervals estimates. Another disadvantage of DFA, is that, to our knowledge, there is still no theory allowing to estimate scaling factor b using DFA.

2.3 Benchmark tests of the estimator methods

Before applying the power-law fit methods to reanalysis climate data we benchmark the methods using Monte-Carlo tests of time series with known power-law behavior. These time series

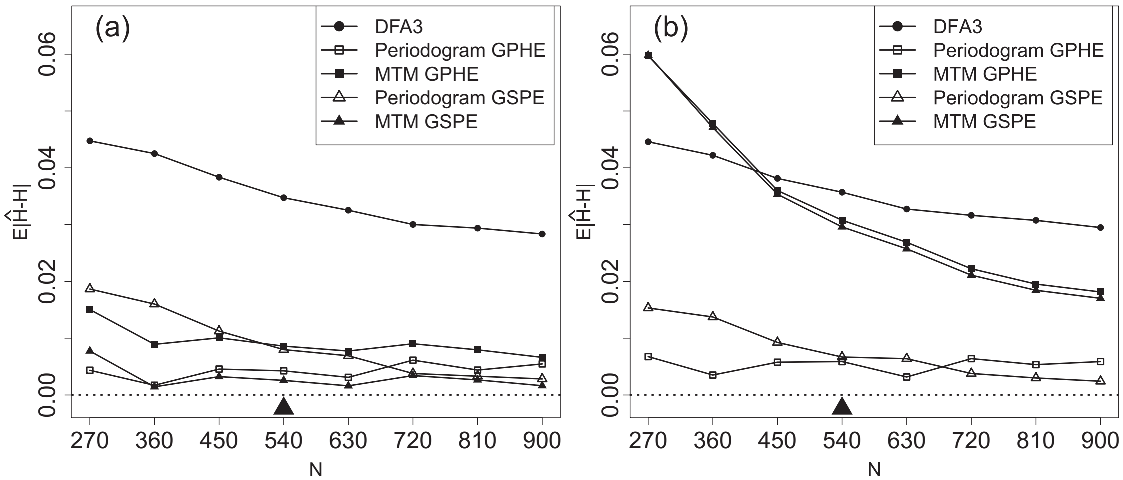

Figure 2.1: Absolute bias of the Hurst exponent estimators as a function of time series length N. Synthetic time series were simulated by ARFIMA(0,0.3,0), i.e. the true Hurst exponent was set to 0.8. Time series length $N = 540$ corresponds to the length of the monthly mean ERA40 data and is marked by a solid triangle. Panel (a) shows the results for the trimmed (t) version of the methods and panel (b) for the all-frequency (a) version of the methods. See Table 2.1 and Section 2.2 for a description of the methods.

are generated from autoregressive fractionally integrated moving average (ARFIMA) models, which are linear models for power-law stochastic processes (Beran, 1994; Taqqu, 2002). By convention, an ARFIMA(0,d,0) time series has Hurst exponent $H = d + 0.5$ (that is, the power spectrum $S(\lambda) \sim \lambda^{-2d}$). To mimic our climate data analysis, we take the ARFIMA time series to represent monthly mean records and estimate the Hurst exponent for frequencies lower than λ_{high}=1/(18 months). We choose λ_{high} =1/(18 months) for consistency with DFA3.

We find that H estimates in climate data are sensitive to the choice of frequency range, and that this sensitivity is method dependent. This is a practical issue encountered when dealing with time series that are not pure power-law stochastic processes with uniform behavior across all timescales. Many of the standard applications of power-law estimates build in these inconsistent ranges, for various reasons. For example, standard practice for DFA is to use $s_{\mathrm{long}} = [N/4]$, for periodogram spectral methods to use $\lambda_{\mathrm{low}} = 1/N$ and for the multitapered methods to use $\lambda_{\text{low}} \approx K/N$ (see Table 2.1 for references to each of these "conventional" methods). To test for the effect of these choices, we benchmark "all-frequency" (denoted "(a)") and "trimmed" (denoted "(t)") versions of the methods. The all-frequency methods set the low-frequency (long time scale) cutoff as low as possible. The trimmed methods cut off some of the lowest frequencies. Table 2.1 lists two versions of the methods we use, and in connection with the table we note the following:

- The multitapered methods conventionally trim the lowest frequencies and the periodogram methods conventionally use all frequencies. We here test trimmed and all frequencies versions of all spectral methods.

- For DFA3, DFA3(t) with $s_{\max} = s_{\text{long}} = [N/4]$ is the conventional method (Kantelhardt et al., 2001). We here test a version DFA3(a), whose time scale range is consistent with the all frequencies spectral methods. DFA3(a) uses $s_{\max} = s_{\text{long}} = N$.

We first test for the convergence of the magnitude of the estimators' bias, $< |\hat{H} - H| >$, as a function of time series length N, where $\hat{H}$ is the estimated value of the ARFIMA time series. Figs. 2.1a and b plot $< |\hat{H} - H| >$ for the trimmed and all-frequency versions of the five methods: DFA3, GPHE, MTM GPHE (i.e. multitapered GPHE), GSPE and MTM GSPE. Here the angle brackets represent the ensemble mean over 10,000 realizations of the ARFIMA model, for $H = 0.8$ and $270 \leq N \leq 900$. We see that the DFA3(t) estimator converges most slowly among the trimmed estimators, the periodogram spectral methods converge most quickly among the all frequencies estimators and neither DFA3 nor the periodogram methods are sensitive to the trimming. The convergence rate of the multitaper spectral methods falls between the periodogram and DFA3 methods for the all frequencies estimators. The increase in bias from tapering for the all frequencies cases is expected from general statistical principles: heavier smoothing, i.e. reduction of the variance from tapering, leads to an increase in the bias (von Storch and Zwiers, 1999). The effect of including the additional low frequency points degrades the convergence of the multitaper methods (Fig. 2.1b). This degradation is not

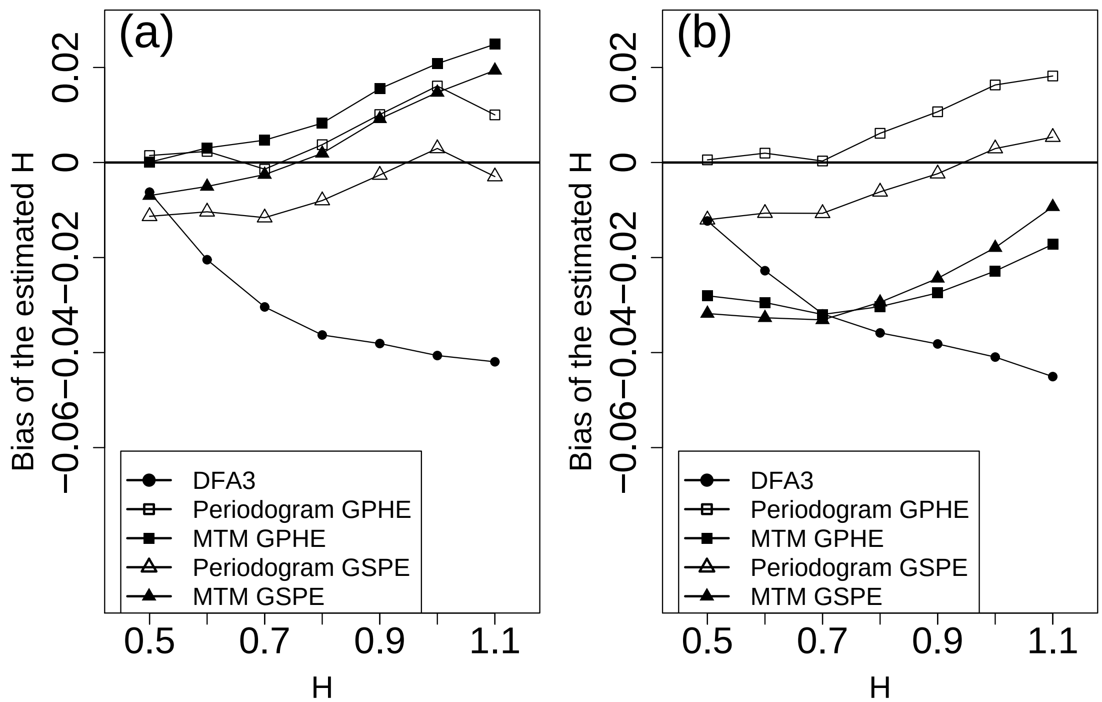

Figure 2.2: Bias of the Hurst exponent estimators as a function of the true Hurst exponent. Time series length was fixed to 540, which corresponds to the length of the monthly mean ERA40 data. The rest of the description is similar to Fig. 2.1.

surprising because the tapering impacts low frequencies most strongly. The impact of trimming on the spectral methods is consistent with theory (Hurvich et al., 1998; McCoy et al., 1998) but to our knowledge the robustness of DFA3 to changing from $s_{\text{long}} = [N/4]$ to $s_{\text{long}} = N$ has not been reported before.

Next we test how the estimators' bias depends on H for $N = 540$, which corresponds to the length of the monthly reanalysis time series analyzed in Chapters 4-6. In Figs. 2.2a and b we plot the bias $< \hat{H} - H >$ over 10,000 realizations. All the methods provide accurate estimates of H, within the estimators' standard deviation (see Fig. 2.3). DFA3 exhibits the largest bias among the trimmed estimators. This bias increases in magnitude with increasing H and is robust to the trimming. The periodogram methods have the smallest bias, which is also

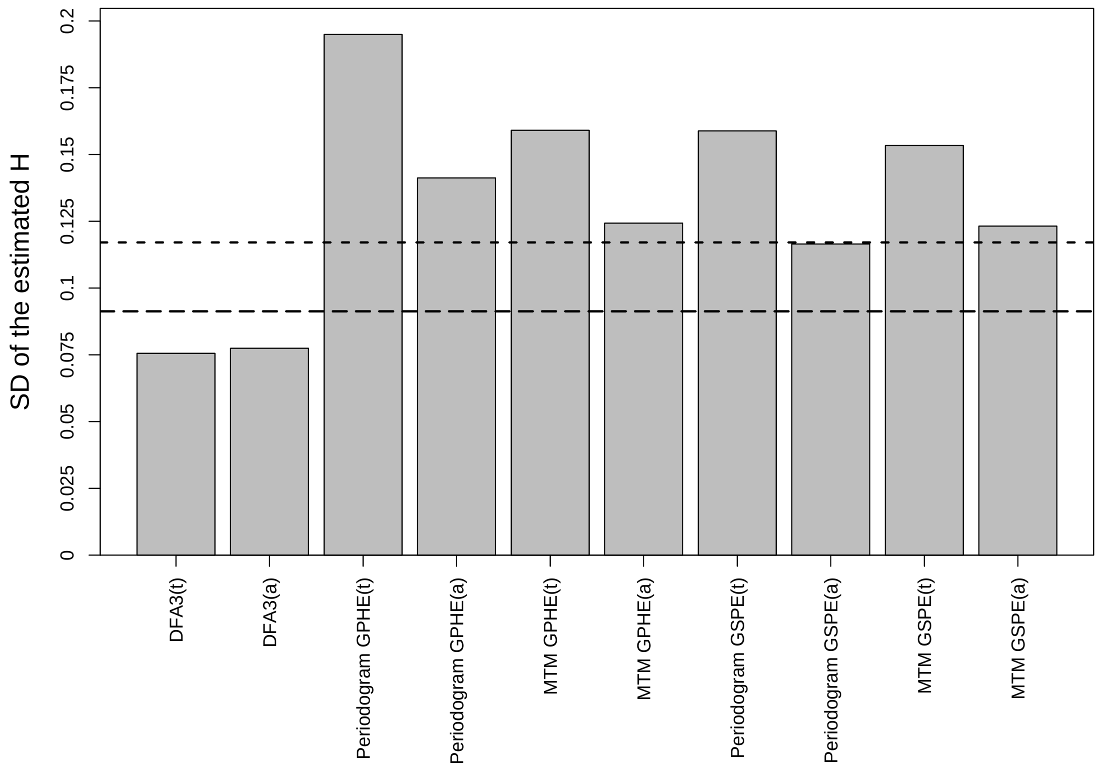

Figure 2.3: Standard deviation of $\hat{H}$ averaged for each method in Table 2.1 over the H values shown in Fig. 2.2. The short and long dashed lines demonstrate the asymptotic confidence intervals for GPHE and GSPE respectively.

robust to the trimming. The bias of the trimmed multitaper methods is comparable to that of the periodogram methods (Fig. 2.2a). But the bias increases for the all frequencies multitaper methods, indicating again that the multitaper H estimate degrades when all frequencies are included (Fig. 2.2b).

Fig. 2.3 shows the relative robustness of the different methods as measured by the estimator standard deviation $\sqrt{\left\langle\left(\hat{H}- < \hat{H} >\right)^2\right\rangle}$ averaged across values of H from 0.5 to 1.1. In this figure we also include the large N asymptotic estimates for periodogram GPHE and GSPE; these estimates are independent of H. DFA3 exhibits the least spread and is not sensitive to the trimming. However all the spectral methods have smaller variance when all frequencies are included. Among the spectral methods the periodogram GSPE(a) exhibits the least spread. The decrease of the $\hat{H}$ variance gained from multitapering is outweighed by the increased variance from the necessary trimming. The experimental standard deviations are consistently greater than the asymptotic ones (shown by the dashed lines in Fig. 2.3) for the time series length of 540, but the asymptotic results provide useful constraints in many applications.

Thus all the methods provide valid approaches to power-law fitting of pure power-law stochastic processes, but each method has distinct characteristics:

- DFA3 is robust (has the smallest variance) but has relatively large biases.
- GSPE is more complicated than GPHE but produces more robust estimates than GPHE and less biased estimates than DFA3.
- The standard trimmed MTM methods (MTM GPHE(t), MTM GSPE(t)) are less robust than the corresponding standard all frequencies periodogram methods (GPHE(a), GSPE(a)).

2.4 Trend variance and the number of years required to detect a trend

For the purpose of trend analysis memory is an issue. It is hard to distinguish a trend from natural variability if a time series is strongly serially correlated. The importance of taking into account LRC in trend analysis was first realized by Bloomfield (1992) during his studies of trends in surface air temperature. He proposed to use an ARFIMA model, introduced independently by Granger and Joyeux (1980) and Hosking (1981), for modeling temperature residuals. The idea is to fit the residuals, obtained after filtering out deterministic components of temperature time series such as seasonal cycle and trend, by ARFIMA and knowing analytical expression for the variance of ARFIMA calculate the variance of the trend. Bloomfield's approach can be classified as sequential full parametric estimation, since one first estimates and filters out the trend and then estimates the parameters of ARFIMA. Joint full parametric estimation, in which the trend and the parameters of ARFIMA are estimated simultaneously, was theoretically justified by Robinson (2005) and applied to Northern Hemisphere SAT anomalies by Gil-Alana (2005). The disadvantage of full parametric approach for trend detection studies is a problem of choosing the correct order of the ARFIMA, which itself is an issue (Beran et al., 1998). An appealing way to overcome the issue of model selection was proposed by Smith (1993). He showed that it is important to fit only the low frequency part of the residuals' spectrum using an asymptotic form of LRC spectral density $b|\lambda|^{1-2H}$ with only two unknown parameters. Then the variance of the trend can be calculated based on these two parameters. We will follow this direction in Chapters 3 and 6. This approach is classified as semiparametric since it requires estimation only of a part of the whole parameter set. Smith and Chen (1996) advocate for joint estimation of the trend and the parameters b and H. Unfortunately this theoretically more correct approach is still missing a solid mathematical foundation. Therefore in this thesis we implement sequential semiparametric estimation, i.e. we first estimate and filter out the trend from the time series and then find b and H for the residuals using semiparametric estimation. The general theoretical justification of this method is given by Yajima (1988).

2.4.1 Estimation of trend variance through autocovariance

Let us consider a general linear estimator

$$\hat{\xi} = \sum_{t=1}^{n} l(t)Y(t). \tag{2.11}$$

The variance of $\hat{\xi}$ may be expressed through autocovariance γ of $Y(t)$

$$\sigma^2(\hat{\xi}) = \sum_{t=1}^{n}\sum_{s=1}^{n} l(t)l(s)\gamma(t-s). \tag{2.12}$$

For example for the statistical model

$$Y(t) = \alpha + \beta y(t) + X(t), \tag{2.13}$$

where $y(t)$ is a certain explanatory variable (covariate) with zero mean and $X(t)$ is a noise, the ordinary least squares slope estimator $\hat{\beta}$ and its variance $\sigma^2(\hat{\beta})$ may be written as follows

$$\hat{\beta} = \frac{\sum_{t=1}^{n} y(t)Y(t)}{\sum_{t=1}^{n} y^2(t)}, \tag{2.14}$$

$$\sigma^2(\hat{\beta}) = \frac{\gamma(0)\sum_{t=1}^{n} y^2(t) + 2\sum_{k=1}^{n-1}\gamma(k)\sum_{j=1}^{n-k} y(j)y(j+k)}{(\sum_{t=1}^{n} y^2(t))^2}, \tag{2.15}$$

where n is the time series length, $\gamma(t)$ is the residuals' autocovariance function.

2.4.2 Approximation of autocovariance by exponential function

The most conventional way to proceed from this point is to use an exponential approximation for the residuals' autocovariance function $\gamma(t)$ for deriving an asymptotic formula for $\sigma(\hat{\beta})$. In principle one can use an estimate of $\gamma(t)$ to numerically evaluate $\sigma(\hat{\beta})$. However due to poor sampling properties of autocovariance function estimates statisticians prefer to use approximations of the sample autocovariance function. To obtain an exponential approximation for the autocovariance function one can fit an autoregressive model of the first order (AR1) to the noise. Symbolically AR1 can be written as follows

$$X(t) = \phi X(t-1) + \varepsilon(t), \tag{2.16}$$

where $-1 < \phi < 1$ is the month-to-month autocorrelation (lag-one autocorrelation coefficient) and $\varepsilon(t)$ is a Gaussian white noise. Let us review some of the AR1 model properties. Autocovariance function of AR1 decays exponentially

$$\gamma_{AR1}(t) = \sigma_X^2 \phi^{|t|}. \tag{2.17}$$

Spectral density of AR1

$$S_{AR1}(\lambda) = \frac{\sigma_X^2}{2\pi}\frac{1-\phi^2}{|1-\phi e^{-i\lambda}|^2} \to \frac{\sigma_X^2}{2\pi}\frac{1+\phi}{1-\phi}, \quad \text{as} \quad \lambda \to 0, \tag{2.18}$$

where σ_X is standard deviation of $X(t)$. Equations 2.17 and 2.18 are analogous to equations 1.2 and 1.3.

In the case we assume an AR1 model for the monthly resolved residuals $X(t)$ and take $y(t) = t - (n+1)/2$, we have

$$\sigma_{AR1}(\hat{\omega}) \approx \frac{\sigma_X}{N^{3/2}} \sqrt{\frac{1+\phi}{1-\phi}}, \tag{2.19}$$

where $\hat{\omega} = 12\hat{\beta}$ is the estimate of the linear trend in unit(Y)/year and N is the length of a considered period in years (Weatherhead and Coauthors, 1998) .

2.4.3 Approximation of autocovariance by power law function

The alternative approach is to use a power law approximation of the sample autocovariance function whose coefficients can be obtained by various estimation methods (see Section 2.2). Substituting $\gamma(0) = {\sigma_X}^2$, $\gamma(t) = at^{2H-2}$ for $t > 0$, and $y(t) = t - (n+1)/2$ into (2.15) and performing asymptotic derivations we obtain

$$\sigma^2(\hat{\beta}) \approx \frac{36a(1-H)}{H(1+H)(2H-1)} n^{2H-4}. \tag{2.20}$$

Scaling factors of the autocovariance and the spectral density, a and b, related as follows (e.g. Smith, 1993)

$$a = \frac{\pi b}{\Gamma(2H-1)\sin(\pi H)}. \tag{2.21}$$

Using this relation and some properties of the gamma function we can rewrite the asymptotic formula for $\sigma(\hat{\beta})$ in terms of b and H

$$\sigma_{LRC}(\hat{\beta}) \approx B(b, H) n^{H-2}, \tag{2.22}$$

where $\sigma_{LRC}(\hat{\beta})$ is the standard deviation of the estimated trend under the LRC hypothesis and

$$B(b, H) = \sqrt{\frac{72b\pi(1-H)}{(1+H)\Gamma(2H+1)\sin(\pi H)}}. \tag{2.23}$$

Formulas (2.19) and (2.22) were used in Fig. 3.8b and Fig. 3.9.

2.4.4 Estimation of trend variance through spectral density

Asymptotic formulas for the standard deviation of a regression coefficient (slope) can be derived only in cases when the explanatory variable $y(t)$ has a relatively simple form such as a linear trend. In other cases one can estimate the standard deviation of the slope only numerically. From the numerical point of view it is more convenient to express the autocovariance through the spectral density. Thus replacing in formula (2.12) the autocovariance by its expression through the spectral density of $X(t)$

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i\lambda k} S_X(\lambda) d\lambda, \tag{2.24}$$

we obtain

$$\sigma^2(\hat{\xi}) = \int_{-\pi}^{\pi} U(\lambda) S_X(\lambda) d\lambda, \tag{2.25}$$

where

$$U(\lambda) = \left| \sum_{t=1}^{n} l(t) e^{i\lambda t} \right|^2. \tag{2.26}$$

The important thing is that for the case of a trend estimator almost all weight of the function $U(\lambda)$ is concentrated near the origin. Therefore for calculation of the trend uncertainty a high frequency part of the spectrum is not important. This fact motivates the implementation of semiparametric (local) instead of full parametric (global) statistical models (Smith, 1993). For example in order to calculate the slope uncertainty in the case $y(t) = EESC(t) - \overline{EESC(t)}$, where $EESC(t)$ is the equivalent effective stratospheric chlorine time series, and the spectral density $S_X(\lambda)$ can be approximated by a power-law $b|\lambda|^{1-2H}$ we use the following equation

$$\sigma^2(\hat{\beta}_{EESC}) = b \int_{-\lambda_{\text{high}}}^{\lambda_{\text{high}}} U_{EESC}(\lambda) |\lambda|^{1-2H} d\lambda, \tag{2.27}$$

where $\lambda_{\text{high}} = \pi/12$ - high-frequency cutoff and

$$U_{EESC}(\lambda) = \left| \sum_{t=1}^{n} \frac{EESC(t) - \overline{EESC(t)}}{\sum_{s=1}^{n} (EESC(s) - \overline{EESC(s)})^2} e^{i\lambda t} \right|^2.$$

Therefore we could neglect intra-annual variability (frequency ranges $(-\pi/2, -\pi/12)$ and $(\pi/12, \pi/2]$ of the total ozone anomalies. Formula (2.27) was used in Fig. 3.8a.

2.4.5 Estimation of the number of years required to detect a trend

The number of years required to detect a trend of specified magnitude $|\omega|$ under the hypothesis that $X(t)$ can be well described by an AR1 model according to Weatherhead and Coauthors (1998) is as follows

$$N^*_{AR1} \approx \left[\frac{(2+z_p)\sigma_X}{|\omega|}\sqrt{\frac{1+\phi}{1-\phi}}\right]^{2/3}, \tag{2.28}$$

where N^*_{AR1} is the number of years required to detect a trend of specified magnitude $|\omega|$ (in particular one may choose $\omega = \hat{\omega}$) and z_p is the p-percentile of the standard normal distribution. In this setup the probability to reject the test hypothesis of zero trend when it is true is equal to 5% and the probability to accept the hypothesis of zero trend when it is false is equal to p. The number of years required to detect a linear trend of specified magnitude $|\omega|$ depends on three key parameters in the AR1 case (ω, σ, ϕ).

From (2.22) we derive an analogous formula for the case when $X(t)$ is long-range correlated

$$n^*_{LRC} \approx \left[\frac{(2+z_p)B(b,H)}{|\beta|}\right]^{\frac{1}{2-H}}. \tag{2.29}$$

In the above formula n is expressed in basic time units of the time series, i.e. days or months, and β has a unit "unit(Y)/(basic time unit)". Let us now transform this formula to the form which is conventionally used in ozone trend analysis when the time to detect the trend has units of years and the trend has units of DU/year. Let $n = TN$ and $\beta = \omega/T$, where T is the length of year in basic time units, i.e. $T = 365$ or $T = 12$, N is the length of the time series in years, and ω is the trend in DU/year. Then from (2.29) we get

$$N^*_{LRC} \approx \left[\frac{(2+z_p)B(b,H)}{|\omega|T^{1-H}}\right]^{\frac{1}{2-H}}. \tag{2.30}$$

This formula is somewhat similar to formula (2.28). However, due to the fact that the exponent in formula (2.30) is greater than the corresponding exponent in formula (2.28), trend error bars tend to be larger under the LRC hypothesis than under the AR1 hypothesis. This means that in the presence of long-range correlations we have to observe the time series longer in order to detect the trend with the same statistical significance. The number of years required to detect a linear trend of specified magnitude $|\omega|$ in the case when $X(t)$ is LRC also depends on three key parameters: magnitude of the trend $|\omega|$, spectral scaling factor b, and the Hurst exponent. It is worth noting that formula (2.30) is a generalization of formula (2.28). Thus for monthly resolved time series ($T = 12$) under assumption of AR1 model we get that

$$H_{LRC} \to H_{AR1} = \frac{1}{2}, \quad b_{LRC} \to b_{AR1} = \frac{\sigma_X^2}{2\pi}\frac{1+\phi}{1-\phi}, \tag{2.31}$$

and $B_{LRC} \to B_{AR1} = \sqrt{24\pi b_{AR1}}$. Therefore formula (2.30) reduces to formula (2.28). The numerical validation of this fact can be noticed by looking at the Southern Hemisphere mid- and high latitudes in Figs. 3.4b, 3.6b, 3.8, 3.9, and 3.11. The Hurst exponent converges to 0.5 as one moves from 30°S to 60°S as shown in Figs. 3.4b and 3.6b. Simultaneously the LRC trend error bars converge to the AR1 errors bars in Figs. 3.8 and 3.9, and the number of years to detect the trend under LRC hypothesis converges to the one under AR1 hypothesis in Figs. 3.10a and 3.11.

2.5 Conclusions

We have described two variants of the five Hurst exponent estimation methods. We have tested them using synthetic power-law time series (ARFIMA(0,d,0)) of various length and H. In the case when the lowest frequencies are trimmed DFA3 shows the largest bias, whereas when all the lowest frequencies are retained the DFA3 has a bias similar to the MTM methods and larger than the periodogram methods. However DFA3 has 1.5-2 time smaller variance comparing to other methods. By default DFA sets the largest used time scale equal to a quarter of the time series length. But we have demonstrated that this time scale range can be extended up to the time series length, because the statistical properties of DFA H estimate almost do not change in this case. These results, to our knowledge, represent a more comprehensive test of the time-domain and spectral-domain methods than has been previously carried out. It is reassuring that DFA3, which is widely used but not well justified, performs relatively well. Indeed, we find that DFA3 is one of the best Hurst exponent estimators. Overall the methods have shown a good performance and final selection of the two best methods, which we recommend to employ in similar studies, will be done in the next chapter based on the results of the methods' application to zonal mean tropospheric and stratospheric air temperature. Some of the Hurst exponent estimators, we have described and tested in this chapter, will be used in each of the following chapters. The equations for trend confidence interval estimation derived in Section 2.4 will be used for estimation of total ozone trends uncertainty through out Chapter 3 and for improving an IPCC result in the conclusions of Chapter 6.

We have assembled the code written in the R statistical language for the power spectrum estimators, Hurst exponent estimators, and their Monte-Carlo benchmarks described in this section into a package, which can be found at the URL http://www.atmosp.physics.utoronto.ca/people/vyushin/PowerSpectrum_0.3.tar.gz. The package, called PowerSpectrum, also includes functions for fitting AR1 model in the spectral domain, for estimation of a linear trend and its confidence intervals based on white noise, AR1, and power-law assumptions for the residuals, for cross spectrum estimation, for a spectral goodness-of-fit test, for Portmanteau tests, etc. In order to make our results easily reproduced and extended, we have made this package open source. A manual for this package is included in the thesis as Appendix B.

Chapter 3

Total ozone trend detection

3.1 Introduction

The problem of the long-term decline of stratospheric ozone (e.g. Stolarski et al., 1992; World Meteorological Organization, 1988) and, in recent years, of ozone recovery (e.g. Newchurch et al., 2003; Reinsel et al., 2005) has received wide attention from both the scientific community and the general public. Statistical models, particularly those based on multilinear regression methods, are commonly used for the detection of ozone changes (see SPARC (Stratospheric Processes and their Role in Climate) (1998) and references therein). Once a statistical model is established, it can be combined with a fitting method, for example ordinary least squares, to find the best fit to the observations. Ozone variations are typically represented as a combination of a long-term trend, natural periodic components (seasonal cycle, solar cycle, quasi-biennial oscillation, etc.), and a random component (the residuals). Knowledge about autocorrelations of the residuals of the regression model is required for a correct estimation of the model parameter uncertainties. Since the earliest ozone assessments (e.g. World Meteorological Organization, 1988) it has been assumed that the residuals can be described by an AR1 model, i.e. that the residual for a given month is proportional to the residual for the previous month plus random uncorrelated noise. In this case the autocorrelation function of the residuals $C(t)$ declines exponentially, i.e. $C(t) \sim \exp(-at)$, and the time series do not contain any significant long-term components other than those included explicitly in the model. In terms of equation 1.2 $a = -\log(\phi)$, where ϕ is a lag-one autocorrelation. Once the model parameters and their uncertainties have been estimated, they can be used, for example, to calculate the number of years required to detect a trend of a given magnitude at a given level of statistical significance (e.g. Weatherhead and Coauthors, 1998, 2000; Reinsel et al., 2002).

As we stated in Introduction chapter, geophysical time series do not always follow the AR1 model, however. They commonly exhibit slow autocorrelation function decay, which can be approximated by a power law, i.e. $C(t) \sim |t|^{2H-2}$, where $0.5 < H < 1$. There are also indications that ozone time series are not always well described by the AR1 model. Toumi et al. (2001) considered daily total ozone records from three west European stations (Arosa, Lerwick, and Camborne) and calculated Hurst exponents for deseasonalized and detrended time series (assuming a linear trend). All three time series exhibited Hurst exponents of about 0.78. However the authors did not remove the QBO- and solar cycle-related components, which could affect the estimate of the Hurst exponent. Varotsos and Kirk-Davidoff (2006) considered total ozone time series for large spatially averaged areas, but also removed only the seasonal cycle and the linear trend. The estimates of the Hurst exponents were calculated using DFA1 (see Section 2.2 for DFA details). (The DFA filters out polynomial trends whose order is less than the order of the DFA applied.) The Hurst exponents for tropical ozone estimated by (Varotsos and Kirk-Davidoff, 2006) were about 1.1, which corresponds to a spectral slope of -1.2 (in log-log coordinates). If we took the naive view that this value characterized the entire spectral range, this would imply an infinite spectral variance, since the integral of the spectral density would diverge. But as we will see later, strongly negative spectral slopes are characteristic of restricted frequency ranges of geophysical data, and these slope estimates can be affected by periodic signals not removed from the ozone time series in the Varotsos and Kirk-Davidoff analysis, like QBO and the solar cycle (see Jánosi and Müller (2005) and Chapter 4).

In recent years it has been established that a sizable fraction of the long-term ozone changes over northern midlatitudes can be related to long-term changes in dynamical processes (e.g. Weiss et al., 2001; Randel et al., 2002; Hadjinicolaou et al., 2005). Estimation of ozone trends requires a proper accounting for the effects of these processes on ozone. One approach is to add more terms to the statistical models used for trend calculations (e.g. Reinsel et al., 2005; Dhomse et al., 2006). However, the physical mechanisms behind these dynamical effects on ozone are often not well understood and therefore it is difficult to account for them properly in a statistical model (see further discussion in subsection 3.3.1). Furthermore such non-periodic components cannot be predicted and thus such models cannot be used to estimate future behaviour. An alternative approach is to consider the contribution of dynamical processes to ozone fluctuations to be part of the noise. In this case, the noise may be LRC and a proper estimation of the residuals' autocovariance is required.

Here we investigate the possible existence of LRC behaviour in total ozone time series and study its effects on ozone trend significance estimates and on the number of years required for trend detection. In this chapter we take a somewhat naive viewpoint that significantly greater than 1/2 values of $\hat{H}$ imply the LRC behaviour. Later, in Chapter 6 we will modify this viewpoint by demonstrating that this condition is necessary but insufficient. Here we employ spectral methods of Hurst exponent estimation instead of DFA because they provide a necessary estimate of the spectral scaling factor b (see subsection 2.2.1 and Section 2.4) together with an estimate of H, which are needed for trend confidence interval evaluation. An introduction to the theory of LRC processes has been given in Section 2.1. Some details of the spectral methods for Hurst exponent estimation have been presented in Section 2.2. Formulas elucidating the implications of LRC behaviour for trend uncertainties and for the number of years to detect linear trends have been derived in Section 2.4. The plan of the chapter is as follows. The total ozone data used in the analysis are described in Section 3.2. Section 3.3 is devoted to the statistical models and their estimates of the noise. We review the theoretical background in subsection 3.3.1. Long-term trends in total ozone are represented in terms of either the equivalent effective stratospheric chlorine (EESC) time series or a piecewise-linear trend (PWLT) with a turning point in early 1996. Evidence of spectral power growth of several total ozone time series, including station data, is given in subsection 3.3.2, while for TOMS/SBUV zonal averages it is quantified in subsection 3.3.3 and compared with AR1 behaviour. The significance of the long-term ozone decline is compared under two different assumptions for the ozone residuals (AR1 vs. LRC) in subsection 3.4.1. The recent positive ozone trend, and the number of years required to detect this trend under the two different assumptions, are compared in subsection 3.4.2 for both the EESC and PWLT derived trends. Some results for TOMS/SBUV gridded total ozone data, showing longitudinal structure, are discussed in Section 3.5. The main results are summarized and their implications discussed in Section 3.6. The material in this chapter has been published, along with relevant material in Chapter 2, in the Journal of Geophysical Research (Vyushin et al., 2007).

3.2 Data

The merged satellite data set used here is prepared by NASA and combines version 8 of TOMS and SBUV total ozone data (Frith et al., 2004; Stolarski and Frith, 2006); it is available from `http://hyperion.gsfc.nasa.gov/Data_services/merged/`. The data set provides a nearly continuous time series of zonal and gridded (10° latitude by 30° longitude grid) monthly mean total ozone values between 60°S and 60°N (higher latitudes have data gaps during polar night) for the period from November 1978 to December 2005. Here we consider only the period from January 1979 to December 2005. Some data, particularly the data for August-September 1995 and May-June 1996, were missing. Zonal averages estimated from ground based total ozone measurements (Fioletov et al., 2002) were used to fill the gaps. In addition, Dobson monthly mean total ozone values from three sites (Mauna Loa, Buenos Aires, and Hohenpeissenberg) were also analyzed here. These data are available from the WMO World Ozone and UV Radiation Data Centre (`http://www.woudc.org`).

3.3 Analysis of long-range correlations in total ozone time series

3.3.1 Statistical methods

A typical statistical model describing observations of monthly mean total ozone can be expressed in the form

$$\Omega(t) = a_0 + A(t) + Q(t) + S(t) + T(t) + X(t), \qquad (3.1)$$

where $\Omega(t)$ denotes total ozone, t is the number of months after the initial time (taken here as January 1979), a_0 is the mean, $A(t)$ represents the seasonal cycle, $Q(t)$ the quasi-biennial oscillation (QBO), $S(t)$ the solar cycle, $T(t)$ the long-term trend, and $X(t)$ are the residuals (noise). We used $A(t) = \sum_{j=1}^{4} a_{2j-1}\sin(2\pi jt/12) + a_{2j}\cos(2\pi jt/12)$, $Q(t) = (a_9 + a_{10}\sin(2\pi t/12) + a_{11}\cos(2\pi t/12))w_{30}(t) + (a_{12} + a_{13}\sin(2\pi t/12) + a_{14}\cos(2\pi t/12))w_{50}(t)$, and $S(t) = (a_{15} + a_{16}\sin(2\pi t/12) + a_{17}\cos(2\pi t/12))S_{107}(t)$, where $w_{30}(t)$ and $w_{50}(t)$ are the equatorial zonally averaged zonal winds at 30 and 50 hPa respectively (`http://www.cpc.ncep.noaa.gov/data/indices/`), and $S_{107}(t)$ is the solar flux at 10.7 cm (`http://www.drao-ofr.hia-iha.nrc-cnrc.gc.ca/icarus/`). We use winds at both 30 and 50 hPa, because they are about 90 degrees out of phase, which allows a better representation of the QBO signal in total ozone. The $\sin(2\pi t/12)$ and $\cos(2\pi t/12))$ terms in $Q(t)$ and $S(t)$ represent seasonal dependence. To describe the long-term trend in total ozone, two commonly used approaches are the equivalent effective stratospheric chlorine time series, $EESC(t)$ (`http://fmiarc.fmi.fi/candidoz/`) (Guillas et al., 2004; Newman et al., 2004; Fioletov and Shepherd, 2005; Stolarski and Frith, 2006; Weatherhead and Andersen, 2006), and a piecewise-linear trend with a turning point that is typically chosen in the second half of the 1990s. Similar to Reinsel et al. (2005) and Miller and Coauthors (2006) we choose a turning point n_0 in January 1996, because of the changes in ozone behaviour and in the EESC tendency in the late 1990s. Therefore, we use either $T(t) = (a_{18} + a_{19}\sin(2\pi t/12) + a_{20}\cos(2\pi t/12))EESC(t)$ or $T(t) = a_{18}T_1(t) + a_{19}T_2(t)$, where $T_1(t) = t$, for $0 < t \leq n$, where n is time series length (324 in our case), and

$$T_2(t) = \begin{cases} 0, & 0 < t \leq n_0, \\ t - n_0, & n_0 < t \leq n. \end{cases} \tag{3.2}$$

In order to provide analytical expressions for trends and their uncertainties, we use relatively simple trend models similar to those used by Reinsel et al. (2002) and Reinsel et al. (2005). In addition, one of key principles of statistical modeling is that the model be parsimonious, namely that it involve a minimum number of free parameters (von Storch and Zwiers, 1999). The more parameters are introduced, the easier it is to fit the time series and there is a risk that an improved fit may be fortuitous. This is particularly critical when the time series are very limited, as is the case with total ozone. We therefore restrict ourselves to Eq. (3.1) and do not, for example, introduce 12 coefficients for each component in Eq. (3.1) to more fully account for seasonal dependences. We have checked that using 12 coefficients for the QBO and/or trend terms does not alter the statistical properties of the residuals.

To test the impact of the El Chichon and Mt. Pinatubo volcanic eruptions we included SAGE aerosol optical depth observations into our regression model. For each eruption the aerosol loading was added to the model with the time lag that maximized the correlation between total ozone residuals and the aerosols. It was found that inclusion of volcanic aerosols only slightly decreases the Hurst exponent north of 30°S. Qualitatively, the Hurst exponent distribution and other results stay the same.

There are several reasons why we included the solar cycle and QBO into Eq. (3.1) but not other explanatory variables, for example, EP flux or tropopause height (see also World Meteorological Organization, 1998). First, ozone changes could affect temperature and other dynamical variables. Clearly, the solar cycle is not affected by ozone. In addition, QBO and solar variations are reasonably well-explained variations. EP flux forcing variations are not - they are part of the climate noise. If LRC manifest themselves through the EP flux forcing and we remove this forcing, then we just transfer the problem to that of understanding LRC in EP flux forcing. Furthermore, the correlation between ozone and dynamical variables could be different at different spectral intervals. The ozone-temperature correlation is a good example: the two fields are positively correlated on daily and monthly time scales but negatively correlated on an annual basis during major volcanic eruptions (Randel and Cobb, 1994). So, the relationship between ozone and such variables cannot be described by a single regression coefficient. This is not an issue for QBO and solar forcing because the variability of the QBO and the solar signal is located in a narrow spectral range. The QBO and solar cycles create maxima in the ozone time series power spectrum that could affect LRC estimates (Jánosi and Müller, 2005). Since we also want to estimate the number of years that is required to detect future changes, we have to make some assumption about the statistical model terms. We cannot predict the future solar and QBO signals, but we have their power spectra estimates. So, their impact on the future trend errors can be estimated. It is hard to make any predictions of dynamical variables or even about their spectral characteristics.

The parameters a_j of the model (3.1) are unknown coefficients identified by multilinear regression on the total ozone observations using least squares. The autocovariance of the residuals $X(t)$ affects the variance of a_j and should be properly accounted for. Certain assumptions are typically made about the behaviour of $X(t)$. For example, the AR1 model assumes that $X(t) = \phi X(t-1) + \varepsilon(t)$, where $\varepsilon(t)$ are independent normally distributed random errors. Similarly, the ARK model assumes that $X(t) = \phi_1 X(t-1) + \ldots + \phi_k X(t-k) + \varepsilon(t)$. The parameter ϕ can be estimated after estimation of the parameters a_j as the lag-one autocorrelation coefficient of the residuals, or it can be included in the model (3.1) directly and estimated simultaneously with the parameters a_j. In this thesis we follow the first approach, i.e. sequential estimation, because the simultaneous approach is still missing a solid mathematical foundation for a semiparametric power-law model, which will be described below. We estimate lag-one autocorrelation coefficient using the Yule-Walker method.

A different methodology is used if the autocorrelation function of X(t) decays by a power law, i.e. $C(t) \sim |t|^{2H-2}$, where $0.5 < H < 1$. The methods we use here are based on the fact that long-range correlations (dependence) in the time domain translate into a particular behaviour of the spectral density around the origin. It follows from the Abelian theorem that if the autocovariance $\gamma(t) \sim |t|^{2H-2}$ as $t \to \infty$, where $0.5 < H < 1$, then the spectral density $S(\lambda) \sim b|\lambda|^{1-2H}$ as $\lambda \to 0$ (see Taqqu, 2002), where by definition

$$S(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma(t) e^{-it\lambda}.$$

In particular, the log of the spectral density is a linear function of $\log(\lambda)$ as $\lambda \to 0$. In contrast, the spectral density of an AR1 process is a constant function of λ under the same conditions (see Section 2.4) and can be considered as a particular case of a more general power-law model. Thus, as shown in Section 2.4, the results we obtain for the LRC model are generalizations of those for the AR1 model and reduce to the latter when H tends to 0.5.

The GPHE and the GSPE are the two methods used in this chapter to estimate the two parameters, b and H, of the spectral density approximation, as described in Section 2.2. We recall that GPHE estimates b and H by means of a linear regression of the $\log$(periodogram) on $\log(\lambda)$ and that GSPE is a maximum likelihood estimator (see Section 2.2 for an estimators description). The variances of the coefficients a_j of the statistical model (3.1) can be expressed as a function of b and H, as discussed in Section 2.4. Furthermore, they can be used to estimate the number of years that is required to detect a statistically significant trend of a given magnitude (see Section 2.4).

The integral of the autocorrelation function from negative infinity to positive infinity, which is one way of quantifying a decorrelation time, is finite for an AR process and infinite for an LRC process. This means that, in contrast to the case with an AR process where the limit $t \gg t_{decorrelation}$ is well-defined, two observations of an LRC process do not become statistically independent in the limit of arbitrarily large time separations (von Storch and Zwiers, 1999). Among many possible mechanisms generating LRC behaviour, which have been listed in Section 1.1, we think that at least two might be relevant for total ozone. One is based on the aggregation of an infinite number of AR1 processes whose time scales satisfy certain conditions (Granger, 1980). In practice, apparent LRC behaviour may obtain from the aggregation of a finite number of AR1 processes whose longest time scale is comparable to the length of the time series (Maraun et al., 2004). This is a definite possibility in the case of ozone time series where the records are comparatively short. A second possible origin of LRC behaviour is a sequence of shocks or pulses with stochastic magnitudes and durations (Parke, 1999). Volcanic eruptions could play such a role, although as noted earlier a direct link between aerosol loading and total ozone for the time period 1979-2005 does not appear to be associated with LRC behaviour. The attribution of LRC behaviour in total ozone is a separate topic which is not addressed here.

3.3.2 Illustrations of long-range correlations

Fig. 3.1 shows time series of the residuals $X(t)$ for Eq. (3.1), obtained by filtering out the mean, seasonal cycle, QBO, solar flux, and EESC trend, for zonal and monthly mean total ozone in various latitude bands from 1979-2005. The corresponding Hurst exponents H, estimated using the GSPE (see Section 2.2 for an estimator description), are also indicated. The latitude bands correspond to local maxima or minima of H (see Fig. 3.4b below), and have been chosen to illustrate the different temporal behaviour that is exemplified by large or small values of H. The time series with larger values of H tend to exhibit greater low-frequency variability with more instances of strong apparent "trends" over decadal timescales. Values of H that are close to 0.5 correspond to behaviour that is not significantly different from AR1, while the larger values of H are clear indicators of LRC behaviour.

To illustrate how the Hurst exponent is calculated by the GPHE we show power spectra of monthly mean total ozone residuals for three ground-based stations, Mauna Loa (19.5°N, 155.6°W), Buenos Aires (34.6°S, 58.5°W), and Hohenpeissenberg (47.8°N, 11.0°E), as well as for the corresponding nearest grid points and zonal averages from the merged satellite data set. For the purpose of comparison the period was limited to 1979-2005 for all data sets. Several months with missing data were filled by linear interpolation in time. Fig. 3.2 shows the periodograms in log-log coordinates of the total ozone residuals for station data (panels (a,b,c)),

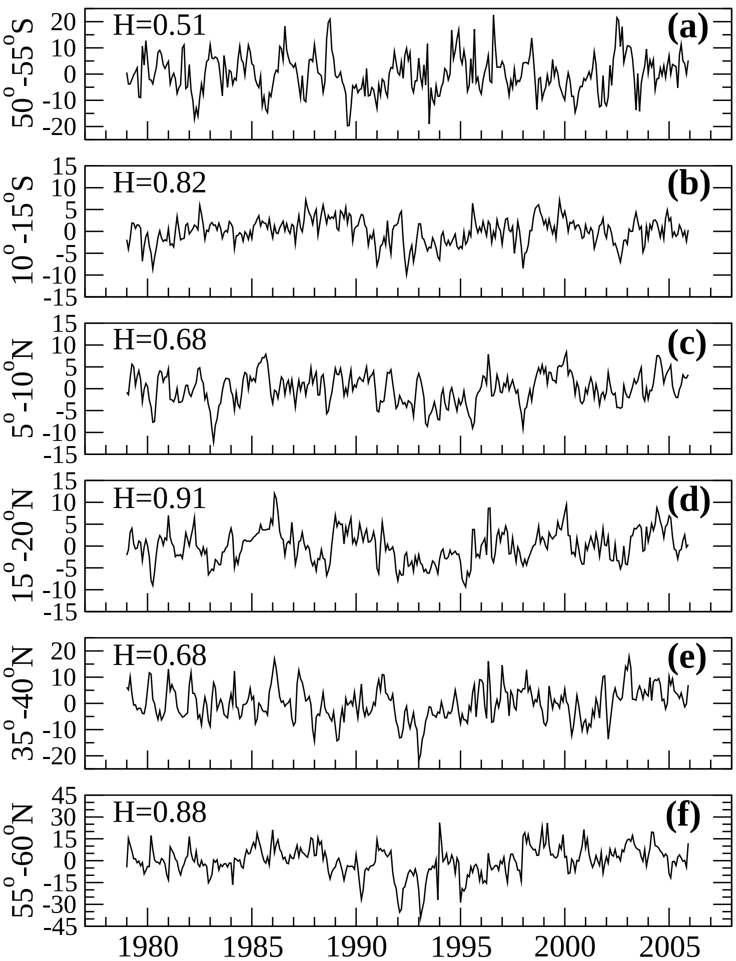

Figure 3.1: Monthly and zonal mean total ozone residuals in DU obtained by filtering out the seasonal cycle, QBO, solar cycle, and EESC fit for various 5° latitude bands, as indicated. The Hurst exponent for each time series is indicated in the top left corner of each panel. Note the differences in the extent of low-frequency behaviour in the time series with different Hurst exponents.

for the nearest grid points from the merged data set (d,e,f), and for the corresponding zonal averages from the merged data set (g,h,i). The periodograms show an increase in variability with a decrease of frequency, which as noted earlier is a manifestation of LRC behaviour. The solid straight lines are the best linear fits of the periodogram in log-log coordinates, corresponding to power law approximations in ordinary coordinates for the frequency bandwidth 1-27 years.

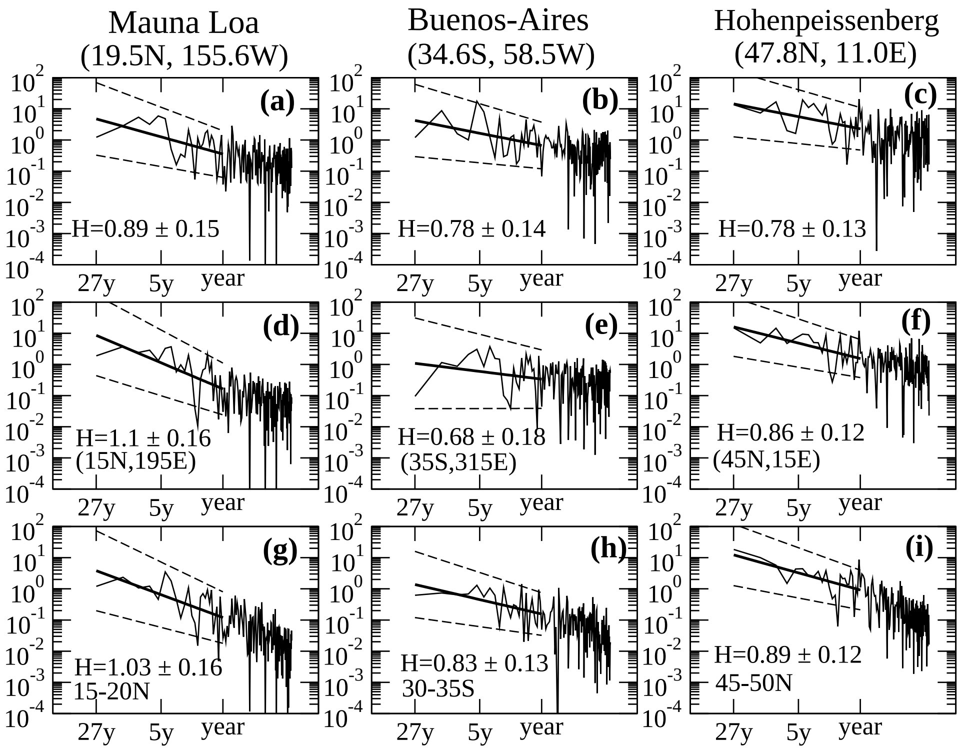

Figure 3.2: The periodograms in log-log coordinates for monthly mean total ozone residuals obtained by filtering out the seasonal cycle, QBO, solar flux, and EESC fit. The solid lines are the best linear fits, while the dashed lines represent the one sigma uncertainty envelope. Panels (a), (b) and (c) show ground-based station data as indicated, panels (d), (e) and (f) the merged satellite data from the grid point nearest to the corresponding station, and panels (g), (h) and (i) the corresponding zonal averages from the merged satellite data.

The dashed straight lines represent the one sigma uncertainty envelope defined by the standard errors for slope and intercept. The linear fit of the periodogram represents the calculation of the Hurst exponent using the GPHE (see Section 2.2 for details). Apart from panel (e) all slopes are statistically significantly less than zero, meaning that H is statistically significantly greater than 0.5 ($H = (1 - slope)/2$). Therefore eight out of the nine power spectra shown reveal that the corresponding total ozone residuals are LRC (and the slopes for panels (b) and (e), which should be comparable, agree within the error bars). The zonal average time series typically have slightly greater Hurst exponents than the grid point time series from the same

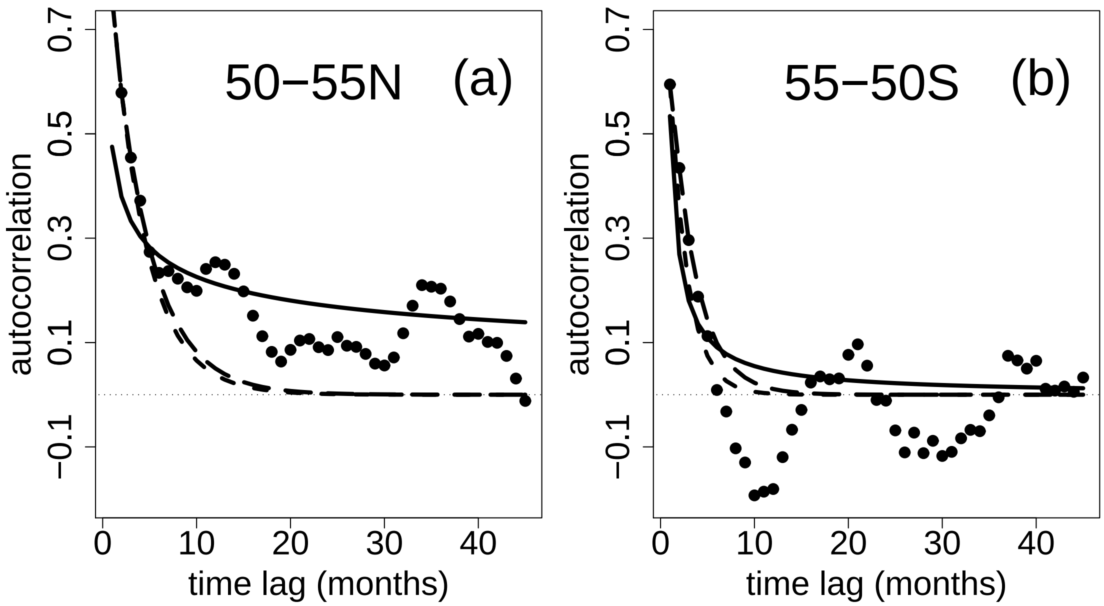

Figure 3.3: Autocorrelation function for two sample latitude bands (solid circles) of monthly and zonal mean total ozone residuals obtained by filtering out the seasonal cycle, QBO, solar flux, and EESC fit, together with that of various approximations (lines). The autocorrelation function of the best-fit AR1 model is shown by the short dashed curve, that of the AR3 model by the long dashed curve, and that of the best-fit power law function by the solid curve. The 50°-55°N latitude band (panel (a)) shows clear evidence of LRC, while the 50°-55°S latitude band (panel (b)) shows no such evidence.

latitudinal belt since the average of several time series tends to have a Hurst exponent equal to the maximum of the Hurst exponents of the individual time series (Granger, 1980). Additional evidence that this is the case for total ozone will be provided in Section 3.5.

To illustrate why it can be important to allow for LRC behaviour in a statistical model, Fig. 3.3a shows the autocorrelation function of the residuals $X(t)$ for 50°-55°N as well as the fits produced by various statistical models. In this case, H is significantly different from 0.5. Fig. 3.3a shows that the AR1 model (the short dashed curve) does not fit well the autocorrelation function of the total ozone residuals (the solid circles) for periods longer than several months. Even using higher order AR models such as AR3 (the long dashed curve) does not improve the fit. The most parsimonious approximation of the solid circles is a simple power law function $a|t|^{2H-2}$ (the solid curve). Autocorrelation functions of the AR1 and AR3 models rapidly decay, while the power law function decays slowly and follows the autocorrelation function of the original time series.

However as seen in Fig. 3.1 there are latitude bands where LRC behaviour is not evident. For example over 50°-55°S, where H is not significantly different from 0.5 (see the appropriate panel of Fig. 3.1 and Fig. 3.4b), the power law function does not provide a superior fit to the autocorrelation function of the total ozone residuals, whereas a reasonably good fit is provided by the AR3 model (Fig. 3.3b). Thus, it is important to establish where LRC behaviour is evident and where it is not (see a more general discussion on this topic in Chapter 6).

3.3.3 Quantification of long-range correlations in zonally averaged ozone

Ozone trend studies are typically performed using zonally averaged data. Part of the motivation for this lies in the approximate zonal symmetry of the stratosphere and thus of quantities such as ozone trends. By taking a zonal average of the data longitudinal fluctuations (eddies) are removed, thereby reducing the standard deviation (noise level) of the time series while keeping the zonally symmetric trend unchanged. Thus zonal averaging usually leads to an increase of the signal to noise ratio. However a drawback of zonal averaging for trend detection is an increase in the strength of serial correlations (see also Section 4.5).

In this subsection we systematically analyze the zonally averaged total ozone residuals in the merged satellite data set as a function of latitude. What is considered a residual depends on how the trend contribution is defined. We thus consider (and compare) two versions of the residuals, which represent the trend contribution in different ways: one with the EESC function and the other with the PWLT. The residuals are then analyzed using the AR1, GPHE, and GSPE approaches. The GPHE and GSPE were applied to the frequency bandwidth from 1 to 27 years. The estimates for the square root of b and H for the case where the EESC trend is removed are shown in Fig. 3.4a and b, respectively. The GPHE and GSPE provide

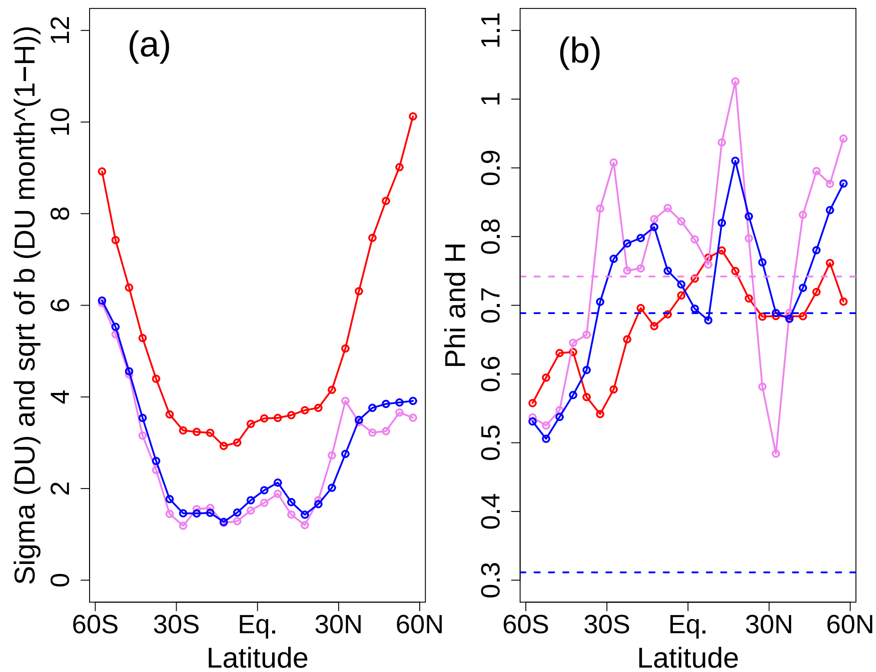

Figure 3.4: Estimates of various statistical parameters for monthly and zonal mean total ozone residuals obtained by filtering out the seasonal cycle, QBO, solar flux, and EESC fit, shown as functions of latitude. (a) Standard deviation σ calculated by AR1 (red circles) and the square root of the parameter b calculated by GPHE (violet circles) and GSPE (blue circles) applied to the frequency bandwidth 1-27 years. (b) Month-to-month lag-one autocorrelation ϕ calculated by AR1 (red circles) and the Hurst exponents calculated by GPHE (violet circles) and GSPE (blue circles). The violet (blue) dashed lines in panel (b) indicate the 95% confidence intervals for GPHE (GSPE). Thus, time series for which the Hurst exponents lie above the dashed lines may be roughly considered as LRC at the 95% significance level (see the comment about conditions of LRC existence in Section 3.1).

consistent estimates for the Hurst exponents, although GPHE is more spatially noisy. (This lends further support to our conclusion, following the analysis in Chapter 4, that GSPE is preferable to GPHE because it typically produces results with reduced disagreement between spatially proximate time series.) All but one Hurst exponent shown in Fig. 3.4b is less than one,

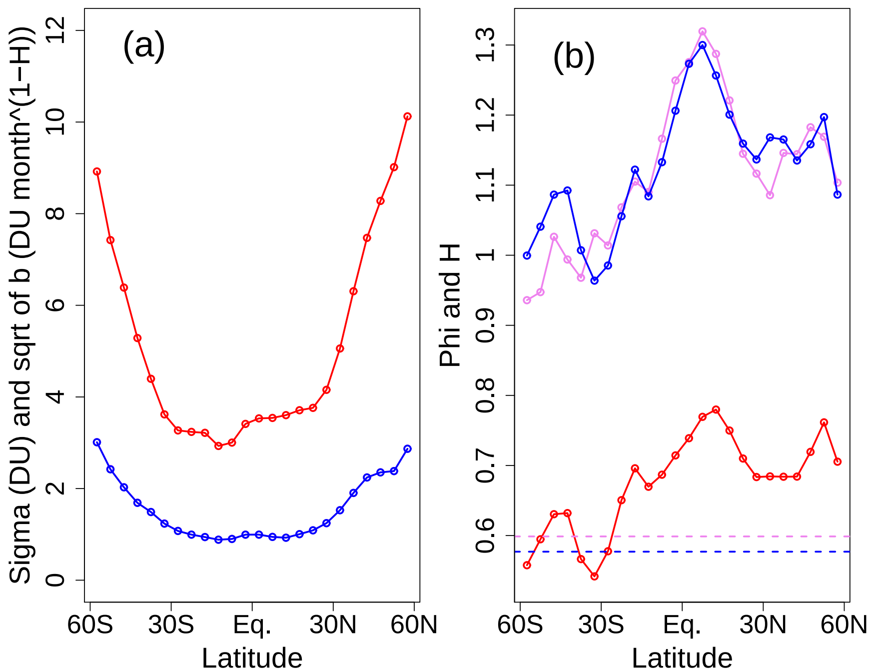

Figure 3.5: Same as Fig. 3.4, but with GPHE and GSPE applied to the frequency bandwidth 2 months-27 years. Note the different vertical scale in panel (b) compared with Fig. 3.4. This analysis yields spurious results, namely Hurst exponents greater than unity (indicating non-stationary behaviour). The figure is included to highlight the importance of choosing an appropriate bandwidth (see text for further discussion).

i.e. the corresponding time series are second order stationary with finite and time-independent mean and standard deviation. The estimates of σ (standard deviation of the residuals) and ϕ (lag-one autocorrelation of the residuals) for the AR1 model are also shown in Fig. 3.4. Although $\sqrt{b}$ and σ, as well as H and ϕ, cannot be compared directly, they represent similar quantities: the first pair is a measure of the variability of the residuals, while the second pair is an indicator of the persistence in the time series. It is interesting to note that the maxima and minima of b and σ, as well as of H and ϕ, tend to occur in roughly the same latitude bands.

The theory for distributions of H estimated by GPHE and GSPE exists only for the asymptotic case $m \to \infty$, where m is the number of frequencies used, with some other additional

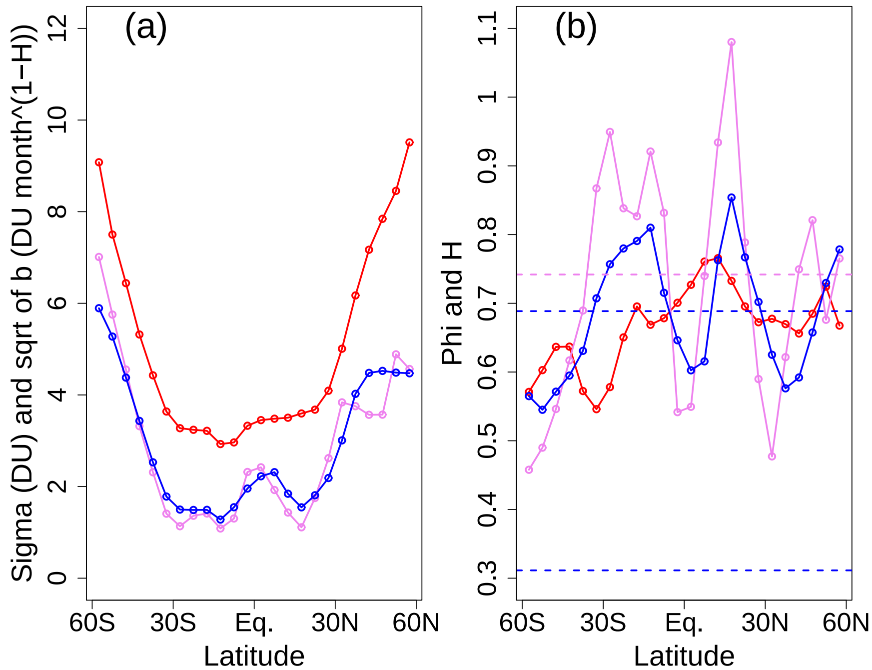

Figure 3.6: Same as Fig. 3.4, but with PWLT filtered out instead of EESC to describe the long-term trend. The Hurst exponents are similar to those in Fig. 3.4 in the Southern Hemisphere, but are reduced in magnitude by about 0.1 in the Northern Hemisphere.

conditions [see *Robinson*, 1995a,b]. Under these conditions the theorems proved by Robinson state that the estimates of H obtained by GPHE and GSPE are distributed normally with a mean equal to the true value of H and variances equal to $\pi^2/24m$ and $1/4m$, respectively (see also Section 2.2.1). Therefore GSPE has a smaller asymptotic variance than GPHE by a factor of $\pi^2/6$. The values $0.5 \pm 1.96\pi/\sqrt{24m}$ and $0.5 \pm 1.96/(2\sqrt{m})$ are indicated in Fig. 3.4b by dashed violet and blue lines, respectively. All Hurst exponent estimates located above these lines may be considered as greater than 0.5 with 95% statistical significance, meaning that the corresponding time series may be parsimoniously described by an LRC model. This applies to just over half the latitudes analyzed. There is clear evidence of LRC at certain latitude bands, while at other latitude bands the autocorrelation behaviour is not significantly different from the AR model. Interestingly, the latitudinal structure of LRC behaviour is quite different in the

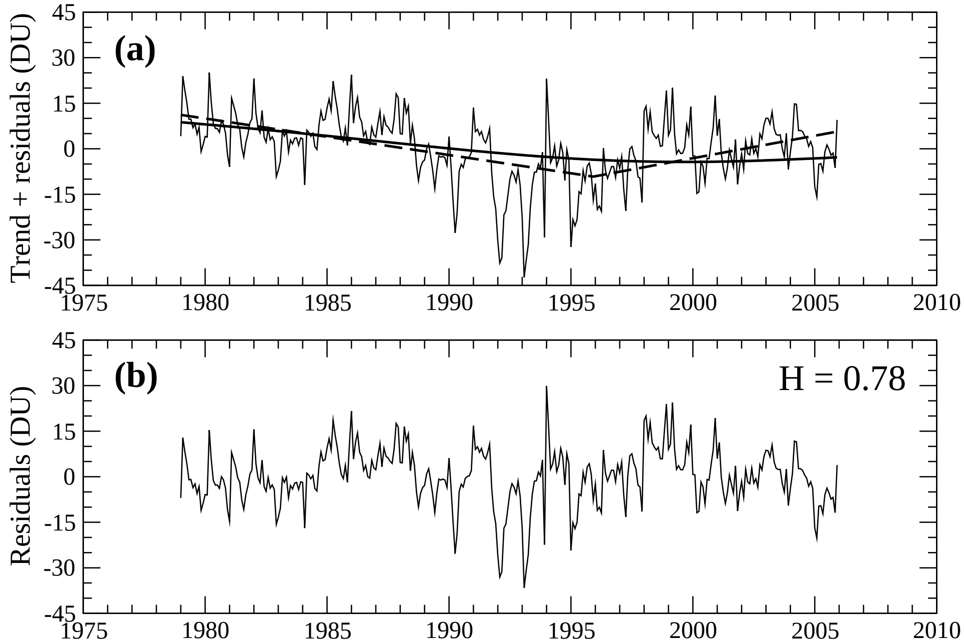

Figure 3.7: (a) Monthly and zonal mean total ozone anomalies for the 55°-60°N latitude band obtained by filtering out the seasonal cycle, QBO, and solar flux, together with the EESC (solid) and PWLT (dashed) fits. (b) The corresponding total ozone residuals when the PWLT fit is removed. Panel (b) may be compared directly with the EESC-based residuals shown in Fig. 3.1f.

two hemispheres.

It should be emphasized that the LRC methods discussed here are based on asymptotic approximations at low frequencies and therefore they could be sensitive to the frequency interval used for the parameter estimation: a wider interval may yield a bias in the estimates, while a narrower interval results in larger uncertainties of the estimates. Fig. 3.5 is similar to Fig. 3.4, except that GPHE and GSPE were applied to the entire frequency bandwidth from 2 months to 27 years. (The results for the AR1 model (red circles) are identical to those shown in Fig. 3.4.) The Hurst exponents shown in Fig. 3.5b are almost everywhere greater than one, i.e. they belong to a nonstationary range. However, Fig. 3.2g,h,i demonstrates that the periodograms for the zonal averages have steeper slopes for the bandwidth 2 months-1 year than for the bandwidth 1 year-27 years. Therefore the fact that the calculated Hurst exponents are greater than one in this case is a result of including the high frequency (sub-annual) part of the spectrum in the fit. We will further discuss the importance of the frequency range choice in Chapter 4.

To investigate the dependence of the residuals on the definition of the long-term trend, the calculations were repeated but with the residuals defined by using the PWLT in Eq. (3.1) instead of the EESC time series. The results are shown in Fig. 3.6, and may be compared with Fig. 3.4. In the Southern Hemisphere the statistical parameters (and their latitudinal variations) are very similar in the two cases. However there is a distinct change in the Northern Hemisphere, where the Hurst exponents decrease by about 0.1. Over a broad region of the midlatitudes H is no longer significantly different from 0.5, implying the loss of LRC in this region; and at the highest sub-polar latitudes, the extent of LRC behaviour is strongly reduced. Inspecting Fig. 3.1f, corresponding to 55°-60°N, it is evident that the major low-frequency variation in the residual defined relative to $EESC(t)$ projects strongly on a piecewise-linear trend with a turning point in early 1996, and its contribution to the residual is therefore substantially reduced when the PWLT function is used to define the long-term trend. This is illustrated by Fig. 3.7, which shows the ozone time series for 55°-60°N (with mean, solar, QBO and seasonal cycle filtered out) together with the EESC and PWLT fits (panel (a)) and the PWLT residual (panel (b)); the latter may be compared with the EESC residual shown in Fig. 3.1f.

3.4 Significance of long-term trends in zonal-mean total ozone

3.4.1 Long-term ozone decline

The statistical model used to describe the noise does not affect the mean trend estimated by Eq. (3.1), but it does affect the estimated uncertainty of the trend. The regression coefficient of the total ozone anomalies on the EESC time series is shown in Fig. 3.8a as a function of latitude for the period 1979-2005. The magnitude of the regression coefficient basically displays the sensitivity of total ozone in that latitude band to the stratospheric abundance of ozone-depleting substances as represented by the EESC. For comparison with other estimates the result is presented in units of DU/year for the time period 1979-1995, during which time the EESC time series is nearly linear with a net change of approximately 1.0 ppb of chlorine. As is well known, the long-term ozone decline has a minimum in the tropics and increases towards the poles, with larger values in Southern Hemisphere as compared with Northern Hemisphere midlatitudes. The strong increase of the Southern Hemisphere trend with latitude is indicative of the large influence of Antarctic ozone loss on the Southern Hemisphere midlatitude long-term decline (e.g. Chipperfield, 2003; Fioletov and Shepherd, 2005).

The error bars in Fig. 3.8a indicate the 95% confidence intervals estimated under AR1 and LRC hypotheses concerning autocorrelation of the residuals. The trend uncertainties under the LRC hypothesis are evidently wider than those under the AR1 hypothesis. The differences are particularly large where H exceeds 0.7, which from Fig. 3.4b occurs basically everywhere north of 35°S. In this region the standard deviation of the trend under the LRC hypothesis is up to 1.5 times larger than that under the AR1 hypothesis. This broadens the range of tropical latitudes over which the trend is not significant at the 95% level, and substantially increases the already large trend uncertainty in northern middle and high latitudes. In contrast, the Hurst exponent is about 0.5-0.6 over southern middle and high latitudes, i.e. the residuals have relatively weak long-range correlations, and in this region the trend uncertainties estimated under the LRC and AR1 hypotheses are nearly identical.

Fig. 3.8b shows the corresponding results for the linear trend from 1979-1995 (the declining part of the PWLT function). The means and standard deviations (including the differences between the latter for LRC and AR1 hypotheses for the residuals) are very similar to those obtained using the EESC fit in Fig. 3.8a, except in northern middle and high latitudes where the PWLT-derived trend is larger. This is consistent with the behaviour already noted in section 3.3.3, where the strong decline in total ozone in northern middle and high latitudes in the

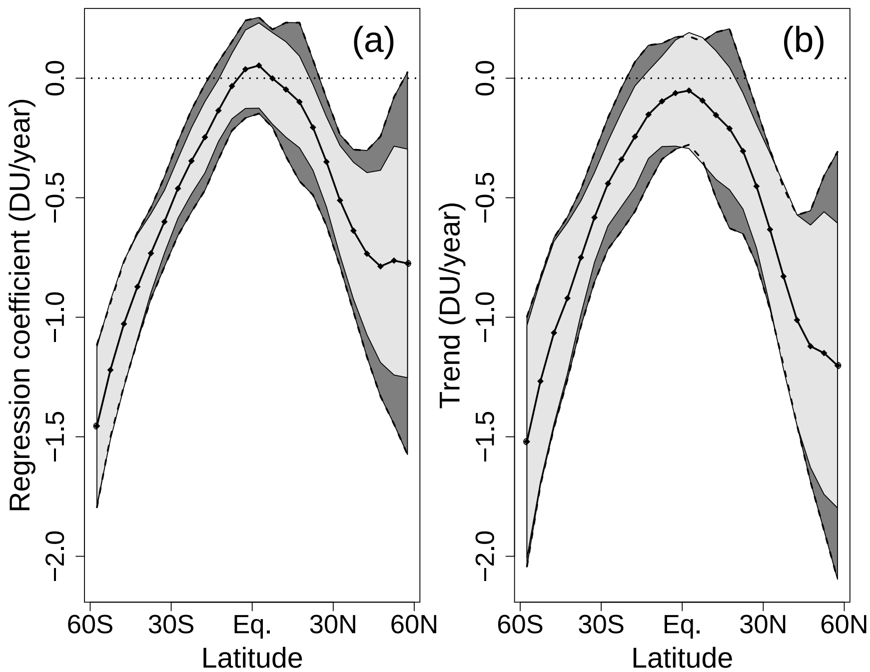

Figure 3.8: (a) Regression coefficients of monthly and zonal mean total ozone anomalies (obtained by filtering out the seasonal cycle, QBO, and solar flux) on EESC for the period 1979-2005. (b) The first (declining) slope of the PWLT fit for the period 1979-1995. The regression coefficients in panel (a) are scaled so as to be comparable to the linear trend over 1979-1995; thus, the two panels represent respectively the EESC-based and PWLT-based estimates of the long-term ozone decline. The 95% confidence intervals shown are calculated under two alternative assumptions: AR1 (light grey region) and LRC (dark grey region bordered by the dashed curves.). Details of the confidence-interval calculations can be found in Section 2.4.

early 1990s and its subsequent increase in the late 1990s is interpreted as LRC noise relative to the EESC time series, but contributes to the long-term decline (with weaker LRC behaviour in the noise) under PWLT.

3.4.2 Recent and future ozone increase

The EESC time series can be well approximated by two linear functions, with the first slope equal to about 1 ppb/decade for the period before the EESC maximum in the second half of the 1990s and the second slope equal to about -0.34 ppb/decade for the period after the EESC maximum.

Therefore it is possible to estimate the expected rate of ozone increase after the late 1990s from the EESC fit: it is just the regression coefficient plotted in Fig. 3.8a multiplied by -0.34. The result is shown in Fig. 3.9 by the dotted-dashed curve. For comparison, the positive trend estimated from PWLT, which is the observed linear trend over the time period 1996-2005, is shown by the diamonds connected by the solid curve together with its uncertainties under both the AR1 and LRC hypotheses. The two trends are fairly similar in southern middle and high latitudes, although the uncertainties on the observed trends encompass zero. In northern middle and high latitudes, however, the observed linear trend is roughly four times the EESC-predicted trend, and is actually statistically significant over 40°-50°N according to the PWLT estimate of the noise. Thus, once again in northern middle and high latitudes we have a major difference between the analysis provided by the EESC and PWLT-based models, although this difference is within the 95% error bars.

Once the analytical relation between the trend uncertainty and the length of the time series is known, it is possible to estimate the number of years required to detect a certain trend with a given error, and its dependence on location. This is important from a practical point of view for designing an ozone monitoring strategy. The number of years required to detect future ozone trends was studied by (Weatherhead and Coauthors, 2000) using the AR1 model. Here we expand on Weatherhead et al.'s results by including an allowance for LRC behaviour. The methods and formulas used here to calculate the number of years are described in Section 2.4.

We first consider the number of years required to detect a trend of a given magnitude, without reference to the magnitude of the expected trend. With the noise estimated relative to the EESC trend function (as in Fig. 3.4), the number of years required to detect a 1 DU/year trend

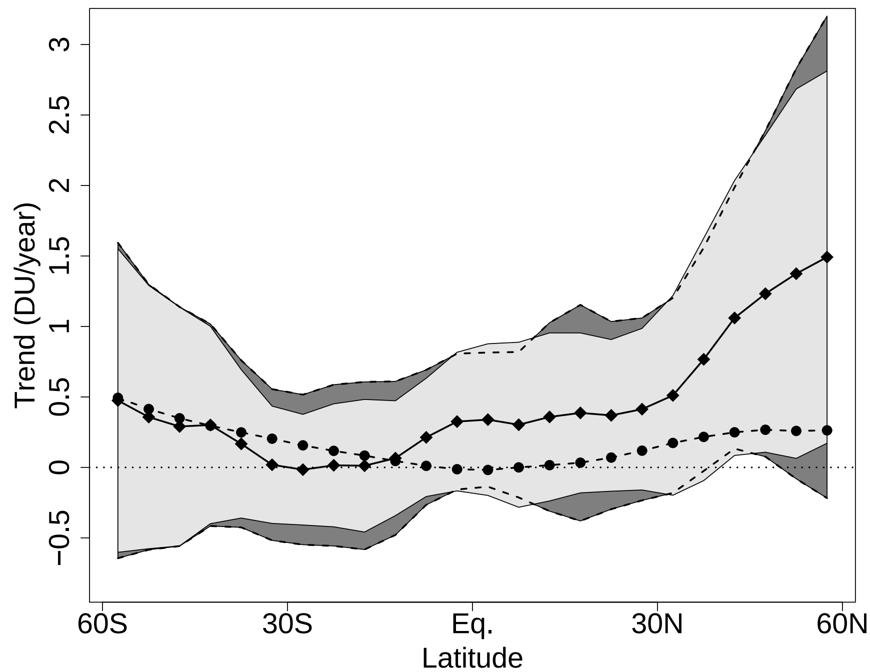

Figure 3.9: The EESC-based linear trend calculated for the declining part of EESC (solid circles connected by the dashed line) is compared with the second (increasing) slope of the PWLT fit for the period 1996-2005 (diamonds connected by the solid line), the latter with 95% confidence intervals calculated under the AR1 (light grey region) and LRC assumptions (dark grey region bordered by the dashed curves).

in zonal mean total ozone at the 95% significance level under the AR1 and the LRC hypotheses is shown in Fig. 3.10a. The latitudinal structure primarily reflects that of the variability (cf. Fig. 3.4a), with the shortest number of years being required in the tropics (30°S-30°N). However the impact of long-term memory (cf. Fig. 3.4b) mainly accounts for the hemispheric asymmetry in Fig. 3.10a, increasing the number of years required in northern as compared with southern latitudes. In those latitude bands for which the Hurst exponents are below 0.7, both the AR1 and LRC models give consistent estimates of the number of years required; whereas in other latitude bands, and especially in northern subtropical and high subpolar latitudes, LRC behaviour considerably lengthens the time required to detect a given trend, by a factor of up to

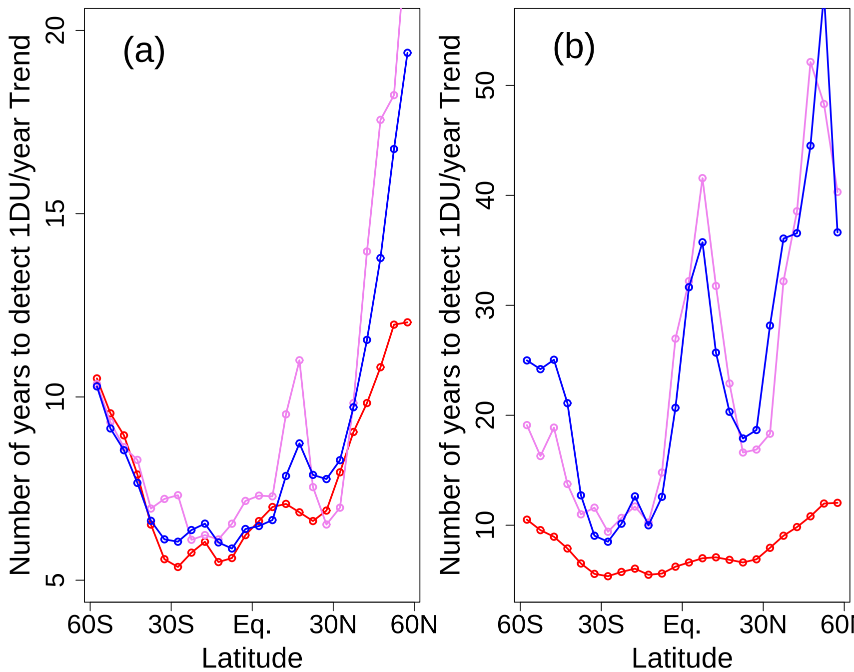

Figure 3.10: (a) The number of years required to detect a 1 DU/year trend at the 95% significance level under the AR1 (red curve) and LRC assumptions (violet curve shows GPHE, blue curve shows GSPE) applied to the frequency bandwidth 1-27 years of the monthly mean total ozone residuals obtained by filtering out the seasonal cycle, QBO, solar flux, and EESC fit. (b) The same as panel (a), but using the frequency bandwidth 2 months-27 years for the LRC estimates (the red curve is the same in both panels). Note the different vertical scales in the two panels. Panel (b) is a spurious result (cf. Fig. 3.5), and is shown to highlight the importance of choosing an appropriate bandwidth for the analysis.

1.5 or so.

If the noise is estimated according to the PWLT trend function (as in Fig. 3.6), then the number of years required to detect a 1 DU/year trend is virtually identical to that shown in Fig. 3.10a in the Southern Hemisphere, but is, as expected, reduced and closer to that estimated from the AR1 model in the Northern Hemisphere (not shown).

Fig. 3.10b shows the same estimates as in Fig. 3.10a, except that in estimating the statistical parameters b and H under the LRC hypothesis, GPHE and GSPE were applied to the frequency bandwidth from 2 months to 27 years (cf. Fig. 3.5). The number of years in this case is several times larger than for the proper bandwidth (1-27 years). We include Fig. 3.10b here to emphasize the importance of a correct bandwidth choice for Hurst exponent estimation under the LRC model.

We now consider the latitudinal dependence of the expected trends, and estimate the number of years required to detect a statistically significant ozone increase if the positive trends are those estimated in Fig. 3.9 according to either the EESC or PWLT models. In both cases the “detection” is here made under the assumption that the trend is independent of the past trend (prior to 2000 for EESC, prior to 1996 for PWLT), and both AR1 and LRC estimates are computed. Consider first the detection of the positive ozone trend expected from the EESC decline, shown in Fig. 3.11a. As was noted by (Weatherhead and Coauthors, 2000), southern middle and high latitudes are the best places to detect ozone recovery according to the AR1 model; the same is seen to be true for the LRC model. In the Northern Hemisphere, there appears to be an optimal region for detection of ozone recovery around 30°-40°N; on either side, there is a strong effect of LRC behaviour on the number of years required, especially at northern middle and high latitudes where, according to the LRC model, it should take about 1.5 times longer to detect the expected trends than estimated under the AR1 model.

The number of years required to detect the observed trend from 1996-2005 at the 95% significance level, according to the PWLT analysis, is shown in Fig. 3.11b. The length of the observed record (10 years) is indicated by the dotted line; at latitudes with points lying below this line, a significant trend has therefore already been detected (cf. Fig. 3.9b). The result is completely different from Fig. 3.11a. According to the PWLT analysis, the best place to detect ozone recovery is northern middle and high latitudes - moreover, in this region a positive trend is either on the verge of being detected or has already been detected - and the second best region is in the equatorial zone. However at southern high latitudes the number of years required, ~ 18, is similar between the EESC and PWLT analyses, and is well estimated

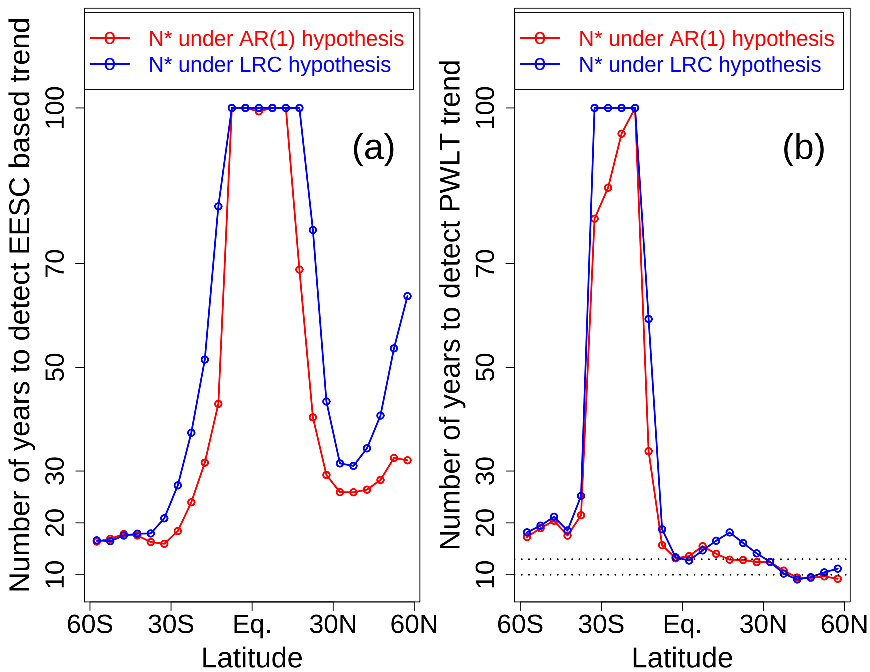

Figure 3.11: (a) The number of years since 2000 required to detect the EESC-based linear trend calculated for the declining part of EESC at the 95% significance level under the two alternative assumptions: AR1 (red curve) and LRC (blue curve, based on GSPE). (b) The same as panel (a), but for the PWLT-based trend calculated for the period 1996-2005, so the value represents the number of years after 1996. Values higher than 100 years are plotted as 100 years. Note the similarity of the two estimates in southern middle and high latitudes, but the large differences in the Northern Hemisphere.

by the AR1 model in both cases. This is expected given the consistency at these latitudes of the EESC-predicted and observed recent trends, the consistency of the EESC and PWLT derived noise estimates, and the absence of LRC.

3.5 Longitudinal structure

In this section we present the latitude-longitude distributions of some of the statistical parameters discussed above for zonal averages. Fig. 3.12a and b show the spatial distributions of

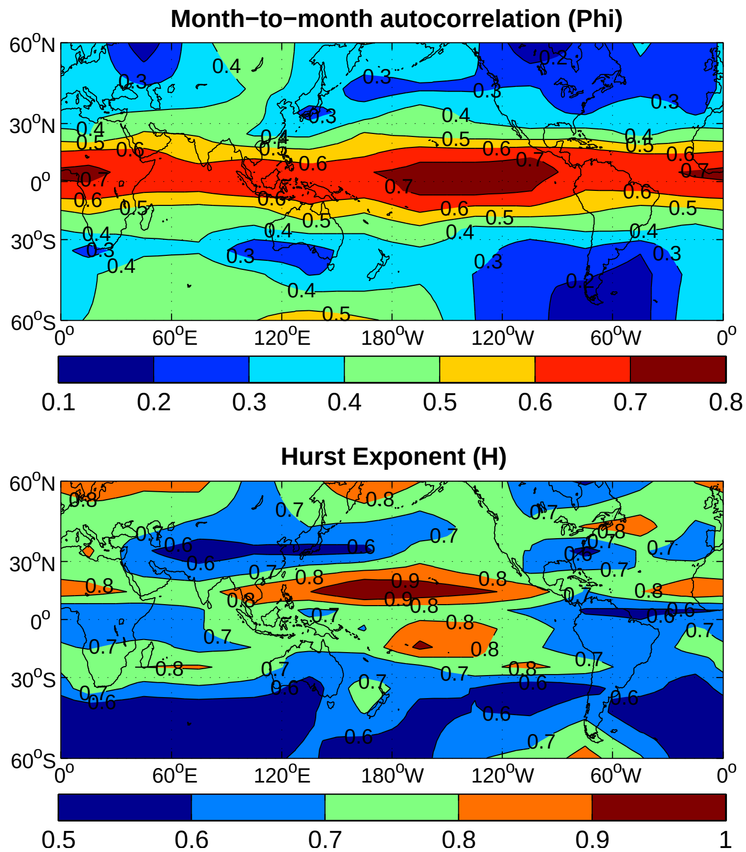

Figure 3.12: Spatial distribution of the month-to-month autocorrelation parameter ϕ (top panel) and the Hurst exponent (bottom panel) calculated by GSPE for monthly mean total ozone residuals obtained by filtering out the seasonal cycle, QBO, solar flux, and EESC fit. This is the two-dimensional version of Fig. 3.4b.

the AR1 month-to-month autocorrelation parameter ϕ and the Hurst exponent H, respectively, when the trend is defined by the EESC function. Both ϕ and H reflect "memory" in the total ozone time series but there is a distinct difference in their spatial structure. The autocorrelation parameter has its maximum at the equator and decreases towards the poles; the values are very similar to those reported by (Weatherhead and Coauthors, 2000) (see their Fig. 3.4 and Plate 3) for 1979-1998. In contrast, the spatial distribution of the Hurst exponent has two maxima, around 15°S over the Pacific and 20°N over the Pacific, eastern Atlantic, and Africa. For both ϕ and H, the maxima have pronounced longitudinal structure. The values of H for zonal-mean

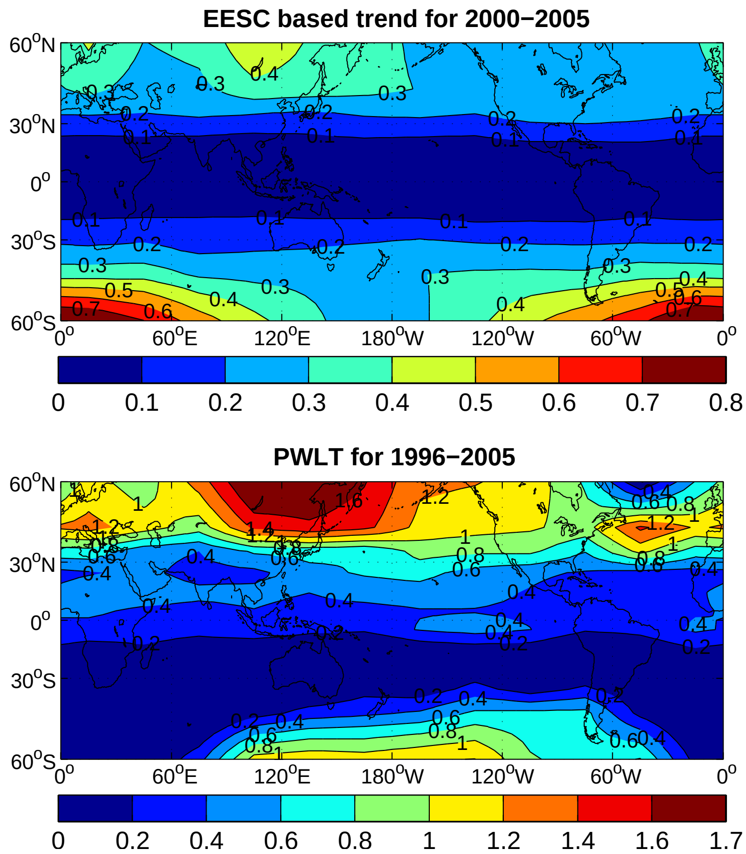

Figure 3.13: Spatial distribution of the EESC-based linear trend calculated for the declining part of EESC in DU/year (top panel) and the second (increasing) slope of the PWLT fit for the period 1996-2005 (bottom panel). This is the two-dimensional version of Fig. 3.9.

ozone (Fig. 3.4b) are approximately equal to the maximum values of H for gridded ozone at the given latitude. We will discuss the relationship between the zonally averaged H and the H for zonal averages in Chapter 4 in relation to temperature. The spatial distribution of the Hurst exponent when the PWLT is filtered out instead of EESC (not shown) is similar to Fig. 3.12b, except the values are lower in the Northern Hemisphere.

Fig. 3.13a and b show the spatial distributions of the recent trends according to the EESC and PWLT functions, respectively; they correspond to the curves in Fig. 3.9 for zonal-mean ozone. In the case of EESC, this represents the positive linear trend expected since the late 1990s based on the fitting of the entire 1979-2005 record by the EESC time series. Based

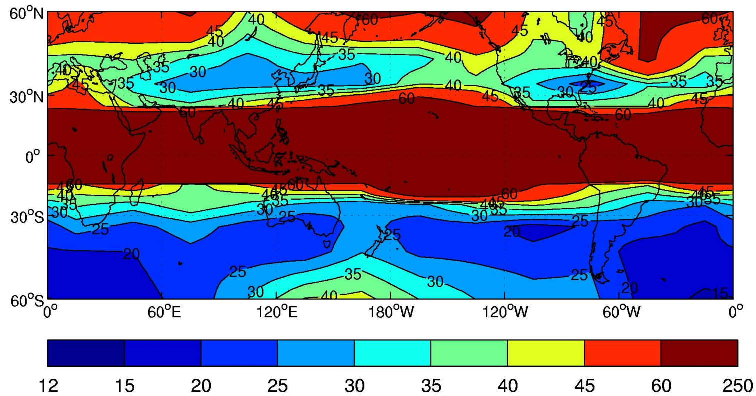

Figure 3.14: The number of years since 2000 required to detect the EESC-based linear trend calculated for the declining part of EESC at the 95% significance level under the LRC assumption (based on GSPE). This is the two-dimensional version of Fig. 3.11a.

on EESC, the ozone recovery rate should be strongest in the Southern Hemisphere sub-polar regions, apart from the south-west Pacific. In the Northern Hemisphere the recovery rate is predicted to be relatively strong over northern Europe and over eastern Siberia. In the tropics and subtropics the expected recovery rate is weak and very zonal. As can be anticipated from Fig. 3.9, the ozone recovery rates based on PWLT (i.e. the observed linear trend from 1996-2005) are also positive everywhere but have a very different spatial distribution and strength. In particular, the PWLT recovery rates are greater in the Northern Hemisphere midlatitudes and sub-polar regions than in the Southern Hemisphere; the trends are especially strong over Siberia, the north Pacific, the subtropical Atlantic, and southern Europe. The only place in the Northern Hemisphere midlatitudes and sub-polar regions where the trends are relatively weak is the north Atlantic. The spatial distribution of the recent PWLT over the Southern Ocean is opposite to that for the EESC trend, with a maximum rather than a minimum over the south-west Pacific.

Finally, Fig. 3.14 presents the number of years from the year 2000 required to detect the expected EESC-based ozone trends shown in Fig. 3.13a, based on the LRC noise estimates computed from the entire time series (with the EESC trend filtered out). The Fig. 3.14 allows us to identify the optimal locations to make long-term ground-based total ozone observations. For example, it would be desirable to have some stations in the southern sub-polar Atlantic since the number of years required to detect ozone recovery has a minimum in that region, where it varies between 12 and 20 years. In the Northern Hemisphere the minimum is located in the zonal band around 35°N, and varies between 20 and 30 years.

3.6 Summary and discussion

The statistical analysis of long-term changes in total ozone has traditionally been performed assuming that the residuals, which represent the noise in the system, are well described by an AR1 model. In this chapter the total ozone record from 1979-2005 has been examined from the alternative viewpoint that the time series is long-range correlated, implying a deviation from AR1 behaviour with an unbounded decorrelation time. The existence of LRC behaviour in total ozone would reduce the statistical significance of a given trend, and lengthen the number of years required to detect a trend, from that estimated using an AR1 model. We employ the merged satellite data set prepared by NASA which combines version 8 of TOMS and SBUV total ozone data (Frith et al., 2004; Stolarski and Frith, 2006), use well-based spectral estimation techniques to quantify LRC paying proper attention to the frequency bandwidth, and filter long-term time-periodic signals (QBO, solar) which can give spurious indications of LRC behaviour. The analysis mainly concerns zonal-mean ozone, although some station data and gridded satellite data are also considered. However the analysis is restricted to 60°S-60°N, as in polar regions the satellite data have gaps during polar night.

We first summarize the results obtained when the long-term total ozone changes are represented in terms of the EESC time series. The large values of the Hurst exponent, which is the necessary condition for the presence of long-range correlations, are found basically everywhere north of 35°S. In southern middle and high latitudes the correlation behaviour is not significantly different (at the 95% confidence level) from that of the AR1 model. In the regions with strong LRC behaviour, uncertainties in the magnitude of the long-term ozone decline attributable to EESC are increased by about a factor of 1.5 compared with those estimated from AR1; this includes northern middle and high latitudes, where the AR1-based uncertainties are already quite large. However the strongest long-term ozone decline is found at southern middle and high latitudes, and there the AR1 estimates are found to be reliable.

Analogous results are found for the number of years (from 2000) required to detect the increase of ozone expected from the anticipated decline of EESC. We confirm (Weatherhead and Coauthors, 2000)'s finding, based on the AR1 model, that southern middle and high latitudes should be the optimal place (within the 60°S-60°N region) to detect ozone increase; at these latitudes we have the combination of the strongest expected trend, the apparent absence of LRC behaviour, and the shortest autocorrelation times. The required detection time (to 95% confidence) is about 18 years for zonal-mean ozone at 60°S, but is even a few years shorter in the sub-polar south Atlantic. (While limited regions have higher noise levels, they also have weaker serial correlations.) The recent observed behaviour of total ozone in these regions is consistent with the EESC-predicted trend, but detection of an ozone increase attributable to EESC is not expected until sometime late in the decade 2010-2020. In the Northern Hemisphere, detection of ozone increase is more challenging. There appears to be a narrow band around 35°N where LRC behaviour is relatively weak and the required number of years is around 30, but in northern middle and high latitudes the required number of years is increased from around 25-35 to around 30-60 by LRC.

Although the representation of long-term ozone changes in terms of the EESC time series is preferred, given the a priori nature of the representation, a commonly used alternative is a piecewise-linear trend (PWLT) with a turning point in the second half of the 1990s. Therefore we compared the results obtained using the two different representations of the long-term changes. In our implementation of PWLT we use a turning point in early 1996. The estimates of the noise and the long-term ozone decline are essentially the same for the two cases in the Southern Hemisphere, but there is a notable discrepancy in the Northern Hemisphere (particularly at northern middle and high latitudes) where the strong decrease in ozone in the early 1990s, and its subsequent increase in the late 1990s, are interpreted mainly as LRC noise relative to EESC, but project strongly on the long-term changes (thereby reducing the strength of the LRC behaviour) relative to the PWLT. This difference affects all subsequent estimates. For example, according to PWLT the long-term ozone decline in northern middle and high (sub-polar) latitudes is comparable in magnitude to that in the Southern Hemisphere; and the recent ozone increase (since 1996) is strongest in this region, and marginally statistically significant (at the 95% confidence level) indicating that a positive ozone trend is already on the verge of being detected.

The natural question is, then, which representation of the long-term changes (and thus of the noise) is correct? We do not attempt to answer this question definitively, but a few comments may be in order. If one adopts the EESC perspective, then the results seem physically sensible: we know that the annual-mean long-term ozone decline, from pre-1980 levels to those characteristic of the 2000 time period, over middle and high (sub-polar) latitudes has been much greater in the Southern as compared with the Northern Hemisphere - roughly 6% as compared with 3% (World Meteorological Organization, 2002). Furthermore we know that Northern Hemisphere ozone exhibits more interannual variability than Southern Hemisphere ozone because of the greater stratospheric dynamical variability in the Northern Hemisphere, which is for well-understood reasons. What remains then to be understood is the physical origin of the LRC, especially in northern middle and high latitudes. If it is the existence of AR time scales comparable to the 27-year observational record, then are these time scales associated with natural variability or with climate change? These questions can likely only be answered with climate models.

If, on the other hand, one adopts the PWLT perspective, then one is forced to consider the strong decline of northern middle and high latitude ozone in the early 1990s, and its subsequent increase in the late 1990s, as part of the signal and account for it. One possibility often considered (e.g. Solomon et al., 1996) is that the increased stratospheric aerosol from the Mount

Pinatubo volcanic eruption in 1991 amplified the EESC-associated ozone loss. The problem with this argument is that there was no corresponding ozone decrease observed in the Southern Hemisphere, even though EESC and aerosol abundances were comparable (Bodeker et al., 2001). Another argument is that the behaviour reflects decadal-scale variations in stratospheric wave forcing (e.g. Randel et al., 2002; Hadjinicolaou et al., 2005), which would affect ozone both through changes in transport and changes in chemical ozone loss, especially in the Arctic which would then affect the annual mean sub-polar ozone abundances through transport of ozone-depleted air. The impact of long-term changes in stratospheric wave forcing on both polar and midlatitude ozone is well established (World Meteorological Organization, 2002). However, attributing the ozone changes to changes in wave forcing merely changes the problem to that of accounting for the variations in wave forcing. In principle, they could be part of the signal or part of the noise. Yet the use of PWLT involves the implicit assumption that the recent strong positive trend in northern middle and high latitude ozone is secular and can be extrapolated; moreover by regarding this trend as part of the signal rather than part of the noise, the estimated noise is reduced and the LRC behaviour weakened, and the estimated significance of the trend thereby increased. So far, no mechanism that could give such a statistically significant positive trend in northern middle and high latitude ozone has been put forward.

3.7 Appendix A: Comparison with ground based measurements for 1979-2008

The analysis of total ozone was repeated for the period 1979-2008 for the merged TOMS/SBUV/OMI data and for the ground based measurements (Fioletov et al., 2002). In comparison to the 1996-2005 period the PWLT for 1996-2008 became significant everywhere in the Northern Hemisphere relative to the AR1 confidence intervals, although it dropped in magnitude north of 40°N and thus became somewhat closer to the EESC trend (see Fig. 3.9,3.15,3.16). The EESC trend for the merged satellite data is similar to the EESC

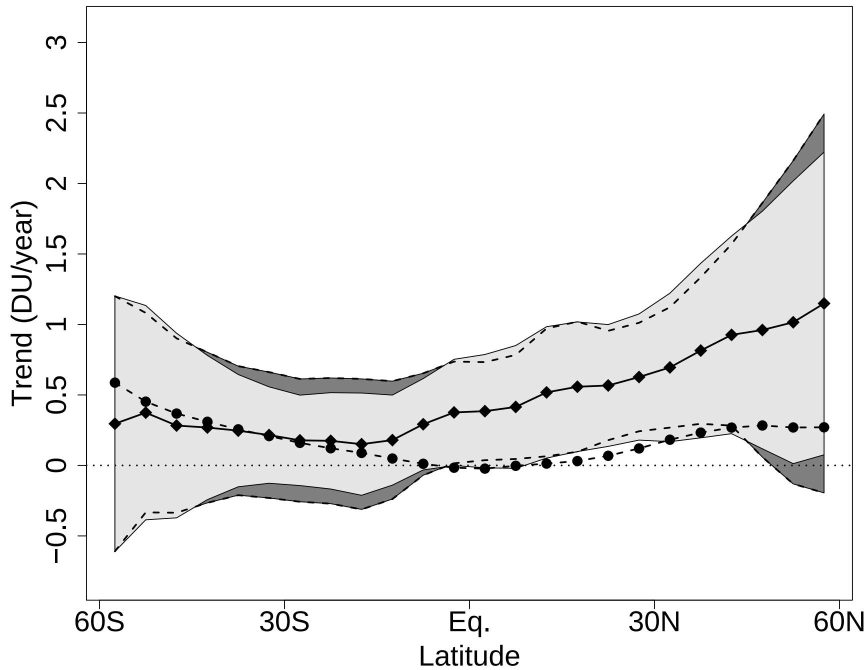

Figure 3.15: Same as Fig. 3.9, but for the period 1996-2008 using the TOMS/SBUV/OMI merged data set.

trend for the ground based measurements (see Fig. 3.15,3.16). In contrast the PWLT disagrees in the two data sets in the Southern Hemisphere, which emphasizes its sensitivity to noise.

The monthly autocorrelations are noticeably lower for the ground based than for the satellite merged data, whereas the Hurst exponents are more similar (see Fig. 3.17b,3.18b), though slightly larger for the former. This probably means that the low-frequency residual variability, represented by the Hurst exponent, is more consistent in the two data sets than the high-frequency variability, represented by the monthly autocorrelations. Trends and autocorrelations demonstrate more zonal variability in the ground based data. The standard deviations are noticeably larger for the ground based measurements (see Fig. 3.17a,3.18a), which leads to larger confidence intervals for the PWLT (see Fig. 3.15,3.16). We tend to think that these differences are presumably caused by data inhomogeneities in the ground based measurements.

One can also notice that the Hurst exponent estimates, especially for the PWLT residuals,

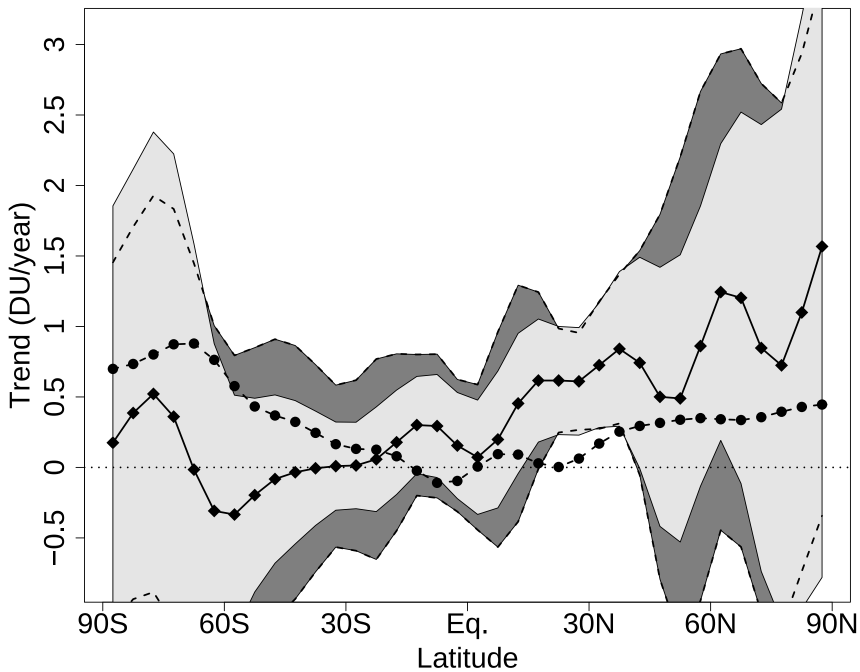

Figure 3.16: Same as Fig. 3.9, but for the period 1996-2008 using the ground based measurements. Please note that the latitudes extent from the South to the North Pole.

fall below 0.5 over the polar regions (see Fig. 3.18), which means that the corresponding time series demonstrate blue noise behaviour, i.e. their spectral power decreases with decreasing frequency. As a result the power-law confidence intervals over the poles are less than the AR1 confidence intervals. It is interesting what physical mechanism stands behind this phenomenon.

Overall the three additional years (2006-2008) of the merged satellite data and the independent ground based measurements support our conclusions based on the TOMS/SBUV merged data for the 1979-2005 period.

3.8 Appendix B: Analysis of Kiss et al. results

(Kiss et al., 2007) (K07 henceforth) analyzed long-range temporal correlations of total ozone measured by the TOMS instrument. K07 estimated the Hurst exponent for the total ozone time series by means of DFA3. Thus the Hurst exponent was estimated as the slope of a

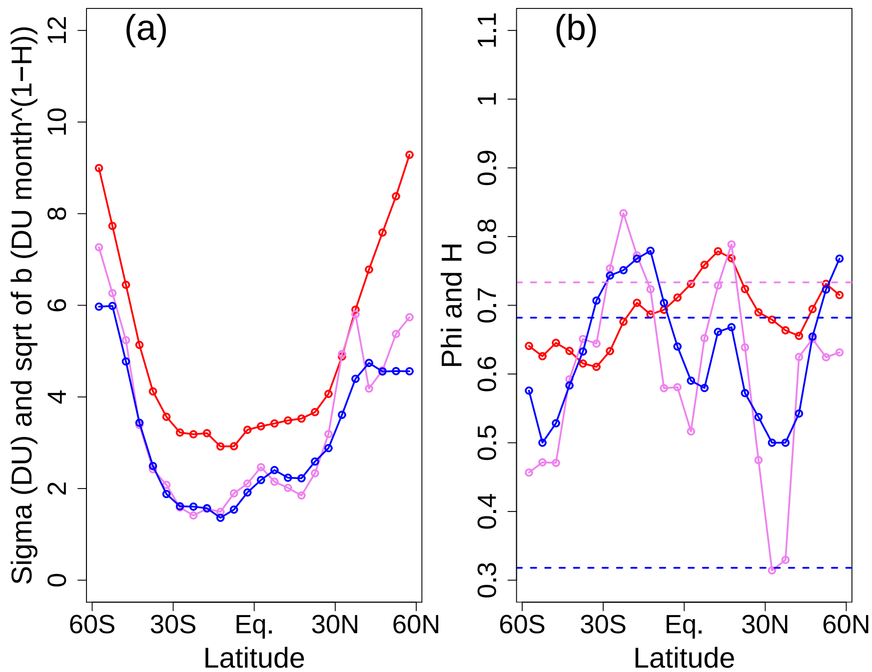

Figure 3.17: Same as Fig. 3.6, but for the period 1979-2008 using the TOMS/SBUV/OMI merged data set.

DFA curve in log-log coordinates, which was measured for the range of scales from about a month to 6 years. The Hurst exponent was estimated for each grid point time series of the gridded TOMS measurements (0°-360°E, 60°S-60N°) and then zonally averaged. The obtained latitudinal distribution of the Hurst exponent estimate has local maxima over high latitudes and the equator and local minima over midlatitudes. This distribution was compared by K07 with the distribution of the Hurst exponent estimated in this chapter and published in Vyushin et al. (2007) (below in this appendix we will refer to the current chapter as V07 for short) for the merged zonally averaged TOMS/SBUV total ozone and found to be different.

In this note we explain the difference in the two distributions. Seven differences in the ways the exponents were estimated are: (a) only TOMS data were utilized by K07, while V07 employed the merged TOMS/SBUV data set; (b) daily data was used by K07 vs. monthly data used by V07; (c) DFA3 was used by K07 vs. the GPHE and GSPE used by V07; (d)

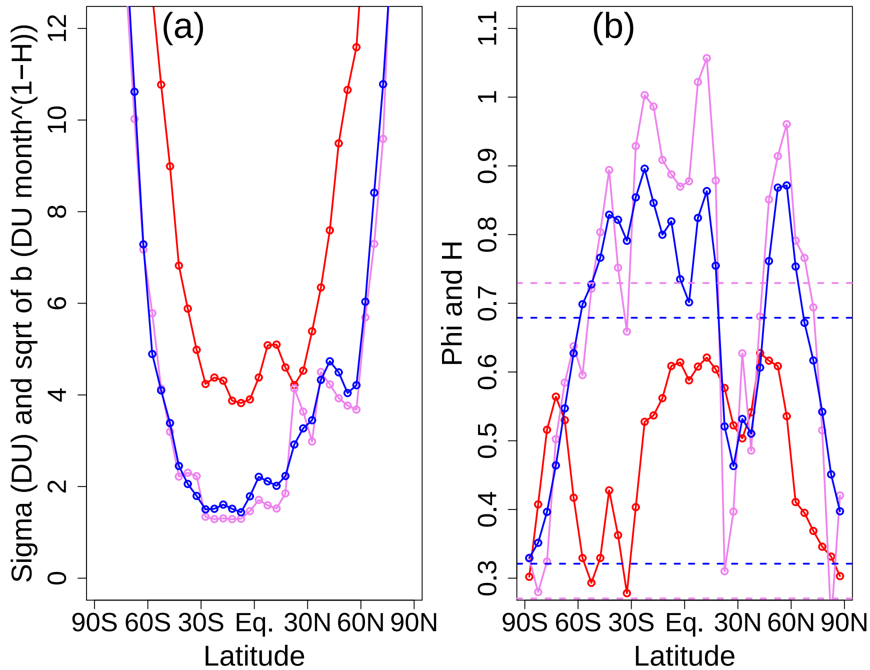

Figure 3.18: Same as Fig. 3.6, but for the period 1979-2008 using the ground based measurements. Please note that the latitudes extent from the South to the North Pole.

only annual cycle and QBO were filtered out by K07 vs. annual cycle, QBO, solar cycle and the EESC trend by V07; (e) scales from about a month to 6 years were used by K07 vs. the frequency range corresponding to 1 to 27 years by V07; (f) V07 reported results for the zonally averaged total ozone, while K07 reported the zonally averaged Hurst exponents for the gridded total ozone; (g) the QBO was filtered by linear regression in V07 and by the Wiener filter in K07. We demonstrate here that the last four differences are key in explaining the different results reported.

First, we checked that the results for monthly TOMS/SBUV data are the same as for TOMS only data. Therefore we excluded item (a). Item (b) is irrelevant in these studies, because sub-monthly frequencies are not used. Providing time series are power-law stochastic processes not contaminated by the presence of trends or periodicities, DFA and spectral methods (GPHE and GSPE) should give close estimates of the Hurst exponent as was shown in Section 2.3.

However climatic and meteorological time series do not usually have pure power-law spectra. Often their spectra have different slopes for high and low frequency ranges (see e.g. Fig 3.2). Moreover, they may contain periodic and quasi-periodic signals of various periods, secular trends as well as data inhomogeneities, caused for instance by changes in instrumentation, etc. In such circumstances an estimate of the Hurst exponent depends on the filtering applied to the time series and the choice of frequency (time scale) range (see Jánosi and Müller (2005); Marković and Koch (2005) and Chapter 4). It will be shown in Chapter 4, using tropospheric and stratospheric air temperature as an example, that DFA and spectral methods give similar estimates provided equal time scales and frequency ranges including the lowest available frequencies are chosen and that trends and periodic and quasi-periodic signals are filtered out. Below we show that these principles are also applicable to the total ozone.

Fig. 3.19 shows estimates of the Hurst exponent for the TOMS/SBUV zonally average total ozone anomalies estimated by DFA, GPHE, and GSPE for different filters and time scales combinations. We start with filtering the components which were filtered in K07, namely the annual cycle and the QBO. In panel (a) the annual cycle and the QBO have been filtered out using linear regression on four annual cycle harmonics and the equatorial zonally averaged zonal winds at 30 and 50 hPa (see Section 3.3.1 for the linear regression details) and the Hurst exponents are estimated for the time scales from one to six years (the intersection of the time scale ranges used in V07 and K07). The Hurst exponents estimated by DFA1, GPHE, and GSPE do not agree in this case. At the next stage (see Fig. 3.19b) we have also filtered out the solar cycle using the solar flux at 10.7 cm and the EESC trend in addition to the annual cycle and the QBO. This brings the three curves closer together, but there are still noticeable differences between them. Comparison of Fig. 3.19a and b reveals the effect of filtering of the solar cycle and the EESC trend, i.e. the effect of item (d).

At the third stage to fully comply with the recommendations of Chapter 2 we extend the time scale range up to 27 years and thus include the lowest available frequencies. Most of the articles, which employed DFA, set the maximum used time scale equal to a quarter of

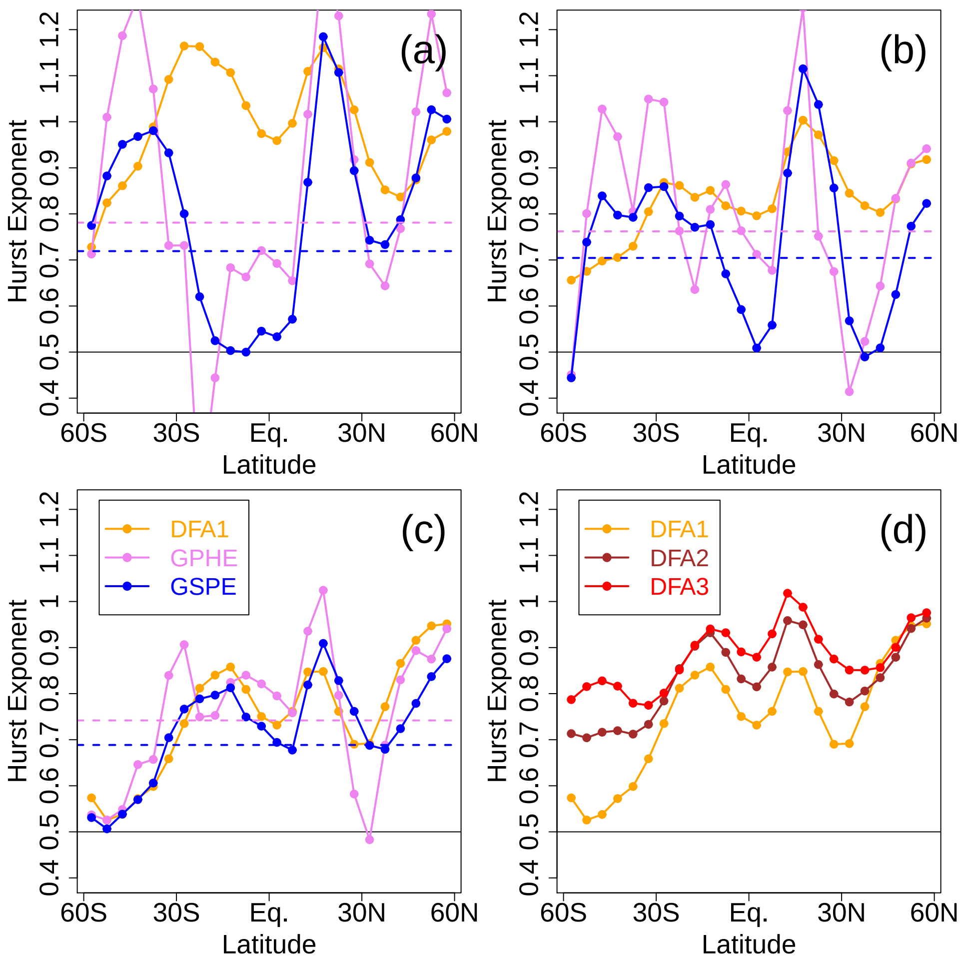

Figure 3.19: Estimates of the Hurst exponent for zonally averaged TOMS/SBUV total ozone anomalies by DFA, GPHE (violet curve) and GSPE (blue curve). Only the annual cycle and the QBO have been filtered out by means of linear regression in panel (a). In all other panels annual cycle + QBO + solar flux + EESC have been filtered out. The time scales from 1 to 6 years have been used in panels (a) and (b) and from 1 to 27 years in panels (c) and (d). The DFA1 results are shown by the orange curves. In panel (d) we also show the results of DFA2 (brown curve) and DFA3 (red curve). The horizontal dashed lines in panels (a-c) show the upper 95% asymptotic confidence intervals for GPHE and GSPE for the Hurst exponent equal to 1/2, which corresponds to stochastic processes with white noise like low-frequency variability.

time series length. This limitation was eliminated in Chapter 2 demonstrating by means of Monte-Carlo simulations that the properties (bias and variance) of the Hurst exponent estimate obtained by DFA do not almost change when the longest available time scales are included. This actually makes DFA based estimates comparable to periodogram based results, for which the lowest available frequency are typically included by default. The inclusion of the longest available time scales is also consistent with the fact that the Hurst exponent is defined only asymptotically (see e.g. Taqqu, 2002). Another motivation for the inclusion of the lowest available frequencies is that for the estimation of trend uncertainty (the trend is supposed to be filtered out prior to the Hurst exponent estimation) only a low-frequency behaviour of the power spectrum matters (Smith, 1993). Fig. 3.19c shows that by the inclusion of the lowest available frequencies we reach a very close agreement between the different methods estimates. The GPHE and GSPE Hurst exponent estimates (the violet and blue curves) in Fig. 3.19c are the same as in Fig. 3.4b, which K07 used for comparison. Comparison of Fig. 3.19b and c reveals the effect of item (e).

In the first three panels of Fig. 3.19 we employed DFA of the first order (DFA1), because it can automatically filter out only discontinuities in time series, but not trends, and therefore is more similar to spectral methods than DFA of the higher orders. K07 used DFA3, which automatically filters out local quadratic trends, and it is important to compare its results with the results of DFA1. We plot in Fig. 3.19d the Hurst exponent estimates obtained by DFA1-3 after the annual cycle + QBO + solar flux + EESC have been filtered out and the time scales from 1 to 27 years have been used. With the exception of the Southern Hemisphere middle and high latitudes the DFA results have qualitatively similar distribution with generally larger estimates obtained by higher orders of the DFA. We will explain the differences between the DFA results below in relation to Fig. 3.20.

Comparing panels (b) and (c) of Fig. 3.19 one can notice that the shift of the low frequency cutoff from 6 to 27 years has decreased the value of the Hurst exponent estimates over the Southern Hemisphere middle and high latitudes. Let us now take a careful look at the power

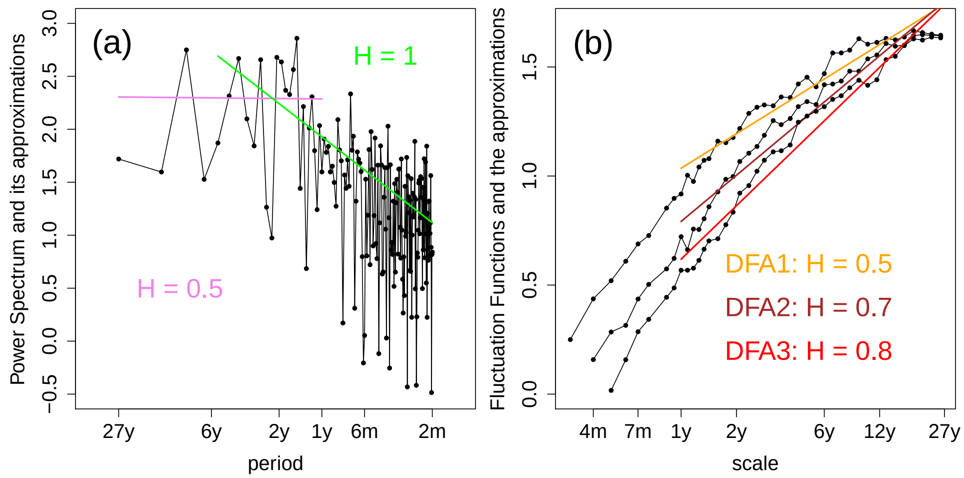

Figure 3.20: Periodogram (panel (a)) and DFA1-3 fluctuation functions (panel (b)) of the TOMS/SBUV total ozone anomalies obtained by filtering out annual cycle + QBO + solar flux + EESC from the 50°-55°S zonal band. The periodogram, fluctuation functions, and their power-law approximations are shown in log-log coordinates. The estimated values of H are rounded to one digit.

spectrum and DFA curves (fluctuation functions) of the total ozone anomalies for the 50°-55°S zonal band. They are plotted in Fig. 3.20. There are two scaling regimes in this power spectrum (see panel (a)). The first is a high frequency one. It ranges from 2 months to somewhere between one and two years. The second, low frequency regime, ranges from about two years to the lowest frequency. If one fits a power-law curve to the frequency range from 2 months to 6 years, which is the scaling range used in K07, then one obtains a Hurst exponent equal to about unity, as illustrated by the green line. In contrast, if a power-law is fitted to the frequency range from 1 to 27 years, as illustrated by the violet line and as was done in V07, then the estimated Hurst exponent is about one half. This explains the difference seen at 50°-55°S during the transition from Fig. 3.19b Fig. 3.19c (the effect of item (e)). The sensitivity of the Hurst exponent estimates to the choice of the frequency range was actually stressed in V07 by contrasting Fig. 3.4 and 3.5.

Fig. 3.20b shows fluctuation functions and their best linear fits in log-log coordinates for

DFA of the first, second, and third order. In agreement with Fig. 3.19d the Hurst exponent estimate increases from 0.5 to 0.8 as one increases the DFA order from the first to the third. It is known that even for a power-law stochastic process DFA curves have two regimes: short scale and long scale (Kantelhardt et al., 2001). The Hurst exponent should be estimated by fitting a power-law function to the long time scales. It is also known that the transition (crossover) point between the two regimes depends on the order of DFA. The higher the order, the larger the transition point (Kantelhardt et al., 2001). This phenomenon can be observed in Fig. 3.20b. The transition point for DFA1 is located around a one year time scale, whereas for DFA3 it is close to two years. Thus using the same time scale range (1-27 years) for DFA1 and DFA3 the Hurst exponent is overestimated by the inclusion of the short scale range regime into the estimation domain for DFA3. We conclude that K07 obtained significantly higher Hurst exponent estimates over the Southern Hemisphere high latitudes, because they used the time scale range located in high frequencies and the third order of DFA. Both of these facts lead to an overestimation of the true Hurst exponent. This paragraph also underlines that care is needed when estimation of the Hurst exponent is performed and that none of the estimation methods should be used in isolation.

The discussion above only partially explains the differences in the shape of the Hurst exponent distributions between V07 and K07. Fig. 3.21 reveals the effect of items (f) and (g). Panel (a) of Fig. 3.21 shows the zonally averaged Hurst exponent obtained for the gridded TOMS/SBUV merged data set after the annual cycle + QBO + solar flux + EESC have been filtered out by linear regression. The time scales from one to six years have been used. Therefore this panel is analogous to panel (b) in Fig. 3.19. One can notice that qualitatively the spatial distributions of the Hurst exponent are somewhat similar in these panels. However the zonally averaged Hurst exponents are generally lower than the Hurst exponents for the zonal averages, in agreement with the theory of power-law stochastic processes (Granger, 1980) and the results for the atmospheric general circulation (see Chapter 4). This phenomenon was also discussed in V07 in respect to Fig. 3.2. Nevertheless K07 compared in their Fig. 7 the Hurst exponent

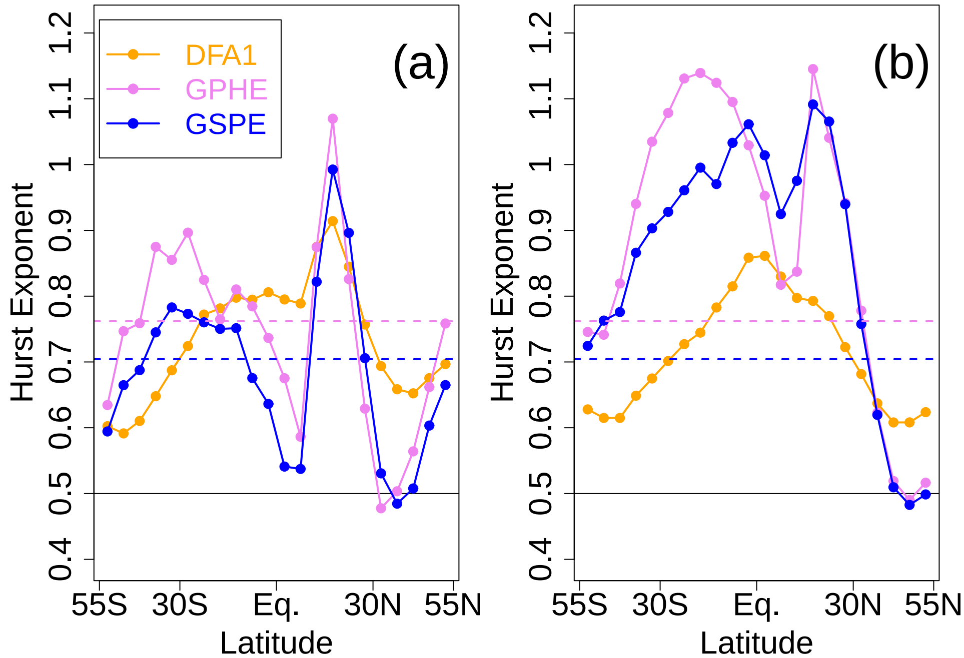

Figure 3.21: Zonal averages of the Hurst exponent estimates obtained for the gridded TOMS/SBUV data by DFA1 (orange), GPHE (violet), and GSPE (blue) after filtering the annual cycle + QBO + solar flux + EESC using the time scales from 1 to 6 years. In panel (a) the QBO has been filtered out by linear regression, whereas in panel (b) it has been filtered out by the Wiener filter as in K07.

estimates from V07 for the zonally averaged total ozone with their zonally averaged Hurst exponent estimates for the gridded data. When we recalculate the Hurst exponents plotted in Fig. 3.21a for the time scale range 1-27 years we get a picture very similar to Fig. 3.19c with somewhat smaller values but with even better agreement between the methods (not shown).

Fig. 3.21b is analogous to Fig. 3.21a, with the only difference is that the QBO has been filtered out using the Wiener filter following K07. K07 linearly interpolated the total ozone anomalies power spectrum for frequencies in the range from 1.1 to 4.3 years. When we apply this filtering method to monthly ozone data the results are significantly affected, as seen by comparing panels (a) and (b) of Fig. 3.21. Remarkably, DFA1 and the spectral methods start to significantly disagree over several regions when the linear regression filtering of the QBO used in V07 has been replaced by the Wiener filter used in K07. All the methods demonstrate a relative boost of the Hurst exponent estimates over the tropics. Although K07 utilized daily data and a wider estimation frequency range, for which the 1.1 to 4.3 years range comprises a smaller portion, their results still could be affected by this rough filtering method.

The major sources of differences between the results of V07 and K07 are the way the data were filtered and which frequency (time scale) ranges were chosen to measure a power-law exponent. V07 employed a low frequency range for two reasons. Firstly, because only a low frequency residual variability affects the uncertainty of a trend, a proper estimation of which was the main goal of V07. Secondly, the mathematical theory of power-law stochastic processes is developed primarily for the asymptotic case, i.e. when the spectral density (auto-correlation function) scales by a power-law for low frequencies (large time lags). K07 chose the intermediate range of scales mainly because of conventional DFA requirements. Therefore the comparison of V07 and K07 results is not appropriate.

Chapter 4

Power-law characteristics of the atmospheric general circulation

4.1 Introduction

In Chapter 3, we applied the two statistical noise models that are the focus of this thesis — the power-law and AR1 models —- to the problem of ozone trends and variability, which is an issue of significant practical interest. In this and the following chapters, we turn to the more general question of how best to characterize internal variability in the climate system. As we have just seen in the ozone context, the choice of noise model can strongly affect conclusions about trend analysis, and this is also the case for detection of periodic signals in climate data (Ghil et al., 2002). But the ozone analysis did not help us settle several of the questions we raised in the Introduction. The questions we focus in this Chapter regard the choice of power-law method and a preliminary attribution of observed spectral power growth to some specific processes.

In this chapter, several power-law exponent estimators are applied to global zonally averaged free atmosphere air temperature data from reanalysis products. The methods employed (detrended fluctuation analysis, Geweke Porter-Hudak estimator, Gaussian semiparametric estimator, and multitapered versions of the last two) agree well for pure power-law stochastic processes as was shown in Chapter 2. But for the observed temperature record, the power-law fits are sensitive to the choice of frequency range and the intrinsic filtering properties of the methods. The observational results converge once frequency ranges are made consistent and the lowest frequencies are included, and once several climate signals have been filtered. Two robust results emerge from the analysis for interannual and decadal time scales: first, that the tropical circulation features relatively large power-law exponents that connect to the zonal-mean extratropical circulation; and second, that the tropical and subtropical lower stratosphere exhibits power-law like behavior that is volcanically forced.

In Chapters 4-6 we employ monthly mean temperature from the ERA40 and NCEP/NCAR reanalyses. The NCEP/NCAR reanalysis is a continually updating gridded data set representing the state of the Earth's atmosphere, incorporating observations and global climate model output dating back to 1948. It is a joint product from the National Centers for Environmental Prediction and the National Center for Atmospheric Research. ERA40 is a similar product by the European Centre for Medium-Range Weather Forecasts, which covers the period from September 1957 to August 2002[4]. For consistency we use only data from this 45 year long period from both reanalyses. The NCEP/NCAR reanalysis has horizontal resolution of 1.9^ox1.9^o and 17 vertical levels between 1000 and 10hPa. ERA40 has horizontal resolution of 2.5^ox2.5^o and 18 vertical levels between 1000 and 10hPa. (ERA40 has five extra vertical levels, which go up to 1hPa, but we do not use them.)

This chapter represents an attempt to systematically fit spectral power growth of the atmospheric general circulation by the power-law. First, we apply the methods to air temperature from reanalysis products for the last half century, focusing on plots of zonal mean cross sections of the Hurst exponent (Section 4.2). Although no single power-law fit technique should be employed in isolation, we identify a pair of techniques that characterize the range of results

[4]Originally the ERA40 data set was planned to be 40 years long, which gave the name to the project. However during realization of the project another five years of data were added.

that might be typically expected (Section 4.3). We attribute robust features of these plots to specific physical climate processes and other less robust features to methodological artefacts (Section 4.4). In Section 4.5 we also discuss the connection between point statistics and zonal mean statistics as the starting point for a more complete physical theory explaining spectral power growth of the general circulation. We provide a summary of this chapter in Section 4.6. The material in this chapter has been published, along with relevant material in Chapter 2, in the Journal of Climate (Vyushin and Kushner, 2009).

4.2 Results for unfiltered data

Having benchmarked the H estimation methods with synthetic data in Section 2.3 we now apply the methods to the monthly mean tropospheric and stratospheric ERA40 air temperature from September 1957 to August 2002 (Uppala and Coauthors, 2005). The annual cycle and three of its harmonics are removed from the temperature. The estimates of the Hurst exponent, $\hat{H}$, are then carried out identically to the benchmark tests in Section 2.3. We calculate $\hat{H}$ at each longitude, latitude and pressure level, and take the zonal mean of the result to obtain a zonal cross section that characterizes the power-law behavior of the global atmosphere. We plot the resulting zonal mean $\hat{H}$ in Fig. 4.1a-d for DFA3(t), DFA3(a), periodogram GPHE(a) and periodogram GSPE(a) (see Table 1 for parameter settings for these methods). By including these various methods and being clear about parameter settings we aim to reconcile existing results for the value of H for air temperature since different studies use different methods and different frequency ranges (see Section 1.2 and Table A.1).

Fig. 4.1 shows that, apart from a common maximum H in the tropical troposphere, there are clear contrasts between DFA3(t) and DFA3(a), and also between DFA3 and the two spectral methods. The differences generally lie well outside the range of biases found in the benchmark tests and one of our main aims is to pin down the source of these differences. DFA3(t) and DFA3(a) show a similar decrease of H from the tropics to the extratropics, but the values of

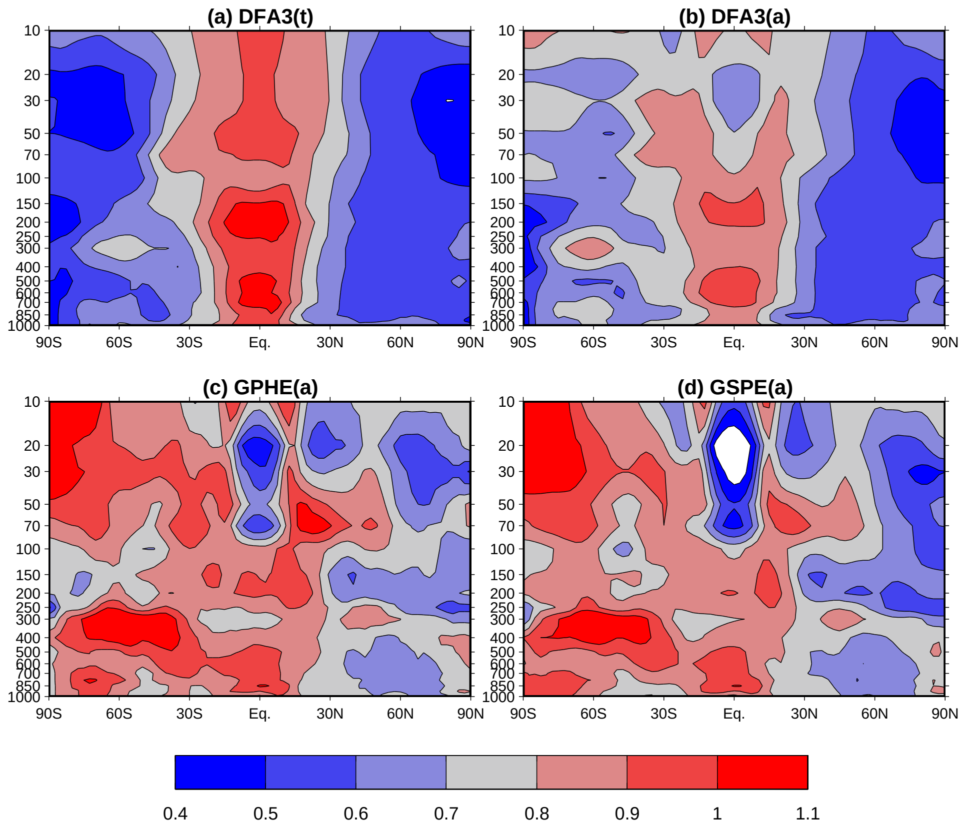

Figure 4.1: Zonal mean $\hat{H}$ of ERA40 air temperature for a) DFA3(t), b) DFA3(a), c) GPHE(a), d) GSPE(a). Values below 0.4 are shown in white.

H are generally lower in DFA3(a) than in DFA3(t). This is particularly true in the tropical lower stratosphere, where the methods disagree most strongly (see Section 4.4). The spectral methods produce noisier plots, as might be expected by their generally larger variance (see Section 2.3). They display pronounced maxima in the Southern Hemisphere that are not found in the DFA3 plots and that we will show are largely tied to linear trends in the data, some of which might arise from data inhomogeneities (Section 4.4). They also show strong minima and even blue-noise (positive spectral slope) behavior in the tropical stratosphere; these features are also discussed below.

4.3 Effect of multitapering and frequency range

We reported in Section 2.2 that multitaper based GPHE and GSPE have somewhat similar biases and standard deviations to the periodogram based GPHE and GSPE. The difference between the multitapered MTM GSPE(a) and the periodogram GSPE(a) for linearly detrended data is shown in Fig. 4.2a. When all frequencies are included, the differences between the methods range from about -0.15 to about 0.1. The asymptotic limit of the GSPE Hurst exponent estimate standard deviation (one sigma) for this frequency range is $1/2\sqrt{30} \approx 0.1$, where 30 is the number of frequencies used. Therefore the differences for the first case are statistically significant only for a few locations and the MTM GSPE plot of $\hat{H}$ is visually similar to Fig. 4.1d (not shown). But we recall that MTM GSPE(a) has a larger bias than MTM GSPE(t) (Section 2.3). Thus, standard practice would suggest that we should compare MTM GSPE(t) and GSPE(a). Fig. 4.2b plots the difference of $\hat{H}$ for MTM GSPE(t) and GSPE(a). The total area where the two methods disagree has increased relative to Fig. 4.2a. The difference between MTM GSPE(t) and GSPE(t) is even larger (see Fig. 4.2c) especially in the observation sparse regions, such as the Southern Hemisphere stratosphere, which points to different sensitivities to data inhomogeneities (Section 4.4) in periodogram and MTM based methods. Given this, it is clear that differences introduced by standard multitapering methods reflect not only the effect of multitapering itself but also a) the selection of frequency range for the power-law fit; b) the effect of data inhomogeneities; and c) the increased variance of H due to trimming.

At this point we are in a position to reduce the number of methods we consider. Together with the results of the Monte-Carlo testing, and additional testing with the GSPE method, we conclude that multitapering adds unnecessary complication to the Hurst exponent spectral estimation procedure. Multitapering might produce graphically smoother spectral plots (as in Fig. 1.2), but it does not provide obviously improved H estimates. We have also found that the all-frequency estimates work equally well for both DFA3 and the periodogram spectral methods. Finally, we find the GSPE method to be similar to but moderately more robust than the GPHE method. We thus proceed to focus mainly on the DFA3(a) and GSPE(a) methods,

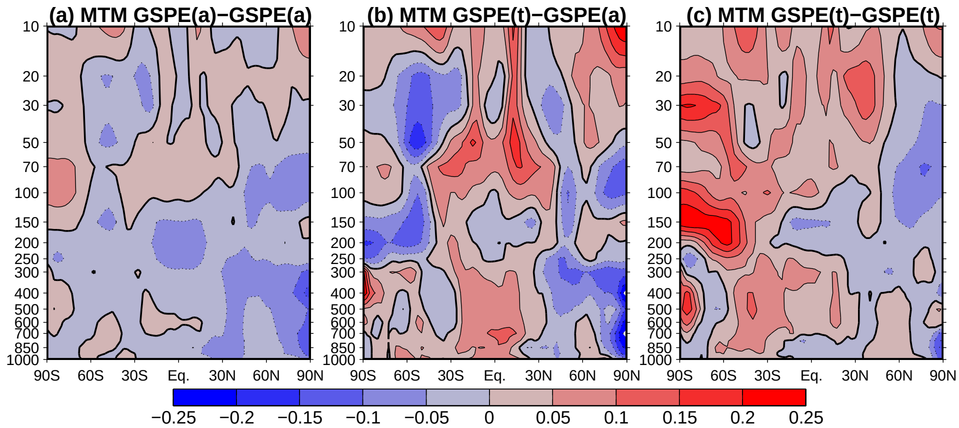

Figure 4.2: a) Zonal mean of $\hat{H}$ for MTM GSPE(a) minus that for GSPE(a). b) Zonal mean of $\hat{H}$ for MTM GSPE(t) minus that for GSPE(a). c) Zonal mean of $\hat{H}$ for MTM GSPE(t) minus that for GSPE(t). All $\hat{H}$ were estimated for linearly detrended (**LTR** filtered, in the notation of Section 4.4) ERA40 air temperature.

and try to explain the robust and non-robust aspects of their $\hat{H}$ portraits.

4.4 Effects of filtering and choice of reanalysis product

We now show that many of the differences between the spectral and time domain methods (e.g. between Figs. 4.1b and d) can be attributed to specific physical processes and methodological artefacts. We consider the effects of detrending, the quasi-biennial oscillation (QBO), ENSO, and volcanic aerosol forcing. The filters we use are

- **LTR:** We remove a simple linear trend from the data.

- **QBO**: We remove a QBO signal by means of multilinear regression using the equatorial zonally averaged zonal winds at 30 and 50 hPa (see Section 3.3.1). We use winds at both 30 and 50 hPa, because they are about 90 degrees out of phase, which allows a better representation of the QBO signal.

- **ENSO**: We remove an ENSO signal consisting of the Nino3.4 index lagged by 4 months by means of linear regression. We choose 4 months lag because it maximizes correlations between Nino3.4 index and tropical troposphere air temperature (Yulaeva and Wallace, 1994; Trenberth and Smith, 2006).

- **VOL**: We remove the effect of volcanic aerosols by regressing air temperature on meridionally and time dependent historical reconstructions of volcanic aerosols optical depth (Ammann et al., 2003).

All signals described above are modulated by the seasonal cycle in our filtering procedure (see Section 3.3.1). We have also carried out additional calculations involving solar and Atlantic multidecadal variability signals, but these did not show significant effects on H estimates. The impact on $\hat{H}$ of removing each signal is plotted in Fig. 4.3. For all rows in Fig. 4.3 the left column corresponds to DFA3(a) and the right to GSPE(a).

The first row of Fig. 4.3 (Figs. 4.3a and b) shows the effect of the linear detrending (filtering **LTR**) on $\hat{H}$. Specifically, $\hat{H}$ with **LTR** filtering is subtracted from $\hat{H}$ with no filtering. As expected, detrending has little effect on the DFA3 based estimate, since this method effectively filters out polynomial trends up to the second order. However the effect is significant for GSPE especially in the Northern Hemisphere lower stratosphere and Southern Hemisphere troposphere. The presence of a linear trend increases $\hat{H}$ by 0.1 to 0.25 for the spectral methods, because a linear trend increases power at low frequencies and therefore steepens the spectral slope. It is well known that climate trends in reanalysis products often reflect data inhomogeneities (Dell'Aquila et al., 2007; Bromwich and Fogt, 2004; Marshall, 2002; Randel and Wu, 1999; Randel et al., 2000), although effect of data inhomogeneities is generally not a trend. We do not aim to evaluate the realism of these trends; instead, we want to point out the relative sensitivities of the H estimation methods to detrending.

The second row of Fig. 4.3 (Figs. 4.3c and d) shows the impact of removing the QBO. Specifically, $\hat{H}$ with **LTR + QBO** filtering was subtracted from $\hat{H}$ with **LTR** filtering. For DFA3(a) and GSPE(a), removing the QBO reduces $\hat{H}$ in the tropical and subtropical lower stratosphere, but the impact is much greater for GSPE(a) than for DFA3(a). Thus the presence of a quasiperiodic signal appears to significantly impact the spectral method. In contrast to the linear trend case, the QBO boosts frequencies near the high-frequency cutoff of 18 months and shallows the spectral slope. Thus, the presence of the QBO reduces $\hat{H}$.

In another analysis, we have found that for the trimmed DFA, DFA3(t), which is the standard method in the literature, the effect of filtering the QBO was to significantly *increase* $\hat{H}$ in the lower stratosphere (not shown), opposite to what is seen in Figs. 4.3c and d. This effect can be attributed to spreading of the QBO signal by the DFA smoothing (e.g. Jánosi and Müller, 2005; Marković and Koch, 2005). The effect is seen in the difference between the DFA3(t) and DFA3(a) plots in Figs.4.1a–b. This again illustrates how sensitive the H estimation methods are to the frequency range choice.

Figs. 4.3e and f show the impact of removing the ENSO signal. The location of the difference is in the tropical troposphere and the sense of the impact is similar to the QBO case. ENSO represents a high (interannual) frequency signal that is significantly correlated with tropical temperatures, and so the ENSO and QBO effects on $\hat{H}$ are analogous. Again, the impact on $\hat{H}$ for DFA3(a) is minimal, but it is more significant for the standard-practice DFA3(t) (not shown).

Unlike for the other filtered signals, the impact of the volcanic signal on $\hat{H}$ is similar for both DFA3 and GSPE (Figs. 4.3g and h). Volcanic forcing appears to increase $\hat{H}$ in the tropical and subtropical lower stratosphere. In climate simulations with and without volcanic forcings, we have been able to reproduce this volcanic signature in $\hat{H}$ (see Chapter 5), and Vyushin et al. (2004) have reported a similar boost of surface temperature $\hat{H}$ from volcanic forcing in climate of the 20th century simulations. The fact that volcanic forcing leads to power-law behavior points to an ambiguity in how to interpret power-law spectra as indicators of long-memory processes. In this case the long-memory process is the geophysical one of volcanism, which leads to intermittent pulses of shortwave forcing, rather than a process internal to the atmospheric general circulation.

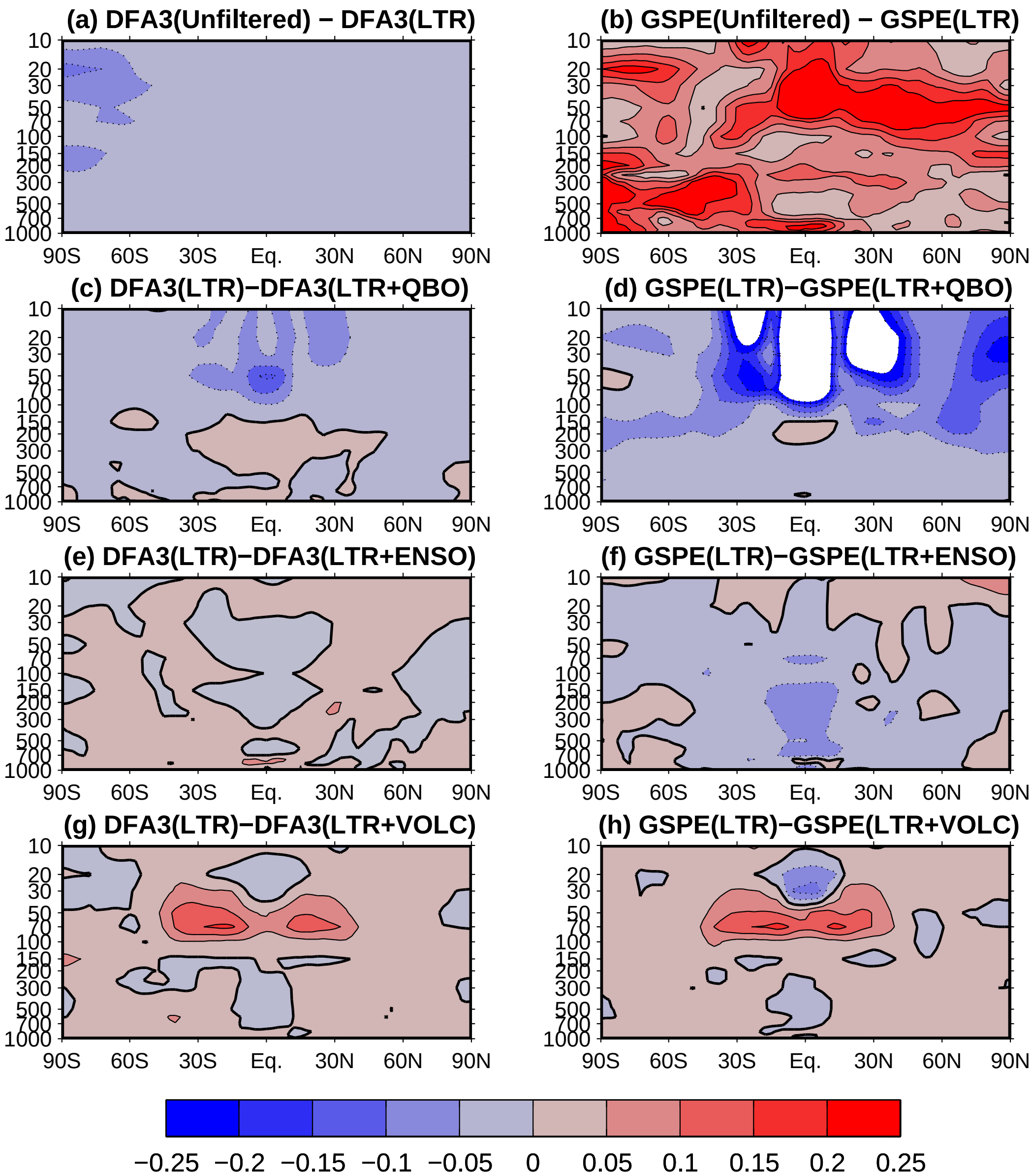

Figure 4.3: Impact on zonal-mean $\hat{H}$ of filtering different climate signals. Difference plots for DFA3(a) are in the left column and difference plots for GSPE(a) are in the right column. First row, a) and b): zonal mean $\hat{H}$ for the unfiltered time series minus that for $\hat{H}$ with **LTR** filtering. Second row, c) and d): zonal mean $\hat{H}$ for **LTR** filtering minus that for $\hat{H}$ with **LTR+QBO** filtering. e) and f): as in c) and d), but for **ENSO** instead of **QBO** filtering. g) and h): as in c) and d), but for **VOLC** instead of **QBO** filtering.

In Fig. 4.1, we saw considerable disagreement between the methods, and especially between the spectral-domain and time-domain methods. Now that we have accounted for the various effects of trends, QBO, ENSO, and volcanoes, as well as considered the effect of using different time scale ranges, we compare again the spectral and time domain methods. Compared to the corresponding plots in Figs. 4.1b and d, the methods have converged considerably. Both methods show relatively large $\hat{H}$ in the tropical troposphere, subtropical lower stratosphere, in the tropical stratosphere above 20hPa, and in the extratropical Southern Hemisphere. The methods still disagree substantially in the Southern Hemisphere stratosphere. Overall, the GSPE provides somewhat larger $\hat{H}$, which is expected based on the Monte-Carlo testing (see Fig. 2.2b).

The Southern Hemisphere stratosphere, where GSPE(a) and DFA3(a) continue to disagree in Fig. 4.4a and b, is a highly problematic area for this kind of analysis because of inhomogeneities in reanalyzed data (Randel and Wu, 1999; Marshall, 2002). For example a red spot in $\hat{H}$ structure at (60^0S,300hPa) is caused by an obvious jump in temperature related to the assimilation of the Vertical Temperature Profile Radiometer data (see Section 5.3 and (e.g. Bromwich and Fogt, 2004; Dell'Aquila et al., 2007)). To test the robustness of the H estimates for ERA40, we calculate $\hat{H}$ using the NCEP/NCAR reanalysis air temperature (Kalnay and Coauthors, 1996) for the same time period and with the same filtering applied. Figs. 4.4c and d show that the main features of the $\hat{H}$ portraits found in the ERA40 data are also present in the NCEP data. But in the data poor Southern Hemisphere polar stratosphere, the four panels disagree significantly. It is known that the Southern Hemisphere stratosphere record has nonlinear temperature trends related to photochemical ozone loss (see Chapter 3) and these cannot be filtered out by **LTR** in GSPE; but this does not explain why GSPE(a) gives different results for NCEP and for ERA40. We thus do not expect to find a robust estimate of H in this region from reanalyzed data.

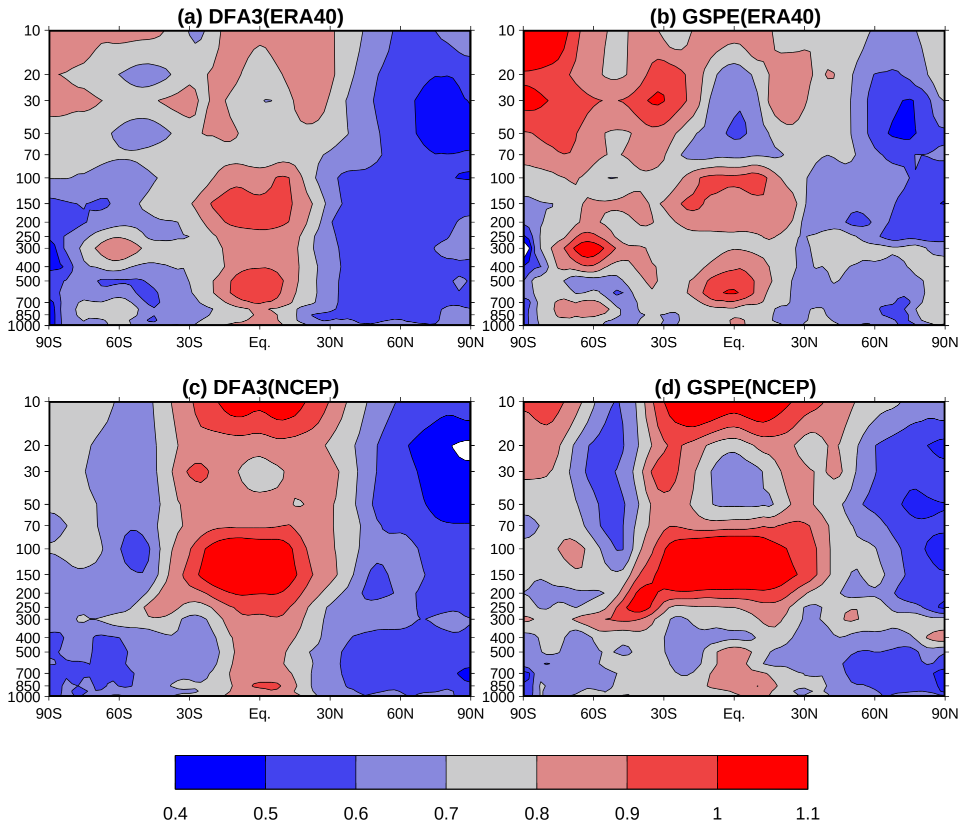

Figure 4.4: Zonal mean $\hat{H}$ with **LTR+QBO+ENSO+VOLC** filtering for a) DFA3(a) and ERA40 data, b) GSPE(a) and ERA40 data, c) DFA3(a) and NCEP data, d) GSPE(a) and NCEP data.

4.5 Hurst exponent estimates of zonal-mean temperature

Figs. 4.1,4.2,4.3,4.4 represent the zonal average of $\hat{H}$ values calculated at each point. But energy and momentum conservation constraints, along with the theory of eddy mean-flow interactions in the atmospheric general circulation (e.g. Lorenz 1967, Schneider 2006), suggest that $\hat{H}$ values of the zonal mean circulation might also be dynamically interesting. With this very general motivation, we show in Fig. 4.5 $\hat{H}$ for the zonally averaged ERA40 air temperature which, analogously to Fig. 4.4, have had all the (**LTR**, **QBO**, **ENSO**, **VOL**) signals removed. We include both DFA3(a) and GSPE(a) estimates in Figs. 4.5a and b, and plot the difference fields ($\hat{H}$ for the zonal-mean temperature minus the zonal mean of $\hat{H}$ for the temperature at each point) in Figs. 4.5c and d.

We see in Fig. 4.5 that the zonal-mean temperature statistics exhibit considerably more power law behavior in the extratropics than the point temperature statistics. For example, the regions with $\hat{H} > 0.8$ are confined between 15^{0}S and 10^{0}N in Fig. 4.4a but between 40^{0}S and 20^{0}N in Fig. 4.5a. For GSPE this difference is even more pronounced. The boost provided by taking the zonal mean first, which is shown in Figs. 4.5c and d, is remarkably robust between the two H-estimate methods. We have also documented that the zonal averages have larger values of $\hat{H}$ than individual grid point time series in the case of total ozone (see Section 3.5).

There is a possibility that some of the boost seen in Figs. 4.4c and d comes about because of an aggregation effect that arises when independent power-law time series are averaged. In particular, when independent power-law time series are averaged, the H of the mean is greater than the mean H of the individual time series (Granger, 1980). Although the temperature time series are spatially correlated this aggregation effect might still operate on sufficiently large scales. We test for the aggregation effect by the following Monte Carlo test: we create a set of independent synthetic temperature time series with values of H equal to the estimated H at each spatial point in the ERA40 reanalysis grid represented in Fig. 4.4b. Therefore we simulate $144 \times 73 \times 18$ mutually uncorrelated time series using the ARFIMA(0,d,0) model. The zonal mean $\hat{H}$ of this dataset is, by construction, the same as that seen in Fig. 4.4b. We note that we do not include spatial correlations in order to focus on the aggregation effect. We then estimate $\hat{H}$ of the zonal averages of these synthetic time series. The obtained spatial patterns (not shown) are noisier than Figs. 4.4a and b but are numerically close to it. Thus for mutually uncorrelated time series the statistical aggregation effect is negligible. This suggests that the boost in H from using the zonal-mean temperature is of dynamical origin and stems from systematic zonal correlations of the eddy fields.

The boost in Figs. 4.5c and d is relatively small in the tropics, consistent with the idea of Sobel et al. (2002) that point temperatures in the tropics are well correlated with the zonal-mean tropical temperature field. The enhanced values of H in midlatitudes, at the surface

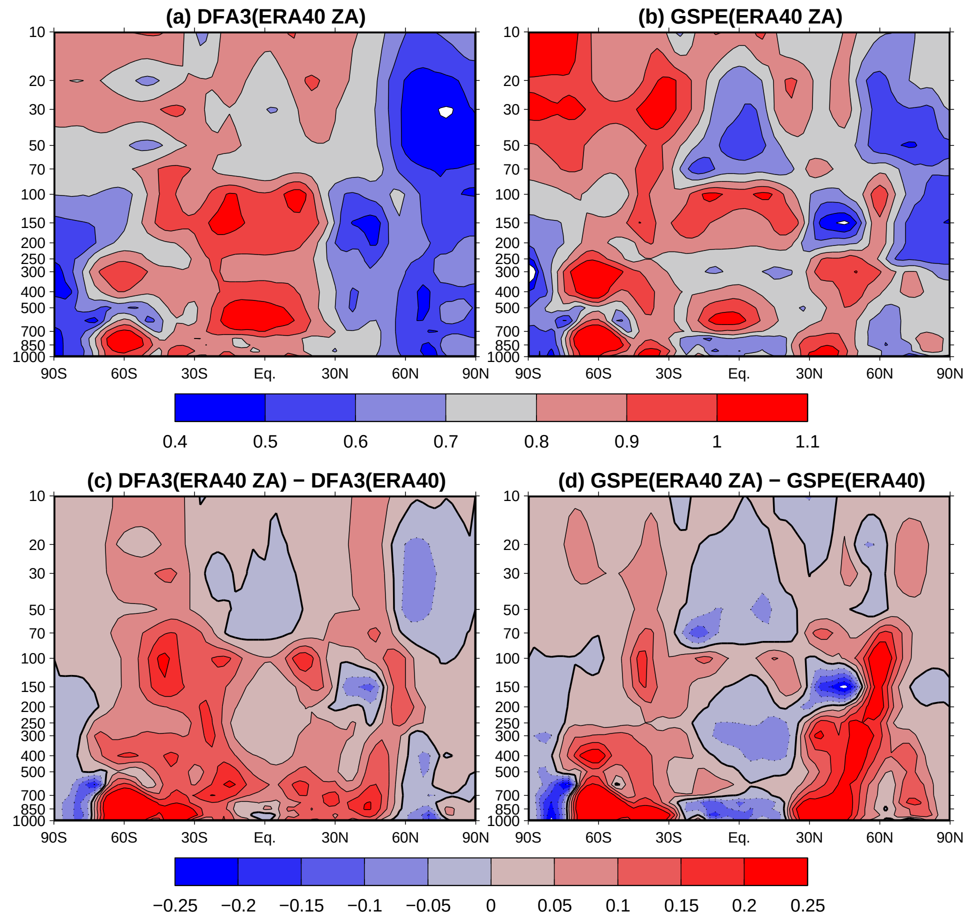

Figure 4.5: a) $\hat{H}$ of the zonal-mean temperature for ERA40 data for DFA3(a). b) As in a), for GSPE(a). c) $\hat{H}$ of the zonal-mean temperature minus zonal mean of $\hat{H}$ for point temperatures, for DFA3(a). d) As in c), for GSPE(a).

and in the lower stratosphere suggest that the long-memory behavior in the tropics is coupled to midlatitudes via the eddy driven zonal-mean overturning circulation (Held and Schneider, 1999).

4.6 Conclusions

Under the working assumption that the atmospheric general circulation exhibits power law behavior, we have estimated the Hurst exponent H for the temperature of the global atmosphere using several statistical methods. Monte-Carlo benchmarking with pure power-law time series reveals no obvious discrepancies between the methods, but when we apply these methods to reanalyzed climate data we find a striking degree of inconsistency among the results. We summarize our current understanding of the methods:

- DFA3 results are insensitive to trends and can be made insensitive to high-frequency periodicities, provided trimming is not applied and all timescales are used, i.e. provided s_{long} is set to N.

- Multitapered and periodogram spectral methods can be made consistent with one another provided consistent frequency ranges are used and the lowest frequencies are included. Since the two methods yield consistent H estimates (Fig. 4.2a), there is no obvious advantage to using multitapering in H estimation, at least in this application.

- The DFA3 and the spectral methods $\hat{H}$ results are quite inconsistent unless filtering is applied, consistent frequency ranges are chosen and the lowest frequencies are included. The spectral methods are sensitive to periodicities and trends, and DFA3 appears to be more robust in this regard.

Given our current understanding, we recommend the use of DFA3(a) and GSPE(a), or alternatively DFA3(a) and GPHE(a), and tests to filtering of well-known climate signals such as the QBO, ENSO and external climate forcings, to provide a representative picture of power-law behavior in climate time series.

Another issue that has arisen is the different sensitivities the methods exhibit to data inhomogeneities, e.g. temperature jumps induced by instrumentation changes. Although previous work suggests that DFA3 is more robust in the presence of such inhomogeneities (Berton, 2004; Chen et al., 2002; Hu et al., 2001) than spectral methods, its results might also be affected. We discuss this question in more details in Section 5.3 below.

Using DFA3(a) and GSPE(a), we have found several robust high H regions in the atmospheric general circulation. In particular, we have found that point temperature statistics exhibit robust power-law behavior in the tropics that decreases with latitude. A connection between the tropics and extratropics becomes evident when $\hat{H}$ is calculated for the zonal-mean temperature. These results may have practical implications for analysis of tropical tropospheric temperature trends (e.g. Santer et al., 2005; Randel and Wu, 2006; Thorne et al., 2007). These trends are highly nonrobust: satellite measurements, radiosondes, and climate model simulations all provide different values for these trends. But if tropical temperatures exhibit power-law behavior, confidence intervals on these trends would very likely be underestimated using AR1 based noise models. Calculation of confidence intervals according to a power-law model for the residuals as carried out in Smith (1993) would lead to a significant increase of the trends confidence intervals (see also Chapter 3 and Section 6.5). Thus at least some of the apparent discrepancies could be accounted for by properly representing long-range temporal correlations in the tropical atmosphere.

Another robust result, found both for DFA3(a) and GSPE(a), is that volcanic forcing increases $\hat{H}$ in the lower tropical and subtropical stratosphere. Volcanic forcing has also been found to have an effect on H at the surface (Vyushin et al., 2004) and it still remains to reconcile the surface and stratospheric H signatures. Furthermore, since the volcanic forcing record can be imprinted in the deep oceanic circulation (Delworth et al., 2005; Gleckler et al., 2006), this result suggests that some of the long-memory behavior seen in the coupled ocean atmosphere system might be attributable to a volcanic forcing effect.

Chapter 5

Reanalysis vs. specialized GCM simulations

5.1 Introduction

In Chapter 1, we discussed how the ability of climate models to capture observed power-law behaviour has lead to some considerable controversy in the previous literature (e.g. Govindan et al., 2002; Fraedrich and Blender, 2003; Vyushin et al., 2004). With the results of Chapters 2 and 4 in hand, we have established the relative sensitivity and reliability of the different estimation methods. We are now in a position to analyze the ability of climate prediction models to simulate temporal scaling behavior. In our view this represents a stringent performance test because it requires the model to capture variability on a wide range of timescales.

In this chapter, we estimate the power-law exponent distribution — i.e. the Hurst exponent distribution — for the global atmospheric circulation of the stratosphere and troposphere during the 20th century, in observations and in climate simulations, and use the climate simulations to gain insight into the distribution. This is a significant extension of the previously cited literature, which has generally been restricted to surface air temperature. We will highlight results that are independent of the Hurst exponent estimation technique. Here we focus on DFA3 results, since it is more robust than GSPE; it automatically filters out linear and quadratic trends; and we do not need to estimate scaling factor b in this chapter. However we verify that DFA3 and GSPE, the two methods we recommend to use in similar studies, results agree. We will mainly use highly constrained climate simulations in which the ocean surface temperatures and different combinations of radiative forcings are prescribed; in Chapter 6, we will extend the analysis to coupled ocean-atmosphere climate models.

This chapter is structured as follows. Section 5.2 compares spatial distributions of $\hat{H}$ for reanalyses and various GCM simulations. The effect of tropical SST forcing is analysed in Section 5.3. The results of the DFA3 are compared with the GSPE in Section 5.4. A simple model for understanding the effect of volcanic eruptions on the temporal spectrum of the lower tropical stratosphere is presented in Section 5.5. Section 5.6 concludes. The material in this chapter has been published in Geophysical Research Letters (Vyushin et al., 2009).

5.2 Specialized GCM simulations

Fig. 5.1 plots DFA3 estimates of H for the reanalysis products and several climate simulations. The $\hat{H}$ distribution displays a characteristic shape that we have verified is robust to different methods of H estimation (see Chapter 4). Both the NCEP/NCAR and ERA40 reanalyses (Figs. 5.1a and b) show maxima in $\hat{H}$ in the tropical to low-extratropical troposphere and in the tropical to subtropical stratosphere and a minimum in the Northern Hemisphere polar stratosphere. But there are differences between the reanalysis products; for example, ERA40 has separate local maxima in $\hat{H}$ in the lower and upper troposphere at 60^0S that will be discussed later in relation to Fig. 5.3. We will also show that even where the distributions appear to agree, they might do so for different reasons.

Fig. 5.1c plots the $\hat{H}$ distribution for a simulation of the GFDL Atmospheric Model (AM2.1 The GFDL Global Atmospheric Model Development Team, 2004) forced by historical SSTs,

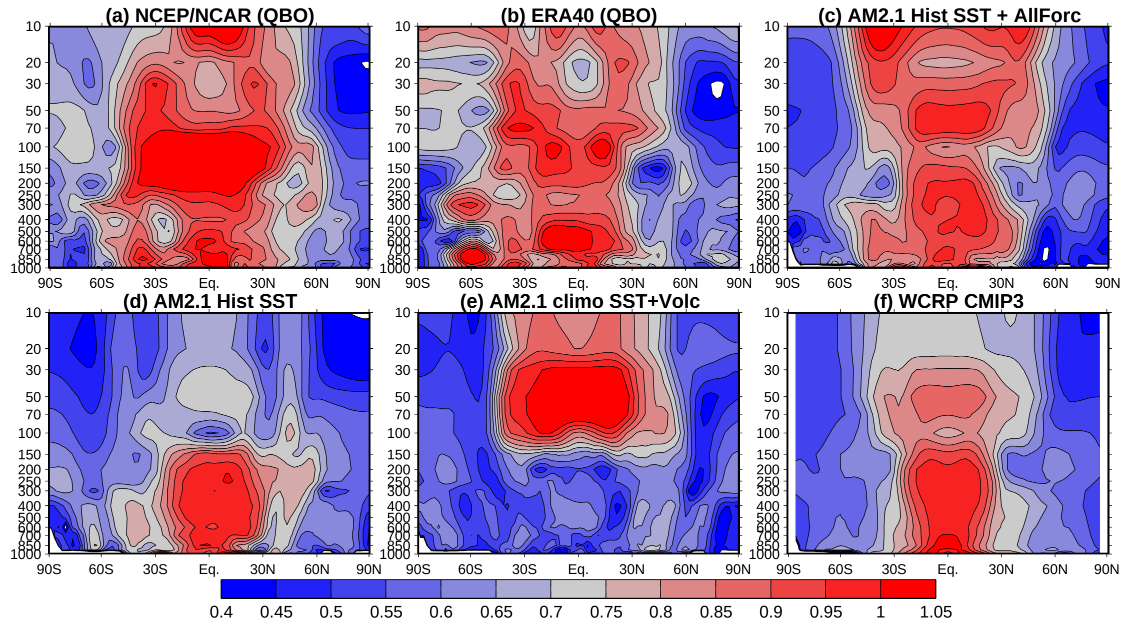

Figure 5.1: $\hat{H}$ distribution for zonal-mean temperature for (a) the NCEP/NCAR reanalysis, (b) the ERA40 reanalysis, (c) the GFDL AM2.1 HistSST+AllForc simulation, (d) the GFDL AM2.1 HistSST simulation, (e) the GFDL AM2.1 Vol simulation, (f) the CMIP3 simulations. Panel (f) represents a multiple model average. As stated in the text, QBO filtering has been applied to the reanalysis temperatures in panels a-b.

anthropogenic greenhouse gases and aerosols, ozone changes, solar flux, and volcanic aerosols (hereafter the "HistSST+AllForc" simulation). The main features of the $\hat{H}$ distribution of this simulation are similar to those displayed in the observationally based Figs. 5.1a-b, including the falloff of $\hat{H}$ as we move from the equator to the poles and separate maxima in the lower stratosphere and the troposphere. Therefore given historical SSTs and the other principal external forcings the GFDL AM2.1 is able to reproduce the continuum of zonal mean temporal temperature variability represented by the Hurst exponent.

Three additional simulations of AM2.1 help explain the physical origins of the $\hat{H}$ distribution. First, a simulation with time-independent radiative forcings and with prescribed climatological SSTs (labelled "Climo") has $\hat{H}$ values close to 0.5, with a range of 0.4 to 0.6

(results not shown). Thus, the atmosphere exhibits a flat spectrum on interannual to multi-decadal timescales in the absence of forcing on these timescales. Second, a simulation with time-independent radiative forcings and with historical SSTs ("HistSST") has a tropospheric pattern of $\hat{H}$ (Fig. 5.1d) that is similar to that in Figs. 5.1a-c. The simulated tropospheric H is thus determined by the SSTs and is consistent with the surface-temperature analysis of Fraedrich and Blender (2003) that suggested that the origin of large $\hat{H}$ in the atmosphere is oceanic. Third, a simulation with time-independent forcings from the radiatively active gases and anthropogenic aerosols and with prescribed climatological SSTs, but with historical volcanic forcing ("Vol") gives rise to a stratospheric pattern of $\hat{H}$ (Fig. 5.1e) that is similar to that in Figs. 5.1a-c. To summarize, the simulations show that the observed $\hat{H}$ distribution is mainly determined by temporal variability of the SSTs in the troposphere and by volcanic forcing in the lower stratosphere.

We briefly demonstrate that current generation climate models can capture the $\hat{H}$ distribution in a less constrained forcing framework. The $\hat{H}$ distribution averaged over the CMIP3 coupled ocean-atmosphere model simulations of the 20th century is shown in Fig. 5.1f; it displays a similar structure to Figs. 5.1a-c but has a narrower meridional extent and a weaker volcanic signature in the lower stratosphere. The simple explanation for the latter is that only 9 of the 17 models considered included realistic volcanic forcings. In Chapter 6 we will compare spatial distributions of $\hat{H}$ for the 20th century simulations of the models with a realistic volcanic forcing to those without it. We will also show that the $\hat{H}$ for the CMIP3 models is essentially the same whether we estimate $\hat{H}$ for the range 18 months to 45 years over the second half of the 20th century or for the range 18 months to 100 years over the entire 20th century.

5.3 Influence of tropical SSTs

We propose that the relatively steep spectral slopes represented by the $\hat{H}$ maximum centered in the tropical troposphere are generated by tropical SST variability. Our test of this idea reveals

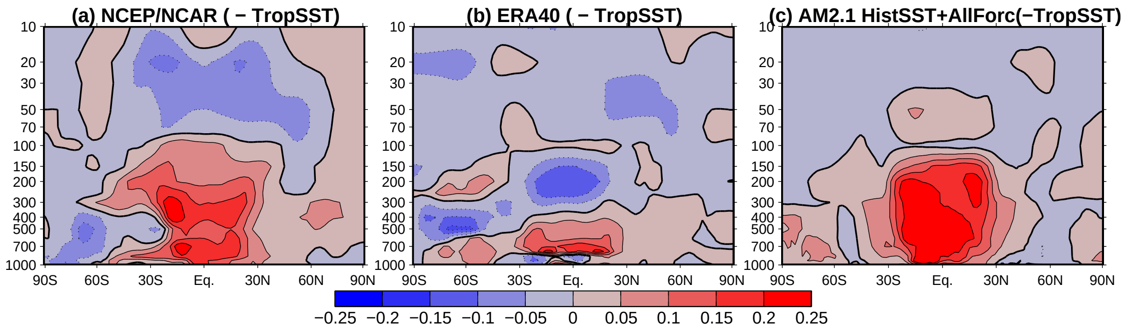

Figure 5.2: $\hat{H}$ without TropSST filtering minus $\hat{H}$ with TropSST filtering, which represents the signature of the tropical SSTs in the $\hat{H}$ field: (a) NCEP/NCAR, (b) ERA40, (c) AM2.1 HistSST+AllForc. QBO filtering has been applied to ERA40 and NCEP/NCAR reanalyses.

a significant discrepancy between the two reanalysis products. To test the idea, we create time series of tropical mean SST in the latitude band 20^0S-20^0N (“TropSST” Smith et al., 2008). We then filter the TropSST signal from the temperature time series using linear regression and estimate H of the result for the NCEP/NCAR and ERA40 reanalyses and for the HistSST+AllForc simulations. Fig. 5.2 isolates the part of the $\hat{H}$ distribution related to tropical SSTs by showing the original $\hat{H}$ minus the TropSST-filtered $\hat{H}$. In the NCEP/NCAR reanalysis (Fig. 5.2a) and in the simulation (Fig. 5.2c), there is a vertically coherent part of the $\hat{H}$ distribution throughout the tropical and low extratropical troposphere that is related to the TropSST signal, as indicated by the positive values. The TropSST $\hat{H}$ signature in the ERA40 reanalysis (Fig. 5.2b) is qualitatively different, being vertically incoherent and of mixed sign.

In Fig. 5.2, the NCEP/NCAR reanalysis and the climate model simulation appear to agree with our hypothesis of tropical SST control, while the ERA40 appears to disagree with it. To understand these inconsistent results we display the residuals of the tropical upper tropospheric temperatures after TropSST filtering has been applied, for the three reanalysis products and for the HistSST+AllForc and HistSST simulations (Fig. 5.3a). A one year running average has also been applied. The ERA40 residuals (shown in red) show much more decadal variance than the NCEP/NCAR and Japanese reanalysis (JRA-25 Onogi and Coauthors, 2007) residuals and the

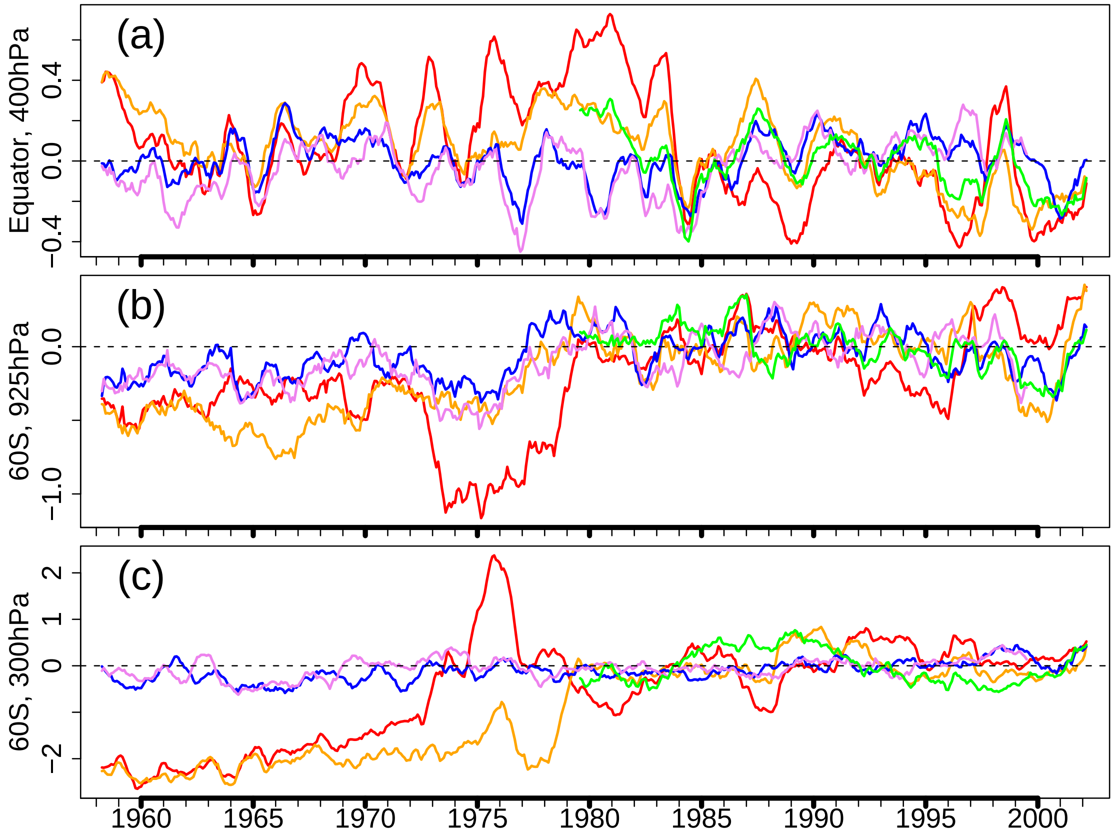

Figure 5.3: The one year running mean of zonally averaged air temperature residuals (a) at (Equator, 400 hPa) with TropSST filtering as described in the text; (b) at (60^0S,925hPa), without TropSST filtering ; (c) as in (b), at (60^0S,300hPa). ERA40 time series are shown in red, NCEP/NCAR in orange, JRA-25 in green, HistSST in blue, and HistSST+AllForc in violet. All time series were adjusted to have zero mean for 1979-2002.

simulations' residuals. Significant fluctuations for the ERA40 include particularly high values during 1975-1983, which are probably related to problems with transition from VTPR to TOVS satellite data (Simmons et al., 2004; Uppala and Coauthors, 2005), and low values for 1986-1991 and after 1992. Similar issues also explain the lower and upper tropospheric $\hat{H}$ maxima at 60^0S that are seen in the ERA40 reanalysis (Fig. 5.1b) but not seen in the NCEP/NCAR reanalysis (Fig. 5.1a) or in the HistSST+AllForc simulation (Fig. 5.1c). Figs. 5.3b and c plot temperature anomalies (without TropSST filtering) from the same five data sets at these locations. There is an obvious jump (negative at 925hPa and positive at 300hPa) in the ERA40 temperature presumably related to problems with assimilation of the VTPR data from 1973 to 1978 (Bengtsson et al., 2004; Simmons et al., 2004). Another striking difference between the models and reanalyses are the strong positive trends at 300hPa. These trends seem to be spurious and stem from the reanalysis models' cold biases combined with a gradual increase in the number of observations in the Southern Hemisphere (Bengtsson et al., 2004; Simmons et al., 2004). Discrepancies in the Southern Hemisphere polar stratosphere have been discussed in Section 4.4. Therefore several data inhomogeneity issues in the ERA40 affect and are revealed by our H analysis. However this does not exclude a possibility of other data problems present in the ERA40 and NCEP/NCAR reanalyses. The data inhomogeneities described above could also probably be identified by comparison of the spatial distributions of standard deviations of the annual mean anomalies of the reanalyses and model simulations. However the Hurst exponent characterizes temporal variability for a range of time scales, whereas the variance is typically dominated by a high-frequency variability of a given time series. Thus to capture possible data inhomogeneities using the variance one should test several temporal aggregations, e.g. to estimate standard deviations of daily, monthly, annual, or decadal means. In contrast, the power-law analysis achieves this goal in one step.

5.4 DFA3 vs GSPE

We have established two primary methods of H estimation, DFA3(a) and GSPE(a), and now test whether some of our key results are method dependent. In the Figs. 5.1 and 5.2 we used the DFA3 time domain H estimator. We have shown in Chapter 4 that DFA3 and other spectral domain methods yield consistent estimates of H provided a consistent frequency range has been chosen and known climate signals have been filtered out. DFA3 effectively filters out linear and quadratic trends in the data, which helps us to focus on internal climate variability. The distribution of the DFA3 $\hat{H}$ seems to be approximately Gaussian (Rybski and Bunde, 2009)

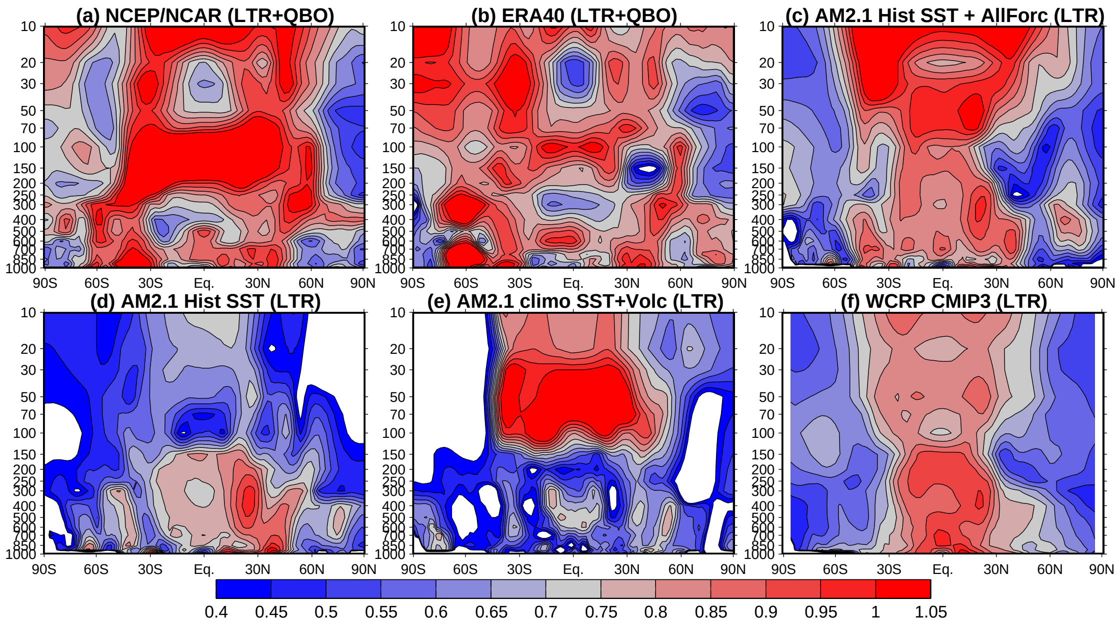

Figure 5.4: $\hat{H}$ distribution estimated by GSPE for zonal-mean temperature for (a) the NCEP/NCAR reanalysis, (b) the ERA40 reanalysis, (c) the GFDL AM2.1 HistSST+AllForc simulation, (d) the GFDL AM2.1 HistSST simulation, (e) the GFDL AM2.1 Vol simulation, (f) the CMIP3 simulations. Panel (f) represents a multiple model average. As stated in the text, QBO filtering has been applied to the reanalysis temperatures in panels a-b. Values of $\hat{H}$ less than 0.4 are shown in white.

with standard deviation ≈ 0.075 for the case of the time scale range of 18 to 540 time units (see Weron (2002) and Section 2.3). Overall, DFA3 provides more robust estimates than other available methods, but for the results reported here, we have found consistent results using the spectral domain Gaussian semiparametric estimator.

Figs. 5.4 and 5.5 apply the Gaussian Semiparametric Estimator (GSPE, Robinson, 1995a) to the same data sets as in Figs. 5.1 and 5.2. GSPE is a maximum-likelihood spectral domain estimator of H. The GSPE $\hat{H}$ is known to be relatively sensitive to the presence of linear and nonlinear trends and to high-frequency spectral peaks compared to DFA3 (see Section 4.4). Because of the known sensitivity to trends we have filtered out the linear trend before calculating the GSPE $\hat{H}$.

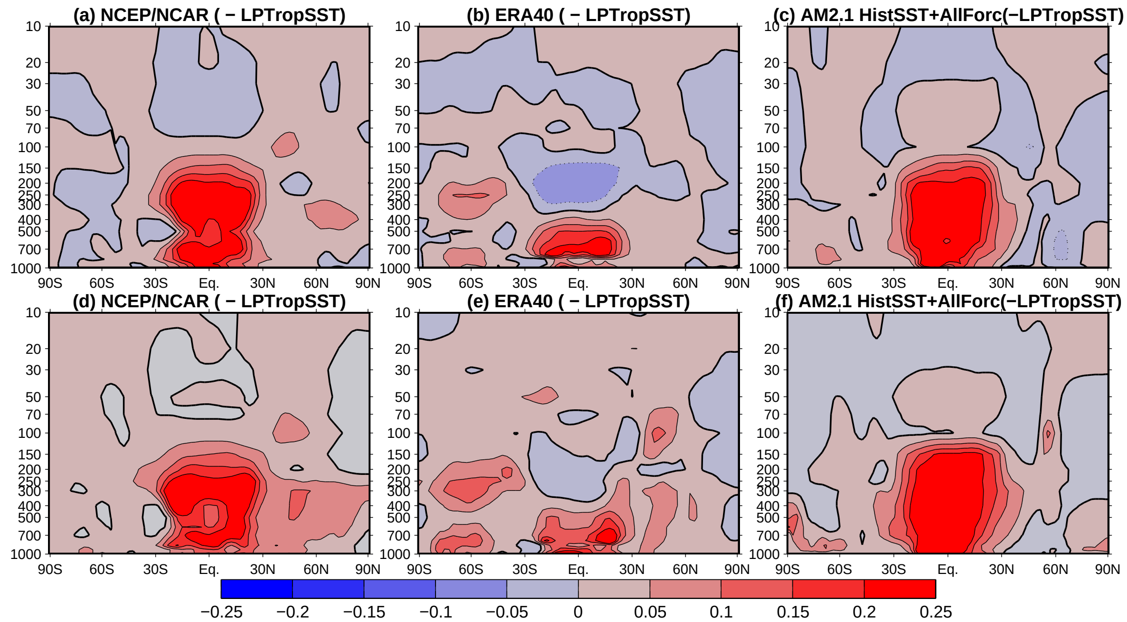

Figure 5.5: $\hat{H}$ without LPTropSST filtering minus $\hat{H}$ with LPTropSST filtering, which represents the signature of the tropical SSTs in the $\hat{H}$ field. First row - DFA3 estimates, second row - GSPE estimates.

In Fig. 5.4 the overall distribution of the GSPE $\hat{H}$ is similar to, but noisier than, the DFA3 $\hat{H}$ in Fig. 5.1. However, significant differences remain. For example, Southern Hemisphere stratospheric values of $\hat{H}$ are larger for GSPE than for DFA3 in the reanalyses due to data inhomogeneities and to nonlinear trends from ozone depletion (see Section 4.4 and Chapter 3). In addition, the differences in the tropical troposphere arise from ENSO related variability that boosts the high frequencies and is known to reduce the GSPE $\hat{H}$ relative to the DFA3 $\hat{H}$ (see Section 4.4). In the CMIP3 simulations this discrepancy is present when we analyze 45 year long time series, but is reduced when we analyze 100 year long time series.

We also have to bear in mind GSPE's sensitivity to high frequency spectral peaks when we try to reproduce the results of Fig. 5.2, which illustrates the sensitivity of $\hat{H}$ to tropical SST variability. The impact on the GSPE $\hat{H}$ of TropSST (Smith et al., 2008) filtering (not shown) looks quite different from that for the DFA3 $\hat{H}$, which is shown in the Figs. 5.2d-f. This difference arises because the high frequency component of the TropSST signal dominates the GSPE $\hat{H}$ response while the low frequency component of the TropSST signal dominates the DFA3 $\hat{H}$ response. But we can put the two methods on a more even footing by focusing on the decadal component of the tropical SST variability, which is the timescale of interest in this thesis. To do so, we construct a 3 year low-pass filtered tropically averaged SST signal ("LPTropSST") and compute the response to LPTropSST filtering in DFA3 $\hat{H}$and GSPE $\hat{H}$ in Fig. 5.5. The first row of this figure is similar to Fig. 5.2, showing that the DFA3 $\hat{H}$ response is robust to the low-pass filtering. The first and the second row of Fig. 5.5 are also remarkably similar, indicating that the tropical SST effect is in good agreement in the two methods.

5.5 Effect of volcanic eruptions

We return to the volcanic signature of $\hat{H}$ in the lower stratosphere. It has been shown theoretically that a sum of stochastic amplitude shocks decaying by a power law has a power law spectrum (Parke, 1999). But the volcanically induced warming of the stratosphere decays exponentially in time (Robock, 2000) and so we cannot expect power-law behavior in temperature except over a limited range of frequencies.

A simple model to capture the behavior is

$$\frac{dT}{dt} = -\frac{1}{\tau}T + V(t) \tag{5.1}$$

where T is temperature, τ is a relaxation time scale and $V(t)$ is the volcanic forcing (Stenchikov et al., 2006) (expressed as aerosol optical depth). Fig. 5.6a shows one-year running mean air temperature anomalies in the tropical lower stratosphere obtained from the GFDL AM2.1 Vol simulation (see Fig. 5.1e), Fig. 5.6b the solutions to (5.1) for various values of τ, and Fig. 5.6c the power spectra for the time series in Fig. 5.6b. As the relaxation time scale gets larger the power spectra saturate at lower frequencies, which would give rise to larger estimates of H. These power spectra demonstrate a combination of power-law behavior between 6 months and 4-10 years and a flat spectrum at the lowest frequencies. When we

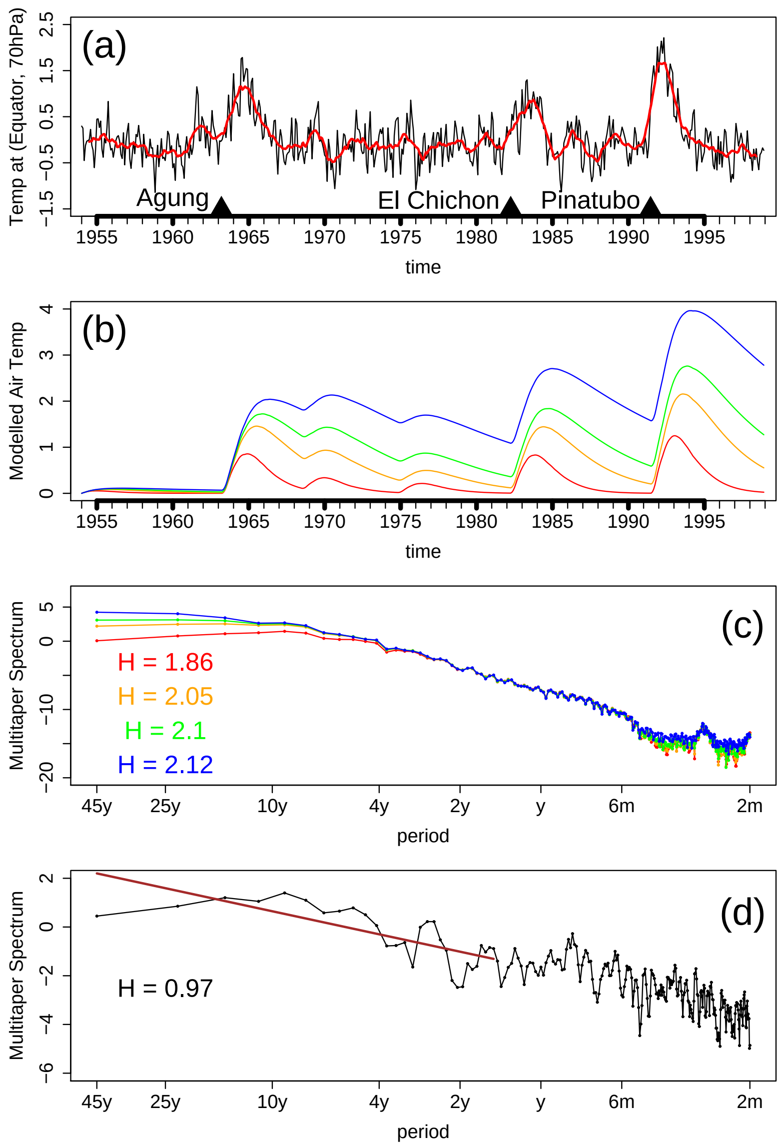

Figure 5.6: (a) Air temperature anomalies at (Equator, 70hPa) from the AM2.1 Vol simulation. The smooth red curve is a one year running average. The timing and the names of the major volcanic eruptions are shown above the time axis. (b) Solutions to equation (5.1) for $\tau = 1$ year (red curve), 3 (orange), 5 (green), and 10 years (blue). (c) Multitapered power spectra of these solutions and their DFA3 Hurst exponent estimates. (d) The power spectrum of the solution for $\tau = 1$ with the weather noise superimposed on it. The best fit power-law curve (line in log-log coordinates) is shown in brown. Panels (c-d) are plotted in log-log coordinates.

repeat the same calculation for the volcanic forcing record of the past 130 years the saturation occurs at lower frequency (not shown).

The DFA3 $\hat{H}$ for the solutions to (5.1) are labelled with colors corresponding to the spectra in Fig. 5.6c. The $\hat{H}$ values are so large because our simple model does not include regular weather noise, which boosts spectral power in high-frequencies and thus decreases the $\hat{H}$. To support this we have plotted the power spectrum of the solution to (5.1) for $\tau = 1$ with the weather noise superimposed on it. We employed the time series of air temperature monthly mean anomalies at the equator at 70hPa obtained from the Climo simulation of GFDL AM2.1 that was forced with climatological SSTs and with time-independent radiative forcings. The Hurst exponent estimate of this time series, obtained using the DFA3 method, is 0.97, which agrees well with the values in the lower tropical stratosphere in Fig. 5.1e.

5.6 Conclusions

To conclude, we find that zonal-mean air temperature on interannual to multi-decadal timescales has a steep spectrum that might be modelled by power-law behavior in the tropical to low-extratropical troposphere and the tropical to subtropical stratosphere. Current generation climate models can capture these features and specialized simulations elucidate their dynamics. We propose that the tropospheric $\hat{H}$ signatures are linked to tropical SST variability and that the lower stratospheric $\hat{H}$ signatures are linked to volcanic forcing. The link to tropical SST variability is clear in the NCEP/NCAR reanalysis. The large $\hat{H}$ values in the tropical upper troposphere in the ERA40 reanalysis appear to arise from data problems that mask the connection to tropical SSTs. The ERA40 H estimates also exhibit tropospheric maxima at 60^0S that appear related to other documented data assimilation issues.

This analysis points to problems in naively interpreting the Hurst exponent distribution as an indicator of long-term memory in climate and care needs to be taken to elucidate the physical basis for a given $\hat{H}$ feature. Data inhomogeneities affect many observational time series and can equally give rise to power-law behavior (Berton, 2004; Rust et al., 2008). Sometimes, such as at 60^0S in the troposphere, it is immediately evident that there is a discrepancy to explain, but at other times, such as in the tropical troposphere, the effort still needs to be made to test the consistency of the power-law behavior under different physical hypotheses. We have found that general circulation models provide a useful tool for such testing.

The frequent presence of power-law behavior, whatever its cause, suggests that statistical testing for significant trends and periodicities should use power-law noise models (see Smith (1993), Chapter 3, and Section 6.5) as well as AR1-models, particularly in the tropical upper troposphere and lower stratosphere where $\hat{H}$ is large and trend evaluation has proven difficult (e.g. Santer et al., 2005). Power-law based confidence intervals are typically larger because they assume more power at lower frequencies. For example, power-law based significance testing has been applied to the problem of stratospheric ozone recovery in the presence of significant stratospheric internal variability, and leads to a lengthening of the projected time for the detection of ozone recovery (Chapter 3).

Chapter 6

Analysis of CMIP3 simulations

6.1 Introduction

In this chapter we study persistence and spectral power growth of the surface and free atmosphere air temperature derived from several observational products and CMIP3 climate model simulations. The CMIP3 project provides an excellent opportunity for verification and generalization of features and mechanisms found in individual observational data sets and models, which were the subject of almost all previous climate persistence studies in climate literature and Chapters 4 and 5. We compare the fidelity and the goodness-of-fit of the AR1 vs the power-law model and show that they provide a lower and an upper bound for climate persistence on monthly to decadal time scales. We test the robustness of the power-law fit by varying the spectral range to which it was fitted and scenarios under which model simulations have been performed. We provide a comparison with previously published results based on individual climate model simulations and paleo reconstructions. The material in this chapter represents a manuscript by Vyushin and Kushner in preparation for submission to Climate Dynamics.

6.2 Data and Methods

In this chapter we use three observational products: the NCEP/NCAR reanalysis (Kalnay and Coauthors, 1996), the ERA40 reanalysis (Uppala and Coauthors, 2005), and the NASA GISS surface air temperature (Hansen et al., 1999). The GISS dataset combines observations of land meteorological stations and sea surface temperature data. It maps the observations on a 2^ox2^o grid and smoothes them over 1200 km. We also employ the pre-industrial control (*picntrl*) and the 20th century (*20c3m*) simulations of 17 atmosphere-ocean coupled general circulation models from the CMIP3 database: CGCM3.1(T47), CGCM3.1(T63), CSIRO-Mk3.0, CSIRO-Mk3.5, ECHAM5/MPI-OM, GFDL-CM2.0, GFDL-CM2.1, GISS-AOM, GISS-EH, GISS-ER, MIROC3.2(medres), MIROC3.2(hires), MRI-CGCM2.3.2, NCAR CCSM3.0, NCAR PCM, UKMO-HadCM3, UKMO-HadGEM1. We also analyze 500 year long *picntrl* simulations of six GCMs (CGCM3.1(T47), ECHAM5/MPI-OM, GFDL-CM2.0, GFDL-CM2.1, GISS-ER, MIROC3.2(medres)).

For each scenario (*picntrl* or *20c3m*) we have found it sufficient to use a single realization from each model. For all the observational products we use the ERA40 period September 1957 to August 2002 and compare this to the *20c3m* simulation period 1955-1999 for the models. We also use the longer *20c3m* simulation period 1900-1999. We note that for the *20c3m* simulations all the models were forced by anthropogenically changing greenhouse gases and aerosols and some models were also forced by changes in stratospheric ozone, solar radiation and volcanic aerosols. Model details have been documented previously (e.g. Santer et al., 2005) and can also be found on the CMIP3 web-site. The seasonal cycle and its first three harmonics are filtered out from all time series. In addition, we have filtered out the effect of the QBO from reanalysis zonal mean air temperature following the methodology described in Section 3.3.1, because none of the CMIP3 models simulates this phenomenon.

We estimate the Hurst exponent by means of detrended fluctuation analysis of the third order (DFA3) (Kantelhardt et al., 2001). DFA3 filters out local polynomial trends up to the second order (Kantelhardt et al., 2001) and therefore typically is not sensitive to human induced secular climate change effects, such as surface warming or stratospheric temperature changes forced by ozone depletion and a subsequent recovery (see Section 2.2.2 and 4.4). We have shown in Section 2.3 that DFA is one of the most robust Hurst exponent estimators, but we have verified the results presented in this chapter by means of the Gaussian Semiparametric Estimator (see Robinson (1995a) and Section 2.2.1), which gives results consistent with DFA provided periodic components and anthropogenically induced trends have been filtered out and equivalent frequency ranges have been used (see Chapter 4). The lag-one autocorrelation coefficient is estimated by the Yule-Walker method and has been verified by a maximum-likelihood fitting of the AR1 spectral density (see Eq. 1.3) to the periodogram (Beran, 1994).

We have already discussed the uncertainty of Hurst exponent estimates in Chapter 2. We recall that unfortunately there is no analytical description of DFA $\hat{H}$ properties, but Monte-Carlo simulations demonstrate that it seems to be approximately normally distributed (Rybski and Bunde, 2009) and that for the frequency range of 18 months to 45 years for 45 year long monthly time series $\sigma(\hat{H}) \approx 0.075$ for DFA3 vs ≈ 0.12 for GSPE (see Section 2.3). For the frequency range of 5 to 45 years Monte-Carlo simulations show that $\sigma(\hat{H}) \approx 0.14$ for DFA3 and $\sigma(\hat{H}) \approx 0.27$ for GSPE.

6.3 Results for the surface air temperature

6.3.1 Time aggregation effect

The ability of climate models to reproduce many aspects of the observed climate variability (e.g. Randall and Coauthors, 2007) helps to answer the question of whether spectral power-law behaviour is an appropriate representation of climate variability. In this subsection we compare the relative validity of the AR1 and power-law statistical models for SAT. We will return to this question again in subsection 6.3.4. Here we will address this question without actually fitting a power-law to power spectrum.

Our comparison exploits the distinctive behaviour of the AR1 and power-law models under *temporal aggregation* as is done when, for example, creating an annual mean time series based on January to December averages of a monthly mean time series. We define the temporally aggregated time series

$$X_j^{(T)} = \frac{1}{T}\sum_{i=1}^{T} X_{i+T(j-1)},\ j = 1, 2, \ldots,\ T \geq 1. \tag{6.1}$$

where $X_i,\ i = 1, 2, \ldots$ is the original time series. In this notation, $X_1^{(12)}$ would be the first value of an annual mean time series aggregated from the monthly mean time series $\{X_1, \ldots, X_{12}, X_{13}, \ldots\}$. Under temporal aggregation, for an AR1 time series with lag-one autocorrelation ϕ, the temporally aggregated time series has lag-one autocorrelation

$$\text{AR1:}\ \phi_{(T)} = \frac{\phi(1-\phi^T)^2}{T(1-\phi^2) - 2\phi(1-\phi^T)},\ 0 \leq \phi \leq 1 \tag{6.2}$$

In (6.2), $\phi_{(T)} = 0$ when $\phi = 0$, $\phi_{(T)} \to 1$ as $\phi \to 1^-$, and $\phi_{(T)} < \phi$ for $0 < \phi < 1$. The shape of $\phi_{(T)}$ as a function of ϕ is shown by the red curves in Fig. 6.1.

By contrast, temporal aggregation has no impact on a power-law stochastic process. More precisely, for a second order self-similar process, which can be regarded as the ultimate case of a power-law stochastic process, we find (Cox, 1984; Taqqu, 2002)

$$\text{Power-law:}\ \phi_{(T)} = \phi,\ 0 \leq \phi \leq 1,\ T \geq 1. \tag{6.3}$$

This property, that the autocorrelation is independent of time aggregation, is certainly rather counterintuitive for climate processes.

Eqns. (6.2) and (6.3) suggest a simple, and to our knowledge novel, test of the relative validity of the AR1 and power-law models: we examine the behaviour of the lag-one autocorrelation under temporal aggregation in comparison with these equations. The results for the SAT from the CMIP3 simulations are the most informative and are shown in Fig. 6.1. Comparison to observations is not straightforward and will be discussed at the end of this subsection.

Figs. 6.1a-c are scatter plots of the CMIP3 ensemble mean annual vs monthly autocorrelations ($\phi_{(12)}$ vs ϕ) for the linearly detrended SAT anomalies. Each point in these scatter plots

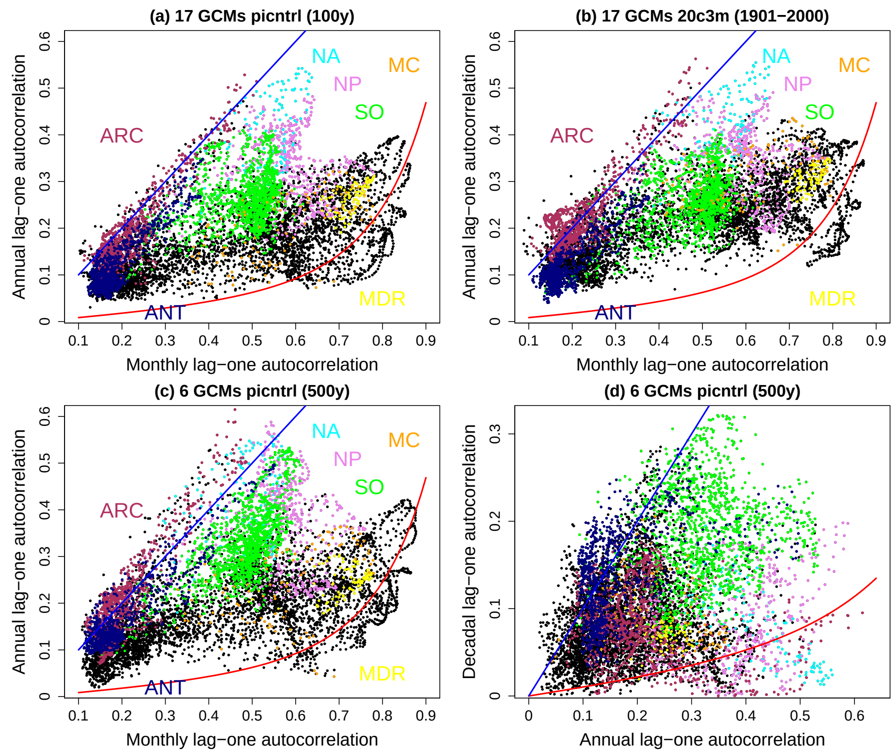

Figure 6.1: a) Scatter plot of $\phi_{(12)}$, which is the lag-one autocorrelation for the annual mean time series, vs ϕ, which is the lag-one autocorrelation for the monthly mean time series, for the linearly detrended SAT. The autocorrelations are ensemble averaged across the *picntrl* simulations of the 17 CMIP3 models. The blue line represents $\phi_{(12)} = \phi$ from (6.3) for the power-law statistical model; the red line represents $\phi_{(12)}$ as a function of ϕ from (6.2) for the AR1 statistical model. The dots are colour coded by region: "cyan": North Atlantic; "violet": North Pacific; "yellow": Main Development Region (MDR); "green": Southern Ocean; "orange": Maritime Continent; "maroon": Arctic; "navy": Antarctica; "black": the rest. b) As in a), for the *20c3m* simulations. c) As in a), for the 500 year *picntrl* simulations of the 6 CMIP3 models. d) As in c), for the decadal versus annual mean time series. Note the different scales used in d). The regions have the following boundaries: the North Atlantic (308°E-350°E, 40°N-60°N), the North Pacific (149°E-230°E, 20°N-57°N), the Southern Ocean (0°E-360°E, 40°S-65°S), MDR (299°E-332°E, 5°N-22°N), the Maritime Continent (98°E-158°E, 5°S-5°N).

represents a grid point on the Earth surface with colour coding for different regions (see the figure caption). The red and blue lines represent (6.2) and (6.3) respectively. Figs. 6.1a and b demonstrate the results for the 100 year long *picntrl* and *20c3m* simulations, with the ensemble mean including the 17 GCMs in those simulations. The annual autocorrelations are slightly larger for the *20c3m* scenario, but otherwise the two plots look quite similar. This suggests that beyond the influence of trends the origins of climate persistence on monthly to inter-annual time scales are internally generated rather than externally forced.

To verify the robustness of this analysis we repeat it for the six 500 year long *picntrl* simulations and plot the results in Fig. 6.1c, which also resembles Figs. 6.1a,b. Finally, we use these long integrations to compare decadal-mean and annual-mean autocorrelations (i.e. $\phi_{(10)}$ vs ϕ). Qualitatively this panel is similar to the first three, but the correlations are reduced overall on decadal scales.

In all four panels most of the points lie below the blue line and above the red line. Therefore, for the simulations, the AR1 model provides a lower bound and the power-law model an upper bound for climate persistence on intra-annual to inter-decadal time scales. The points from the Arctic are located closely to the blue line $\phi_{(12)} = \phi$ in Figs. 6.1a-c; strikingly, for the Arctic points, the annual mean time series has a similar lag-one autocorrelation to the monthly mean time series. But this behaviour, for which we have no simple explanation, does not extend to decadal means: most of the Arctic points fall below the blue line in Fig. 6.1d.

We have produced similar scatter plots for the SAT from the observational products; these are not shown. Although these plots are noisy due to a reduced ensemble averaging effect, qualitatively they look similar to Fig. 6.1 with a majority of the points located between the two curves. However there are noticeably more points located above the blue line and less dependence between annual and monthly autocorrelations, i.e. coloured regions tend to be organized in vertical stripes. This suggests that further study, involving a closer comparison between observations and simulations at individual points, is needed to determine if subannual time scale persistence is an effective predictor of annual to decadal scale persistence.

The fact that the AR1 and power-law models provide bounds for climate persistence on intra-annual to inter-decadal time scales and that the current standard practice in climate science is dealing only with the first of these bounds (e.g. Intergovernmental Panel on Climate Change, 2007; World Meteorological Organization, 2007) motivates us to consider in details the second bound.

6.3.2 Spatial patterns

We start our analysis of the power-law spectral approximation with an analysis of the Hurst exponent spatial distribution for the observed and simulated SAT for the second half of the 20th century. We first estimate H for the time scale range of 18 months to 45 years, which was the time scale range of focus in Chapter 5. Fig. 6.2a shows $\hat{H}$ of SAT calculated for the ERA40, NCEP/NCAR, and GISS datasets, and then averaged together. We refer to averaging the spatial distribution of $\hat{H}$ over different datasets or simulations as "ensemble" averaging. (The GISS SAT is spatially complete northward of 50°S; poleward of 50°S only the ERA40 and NCEP/NCAR data figure in the ensemble average). We generally see larger values of $\hat{H}$ at lower latitudes than at higher latitudes, and larger values of $\hat{H}$ over ocean than over land. We see a gradient in tropical Pacific and Atlantic oceans with larger values in the western part of each basin. Separate local maxima can be found in the North Pacific, North Atlantic and extratropical Southern Ocean. The main features in the average can be found in each observational product (see also Fig. 6.3).

Figs. 6.2b and 6.2c show the 18m-45y $\hat{H}$ for the model ensemble mean *20c3m* and *picntrl* SAT. (A single 45 year segment was used for each model). The fields are smoother than in Fig. 6.2a, primarily because of the ensemble averaging across the 17 models. The main features of the three panels agree, with significant discrepancies in parts of the Indian Ocean and in the southwestern Pacific, where $\hat{H}$ for the observations is greater than that for the simulations. Recall that for DFA3 for the frequency range of 18 months to 45 years $\sigma(\hat{H}) \approx 0.075$. Assuming that $\hat{H}$ for different climate models are not correlated, the standard deviation of the

17 models' ensemble mean $\hat{H}$ is $\sigma(\hat{H}_{CMIP3}) \approx 0.075/\sqrt{17} \approx 0.02$. It is harder to justify a similar assumption for the observational products and thus we obtain only a lower bound on $\sigma(\hat{H}_{Obs}) \approx 0.075/\sqrt{3} \approx 0.04$. Assuming independence of the observed and simulated $\hat{H}$ we have $\sigma(\hat{H}_{Obs} - \hat{H}_{CMIP3}) = \sqrt{\sigma(\hat{H}_{Obs})^2 + \sigma(\hat{H}_{CMIP3})^2} = \sigma(\hat{H})\sqrt{1/3 + 1/17} \approx 0.05$. Thus we can consider the differences between the observed and simulated $\hat{H}$ for the 18 months to 45 years range as approximately significant at the 2σ level if they are greater than 0.1. (For the frequency range of 5 to 45 years this threshold is 0.17.) Therefore as a rule of thumb, values of $\hat{H}$, when the observations and CMIP3 models are compared in Fig. 6.2, should not be considered significantly different when they differ by only a single contour level. More detailed regional comparisons between the models and observations will be done below in relation to Fig. 6.3.

The results for the *20c3m* and *picntrl* simulations are very similar, which underlines the origins of the spectral power increase with decreasing frequencies, characterized by H, in internal climate dynamics and which also suggests a minor role of natural external forcings (solar and volcanic) for the inter-annual and decadal SAT variability. This conclusion applies to regions where the climate models and the observations agree. The anthropogenic forcings obviously boost power in low-frequencies, but firstly their effects are filtered out by DFA3 and secondly they should be modeled deterministically, rather than stochastically, and therefore are not of interest in our study. We have also compared the spatial distribution of the $\hat{H}$ for the *20c3m* simulations which included natural forcings with those which did not. A significant difference between these two types of the simulations is found only over the Maritime Continent (not shown), where the simulations with natural forcings demonstrate larger $\hat{H}$ and which will be discussed below. The origin of this effect remains unknown and requires an additional research.

For the CMIP3 simulations, we can also test the sensitivity of $\hat{H}$ to the low-frequency cutoff by shifting it from 45 to 100 years. The results for the *20c3m* simulations for the time scale range of 18 months to 100 years, corresponding to simulations from 1900-1999, are shown in Fig. 6.2d. The average $\hat{H}$ decreases somewhat over this time scale range, indicating

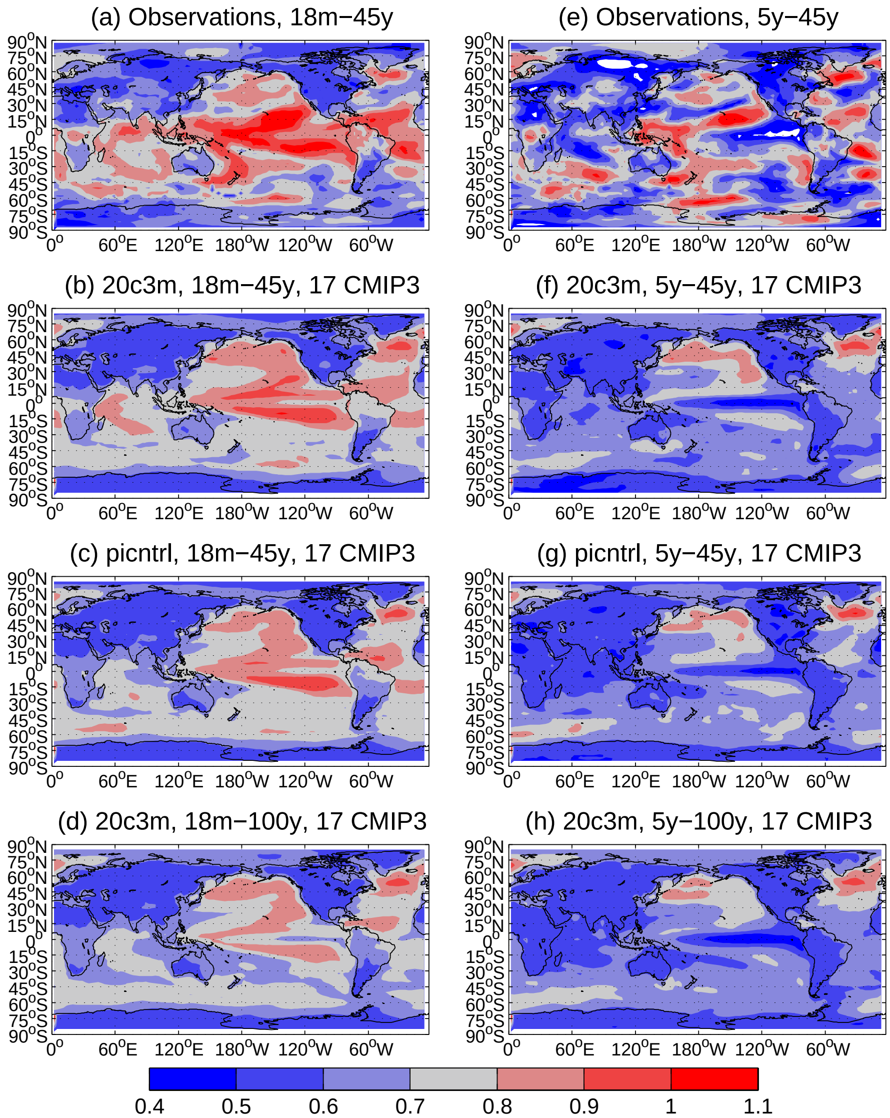

Figure 6.2: The DFA3 Hurst exponent estimates ensemble averaged across the NCEP/NCAR reanalysis, ERA40 reanalysis, and GISS SAT (a,e) and the 17 CMIP3 models (b-d,f-h). In the ensemble average, first the $\hat{H}$ is estimated for each observational product/model and then averaged across the observational products/models. The 18 months to 45 years time scale range is used for estimation of H in (a-c), 18 months to 100 years in (d), 5 to 45 years in (e-g), and 5 to 100 years in (h).

a somewhat shallower slope. We have repeated this for 100 year long *picntrl* runs from the 17 GCMs identified in Section 6.2 and 500 year long *picntrl* runs from the 6 GCMs set; $\hat{H}$ is almost indistinguishable in these cases. The similarity between Figs. 6.2c and 6.2d, which also holds for individual models, suggest that the main features of the $\hat{H}$ distribution can be extracted from 50 years of data. This implies that the current observational record is sufficiently long to characterize H on annual to multidecadal time scales in regions where the climate models and the observations agree.

Although the $\hat{H}$ distribution appears robust to changes in the low-frequency cutoff, it is quite sensitive to changes in the high frequency cutoff. Such a sensitivity was previously noticed by Fraedrich and Blender (2003); Blender and Fraedrich (2003); Blender et al. (2006), who showed that in climate simulations the spectral slopes in the tropical ocean were sensitive to where the high frequency cutoff is located. The right hand column of Fig. 6.2 is the same as the left hand column of Fig. 6.2, but with the short time scale cutoff changed from 18 months to 5 years. In the observed SAT, a sharp drop in $\hat{H}$ occurs in the eastern tropical Pacific and Atlantic and in the central Indian ocean (see Fig. 6.2e). In the eastern Pacific, $\hat{H} < 0.5$, indicating positive spectral slope over the 5 to 45 years time scale range. The high $\hat{H}$ values remain robust over the extratropical oceans.

In the CMIP3 models (Figs. 6.2f-h), $\hat{H}$ is also sensitive to the high-frequency cutoff. The large drops in $\hat{H}$ over the eastern tropical Pacific and Atlantic are seen as in observations, thus cross-validating the observational finding. But the overall $\hat{H}$ for this lower frequency band is biased low compared to the observations, particularly in the Southern Hemisphere, tropical Atlantic and over the Maritime Continent. Kravtsov and Spannagle (2008) have documented discrepancies between the multidecadal variability of regionally averaged observed and CMIP3-simulated SAT. Thus some of the regions for which they found that the models noticeably underestimate natural climate variability, for instance the tropical Atlantic, are the regions of significant disagreement between the observed and modeled $\hat{H}$ (see Figs. 6.2e-h). Kravtsov and Spannagle (2008) suggest that these discrepancies might be related to inability of the CMIP3 ensemble mean to reproduce the observed Atlantic Multidecadal Variability (AMV). We will return to one of these discrepancies during discussion of Fig. 6.3e. $\hat{H}$ in Figs. 6.2g-h is insensitive to the character of the external forcings, to the shift of the low frequency cutoff from 45 to 100 years, and to the shift of the high frequency cutoff from 5 to 7 years (not shown).

The spatial distributions of the CMIP3 model and observational ensemble mean of the monthly SAT lag-one autocorrelation (not shown) are qualitatively similar to the corresponding spatial distributions of the $\hat{H}$ estimated for the range of 18 months to 45 years. The same is true for the spatial distributions of the annual SAT lag-one autocorrelation (not shown) and the $\hat{H}$ estimated for the range of 5 to 45 years. This fact probably means that a) different climate processes are dominant at different time scales, because the estimates of the memory parameters, $\hat{\phi}$ and $\hat{H}$, depend on the aggregation time scale and the high-frequency cutoff respectively, and b) the lag-one autocorrelation and the Hurst exponent provide different viewpoints on the same phenomenon.

The comparison between the models and observations in Fig. 6.2 is incomplete because the model ensemble mean does not always represent individual models and because the confidence in estimates of H varies among regions and decreases for narrower frequency ranges. To assist in making a more informative comparison we present in Fig. 6.3 the average $\hat{H}$ over several regions, grouped by frequency range, and within each range split into individual *picntrl* simulations, observational products, and individual *20c3m* simulations. The boundaries of the regions are listed in the caption of Fig. 6.1.

The confidence intervals were obtained in the following way. Firstly, for each time scale range we calculated the standard deviation of DFA3 H estimate (averaged across values of H between 0.4 and 1.1), $\sigma(\hat{H})$, using 8x10,000 synthetic time series as in Section 2.3. We averaged across the several values of H because it has been shown (e.g. Taqqu et al., 1995) that $\sigma(\hat{H})$ very weakly depends on H. Secondly, in order to take into account the effect of spatial averaging we divided $\sigma(\hat{H})$ by the square root of the number of grid points in each region

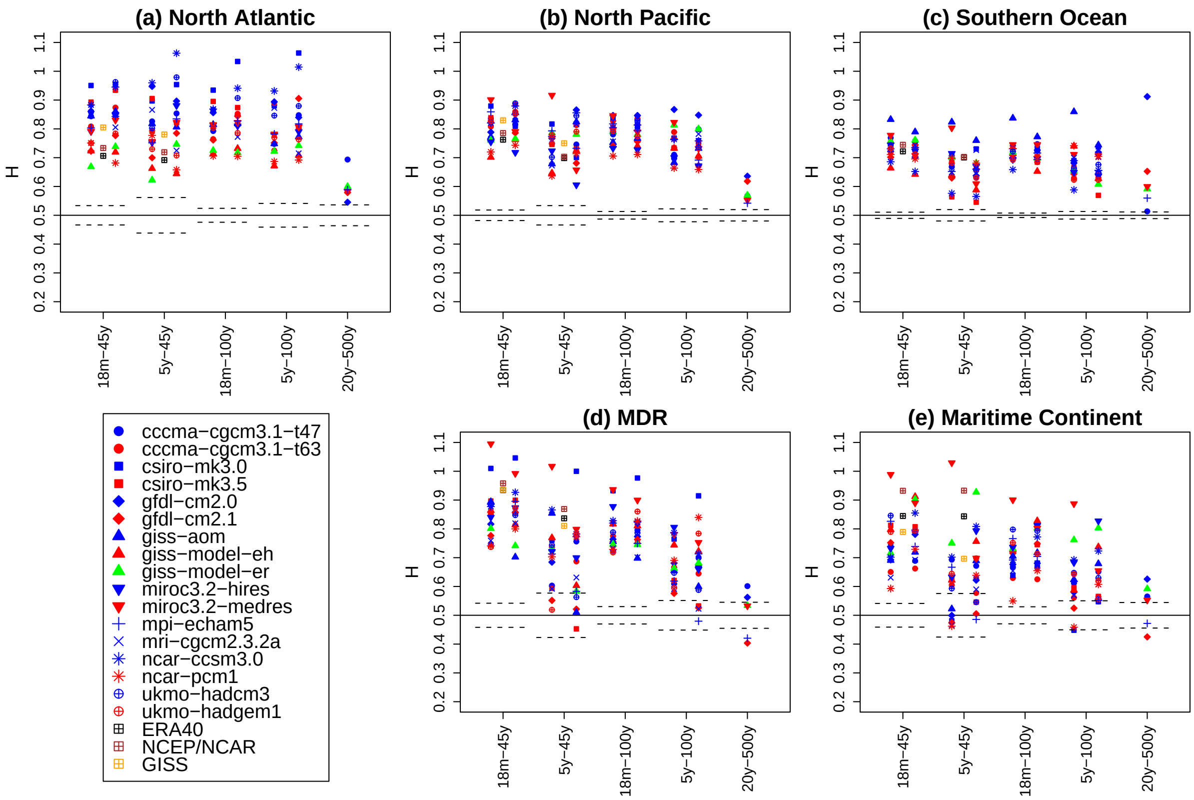

Figure 6.3: The averages of the DFA3 $\hat{H}$ for the various regions and time scale ranges. For the first four time scale ranges the left column represents the results for the *picntrl* and the right column for the *20c3m* simulations. The observations are shown in between those two columns. The averaged $\hat{H}$ for the GISS SAT is not estimated for the Southern Ocean due to its poor coverage. For the 20 to 500 years range only *picntrl* simulations of the six climate models are available. The horizontal dashed lines demonstrate the 2σ confidence intervals for the $\hat{H} = 1/2$. The confidence intervals for $\hat{H} > 1/2$ are roughly similar. See the text for their description. The region boundaries are given in the caption of Fig. 6.1.

divided by zonal and meridional decorrelation scales of SAT. In this procedure we made rough assumptions that temperature spatial correlations decay exponentially and that decorrelation scales of SAT are equal to decorrelation scales of H. We also roughly assumed that the zonal (meridional) decorrelation scale is equal to 3 (2) grid points, which approximately corresponds to 9° (5°) (Molinari and Festa, 2000; Romanou et al., 2006). The obtained effective standard deviations of $\hat{H}$ were multiplied by ± 2, added to 1/2, and plotted as horizontal dashed lines in Fig. 6.3.

For the North Atlantic, the models have just slightly larger estimates of H than the observational products, with CSIRO-Mk3.0 and NCAR CCSM3.0 having the largest estimates and NCAR PCM, GISS-EH, GISS-ER having the smallest estimates. The $\hat{H}$ for the North Atlantic are robust to changes of the high and low-frequency cutoffs within the 18 months to 100 years time scale range. The spectral power grows at a slower rate on multidecadal to centennial time scales for the *picntrl* simulations as demonstrated by the smaller $\hat{H}$ for the 20 to 500 years range.

The models show even better agreement with the observational products over the North Pacific, with GFDL-CM2.0 and MIROC3.2(medres) at the top and NCAR PCM, GISS-EH and MIROC3.2(hires) at the bottom of the model $\hat{H}$ distribution. Here the estimates for the 5 years cutoff are somewhat smaller than for the 18 months cutoff. The distributions of the $\hat{H}$ for the Southern Ocean have a similar spread to the North Pacific, despite larger area, with GISS-AOM being always at the top and GISS-EH, NCAR CCSM3.0, and CSIRO-Mk3.5 typically at the bottom. The agreement with the reanalyses is also quite good. As for the North Pacific the $\hat{H}$ with the 5 years high-frequency cutoff are slightly less than those for the 18 months cutoff. The spreads over the North Pacific and the Southern Ocean are smaller than the spread over the North Atlantic at least partially due to larger areas of the former regions as reflected by the narrower confidence intervals.

As discussed before in the context of Fig. 6.2, $\hat{H}$ is very sensitive to the high-frequency cutoff on time scales between 18 months and 100 years in the tropics, identified by the Main

Development Region (Northern Tropical Atlantic) and the Maritime Continent in Fig. 6.3. Fig. 6.3d demonstrates a drop of the model ensemble mean $\hat{H}$ from about 0.9 (strong long-memory) for the 18 months high-frequency cutoff to about 0.65 (weak long-memory) for the 5 years cutoff. The observational products show a smaller drop (from 0.95 to 0.85) and strong long-memory for both cutoffs, which is consistent with conclusions of Kravtsov and Spannagle (2008) that CMIP3 models underestimate natural multidecadal climate variability in the Main Development Region. The spread between the models in the tropics is larger than over the extratropical oceans, which perhaps represents the consequence of inconsistency in representing interannual (multidecadal) variability related to ENSO (AMV) among the models (e.g. AchutaRao and Sperber, 2006; Kravtsov and Spannagle, 2008) as well as relatively small areas of the considered in Fig. 6.3d,e regions and strong spatial correlations in those regions (Molinari and Festa, 2000; Romanou et al., 2006).

The Maritime Continent has a similar spread between the models as the Main Development Region but a smaller drop in the $\hat{H}$ for the 5 years high-frequency cutoff. However it exhibits the largest spread between the observational products, especially for the 5 to 45 years range. For this range there is also a large disagreement between the models and observations in the two considered tropical regions. For the Maritime Continent the disagreement could be related to data inhomogeneities in the observations or problems with reanalyses models' parameterizations, which are probably manifested in the large spread between different products for the 5 to 45 years range, and difficulties in modelling this region (see e.g. Neale and Slingo, 2003). In addition, the Maritime Continent is the only region in Fig. 6.3 for which the model mean $\hat{H}$ for the *20c3m* simulations (the right column for the first four time scale ranges in Fig. 6.3) is consistently larger than for the *picntrl* simulations (the left column). This might be a regional effect of natural forcings: those models whose *20c3m* simulations included natural forcings demonstrated a consistently larger $\hat{H}$ over the Maritime Continent than those whose simulations did not include natural forcings (not shown).

The model ensemble mean $\hat{H}$ based on the *picntrl* simulations for the 20 to 500 years range is about 0.5 for the two tropical regions, which contradicts to the results of Huybers and Curry (2006), who found significantly steeper slopes for this time scale range in paleo-proxies. For the MDR and the Maritime Continent MIROC3.2(medres) and GISS-EH almost always show the highest exponents and NCAR CCSM3.0 and NCAR PCM the lowest. This might be related to the frequency and the amplitude of the ENSO peak in the corresponding model power spectra (see e.g. AchutaRao and Sperber, 2006).

One of the open questions in climate science is the interaction between the tropical Pacific and the extratropical North Pacific. There are at least two possible hypotheses for generating interdecadal variability in the North Pacific. The first one proposes that this variability is generated locally by coupled atmosphere-ocean feedbacks (e.g. Latif and Barnett, 1996). The second hypothesis suggests that it is forced by a tropical multidecadal variability (e.g. Deser et al., 2004). The values of the CMIP3 $\hat{H}$ for the 5 to 100 years ranges are typically higher for the North Pacific than for the Maritime Continent and especially for the Nino3 region (not shown). This fact underlines the importance of extratropical dynamics for generating or at least modulating the interdecadal variability in the North Pacific, provided the CMIP3 models correctly capture this variability.

6.3.3 Comparison to previously published results

Our results for the CMIP3 simulations are largely consistent with previously published results, to within methodological differences. For example, the results for the 18 months to 45 years time scale range are within the error bars of those for 1 to 5 year time scale range for the NCEP/NCAR reanalysis and 1 to 15 year time scale range for a 1000 year long ECHAM4/HOPE control run and IS92a (business as usual) global warming runs of HADCM3 and ECHAM4/OPYC reported by Fraedrich and Blender (2003); Blender and Fraedrich (2003). In addition our results for the 5 to 45 years time scale range seem to be consistent in the extratropics with the ones for the 5 to 40 year time scale range for a 10,000 year long run of CSIRO-Mk2 analyzed by Blender et al. (2006). In the tropics their H estimates are larger than ours, especially over the tropical Pacific and Atlantic. The $\hat{H}$ for the 18-220 years time scale range for a 1000 year long control run of ECHO-G model studied by Rybski et al. (2008) are consistent with our results for the 20 to 500 years time scale range, except in two regions. In the North Atlantic Rybski et al. (2008) report the values of $\hat{H} > 0.72$ and in the Southern Ocean their values are somewhat consistent with the anomalous results we get for GFDL-CM2.0. Rybski et al. (2008)'s results agree with the results of Fraedrich and Blender (2003) for the 15 to 150 year time scale range of the ECHAM4/HOPE 1000 year long control run. We note that Fraedrich and Blender (2003); Blender and Fraedrich (2003); Blender et al. (2006); Rybski et al. (2008) used DFA2 in their studies. The models they analyzed belong to the previous generation and are not part of the CMIP3 archive.

A linear regression of a logarithmically binned logarithm of a multitaper spectral estimator against a logarithm of the frequency was employed by Huybers and Curry (2006) to estimate H. We have found that their H estimates for the NCEP/NCAR reanalysis for the 2 month to 30 year time scale range are larger than ours over the tropical oceans because of the smaller high-frequency cutoff they implement (comparison not shown). As we have shown above and as can be seen in figures of Fraedrich and Blender (2003); Blender and Fraedrich (2003); Blender et al. (2006), the $\hat{H}$ over the tropical oceans is larger for smaller high-frequency cutoff time scale.

Although the simulated and observed $\hat{H}$ agree over large regions on annual-to-decadal timescales, we find suggestions that the simulated $\hat{H}$ is consistently smaller than observed on multidecadal to centennial time scales. In particular, for the time scale range of 20 to 500 years the $\hat{H}$ is typically between 0.5 and 0.7 for the extratropical oceans according to the six models' *picntrl* simulations, which manifests weak long-memory behaviour on those time scales. But paleo-climate reconstructions consistently exhibit large H on these time scales. For example, Pelletier (1997) reported $\hat{H} = 0.75$ for the Vostok Antarctic ice core on the time scale range between several decades and several millennia, Blender et al. (2006) estimated the $\hat{H} = 0.7$ for GRIP and the $\hat{H} = 0.84$ for GISP2 Greenland ice cores for the range between around 30 and 1000 years, and Rybski et al. (2006) showed that various reconstruction of the Northern Hemisphere SAT have $\hat{H} > 0.8$ for the range between a decade and several hundred years. Huybers and Curry (2006) demonstrated a transition from the values of $\hat{H}$ below 1.0 to values greater than 1.0 at around a 100 years time scale for several paleo-proxies. Pelletier (1997) also documented such a transition for the Vostok Antarctic ice core, but at around the several thousand years time scale. There are several processes missing from the *picntrl*, such as the natural radiative forcings (solar and volcanic), glacial dynamics, and interactions with the biosphere, which might boost the H in the simulations. For now, we conclude that these simulations provide a lower bound for H on multidecadal, centennial, and longer time scales. However potential problems with paleo-climate reconstructions, such as data inhomogeneities, which could increase low-frequency time series variability and therefore H, cannot be completely excluded.

6.3.4 Goodness of fit tests of power-law and AR1 models

We now compare the performance of the AR1 vs power-law model in terms of a spectral goodness-of-fit test (Milhoj, 1981; Beran, 1992). This test estimates a standardized overall measure of the discrepancy between the periodogram and the fitted spectrum (e.g. AR1 or power-law). The spectral goodness-of-fit test also provides an approximate p-value, i.e. the probability of obtaining a deviation from the fitted spectrum at least as extreme as the one that was actually observed. Percival et al. (2001) applied this test to North Pacific atmospheric variability and found that neither model was clearly superior. Their conclusion was that the observational record was too short to distinguish between the two models. In contrast, we applied this test to the three century Central England Temperature time series in Section 1.1 and found $p = 0.67$ for the power-law model and $p = 0.2$ for the AR1 model, which favors the power-law model as a significantly superior fit to this time series.

Here we extend these results by applying the spectral goodness-of-fit test to SAT in the CMIP3. The power-law fit to the periodogram is based on GSPE and the AR1 spectral density

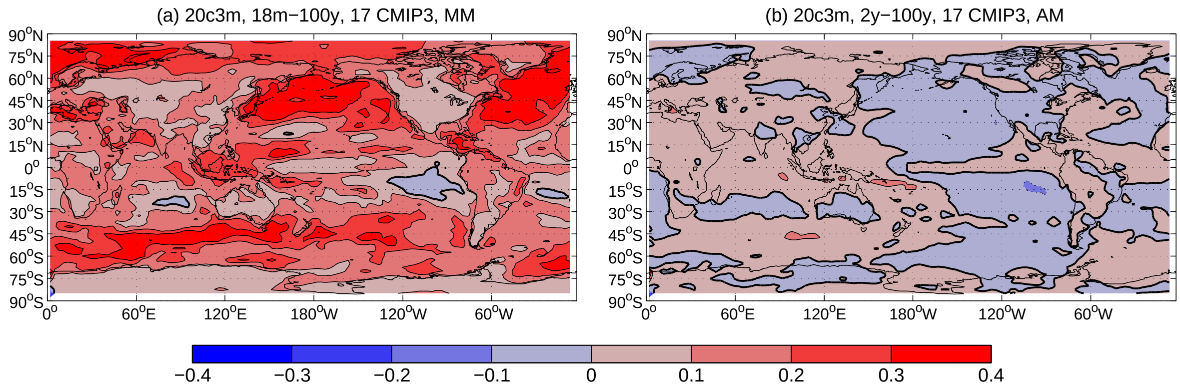

Figure 6.4: The spectral goodness-of-fit test p-value for the power-law fit minus the p-value for the AR1 fit for the *20c3m* CMIP3 simulations. Panel (a) shows the results for the monthly means and (b) for the annual means. The time scale range of 18 months to 100 years have been used for the power-law fitting in panel (a) and 2 to 100 years in (b). All the available frequencies, 1/(2 months) to 1/(100 years) in panel (a) and 1/(2 years) to 1/(100 years) in panel (b), have been employed for fitting the AR1. The results have been averaged across the 17 CMIP3 climate models. The power-law fit is superior (inferior) in the red (blue) areas.

is also fitted to the periodogram using a maximum-likelihood approach (Section 6.2). Here we use GSPE, because in contrast to DFA it allows us to use all available frequencies. (DFA3 has a limitation that its short time scale cutoff should be greater or equal than 18 time units (see Section 2.2.2).) In Fig. 6.4a, we fit the AR1 model to the linearly detrended *20c3m* SAT, calculate p_{AR1} for the AR1 model for each grid point and each GCM, and take the model ensemble mean. We calculate the ensemble average $p_{power-law}$ in a similar way for the 18 months to 100 years power-law model and plot in Fig. 6.4a the difference $p_{power-law} - p_{AR1}$. We see that this difference is positive almost everywhere, especially over the extratropical oceans. Thus the spectral goodness-of-fit test favors the power-law model in this application or in other words the power-law model is more strongly supported by the data under consideration than the AR1 model.

But we note in Fig. 6.4a that there is an inconsistency in the time scale range used in fitting the two models: the AR1 model is fit from 2 months to 100 years, while the power-law model is fit from 18 months to 100 years. We considered this example because it is a standard practice, for instance in studies of total ozone trends (see e.g. World Meteorological Organization, 2007), to fit the AR1 model to monthly mean residuals of a regression model. Fig. 6.4a confirms conclusions of Chapter 3 that the power-law model fitted at low-frequencies provides a better or similar quality fit at low-frequencies than the AR1 model fitted to all available frequencies and thus the former is preferable for trend confidence intervals estimation. When we use consistent frequency ranges, which puts the two statistical models on an equal footing, the results change. In particular, if we fit the two models to the linearly detrended *annual mean* SAT time series over the same time scale range of 2 to 100 years, the two models show equal performance for the *20c3m* (Fig. 6.4b) and the *picntrl* (not shown). This approach is consistent with previous applications of such comparisons (e.g. Percival et al., 2001). The similar performance of the two models, when consistent frequency ranges are used, meshes with the qualitative impression from Fig. 6.1, which shows that persistence in SAT falls about midway between the AR1 and power-law models. This behaviour extends similarly to other time scale ranges and to the *picntrl* integrations, including the decadal-to-centennial range in the 500 year long *picntrl* integrations (not shown).

In summary, from the temporal aggregation analysis (Fig. 6.1) and the spectral goodness-of-fit test (Fig. 6.4), we have reached a key conclusion: there is no objective evidence that the power-law model is superior to the AR1 model in the CMIP3 simulations on interannual to multidecadal timescales. Instead, we see a behaviour in the simulations that falls between the two statistical models, showing that neither provides a complete description of natural climate variability. This opens a possibility that high order models, e.g. autoregressive models of order greater than one, might provide a better fit. The significance of the above mentioned conclusions depends in part on the match between the simulations and observations. In our evaluation, the models do sufficiently well in regions like the North Pacific and the North Atlantic to trust that this conclusion would also apply to the real climate system. Thus, Percival et al.'s (2001) conclusion that AR1 and power-law models perform comparably for the North Pacific circulation might not depend on the length of the observed record but on the fundamental character of natural climate variability in that region. But this conclusion could very well change on longer time scales as Earth System models are developed that properly capture biosphere-climate and cryosphere variability, given the evidence of power-law like behaviour in paleo reconstructions (e.g. Pelletier, 1997; Huybers and Curry, 2006).

6.4 Results for the free atmosphere air temperature

In Chapters 4 and 5 we analyzed observed and simulated zonal mean free atmospheric temperature $\hat{H}$ for the time scale range of 18 months to 45 years; the simulations in Chapter 5 included specialized atmospheric general circulation model simulations and the CMIP3 simulations. In this section, we further explore the CMIP3 simulations, in the context of the findings of the previous section.

The first row of Fig. 6.5 plots the $\hat{H}$ for zonal mean air temperature for the time scale range of 18 months to 100 years. Panel (a) is similar to Fig. 5.1f, which showed the results for the *20c3m* simulations for the range of 18 months to 45 years. This is consistent with the results for the SAT, which are robust to the low-frequency cutoff shift from 45 to 100 years. The tropospheric part of the $\hat{H}$ distribution for the *picntrl* simulations (Fig. 6.5b) is similar to Fig. 6.5a. This agrees with the conclusions of Chapter 5 that the tropospheric structure of the $\hat{H}$ is caused by the internally generated tropical SST variability.

The tropical lower stratosphere maximum in $\hat{H}$ was shown by means of linear regression in Section 4.4 and by a volcanic forcing GCM simulation in Section 5.2 to be caused by the volcanic forcing. The same general effect is found in the CMIP3 models: we plot in panel (c) the difference between the $\hat{H}$ for the nine *20c3m* simulations that had historical volcanic forcing and the eight *20c3m* simulations that did not. This difference resembles the response of the atmospheric GCM experiment forced by climatological SSTs and a historical volcanic

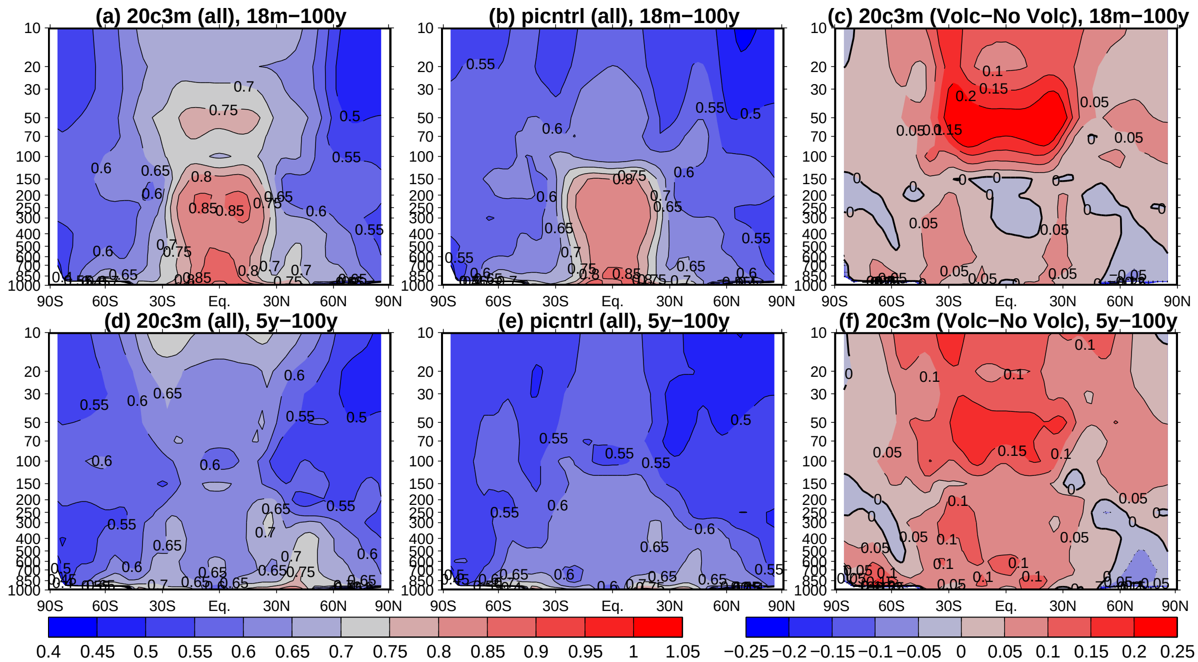

Figure 6.5: The DFA3 Hurst exponent estimates for the 17 CMIP3 models (a,b,d,e). The $\hat{H}$ is estimated for each model first and then averaged across the models. The 18 months to 100 years time scale range is used for estimation of H in (a-c), 5 to 100 years range in (d-f). Panels (c) and (f) show the difference of the $\hat{H}$ between nine models with and eight models without historical volcanic forcings.

forcing seen in Fig. 5.1e.

As for the SAT we test the sensitivity of the estimated H to a change of the high-frequency cutoff from 18 months to 5 years. The results are shown in the second row of Fig. 6.5. One can immediately notice a large drop of the $\hat{H}$ in the tropical troposphere for both scenarios (see also Fig. 6.6e). In principle such a drop could be anticipated given the results for the SAT for this time scale range, which show the values of the ensemble mean $\hat{H} < 0.7$ in the tropics (see Fig. 6.2h and Fig. 6.3d,e), and the conclusions of Chapter 5 that the tropospheric distribution of the $\hat{H}$ is controlled by the tropical SSTs. The difference of the $\hat{H}$ for the range of 5 to 100 years between simulations with and without volcanic forcings (see Fig. 6.5f) is slightly weaker (larger) in the stratosphere (troposphere) than in Fig. 6.5c (see also Fig. 6.6b).

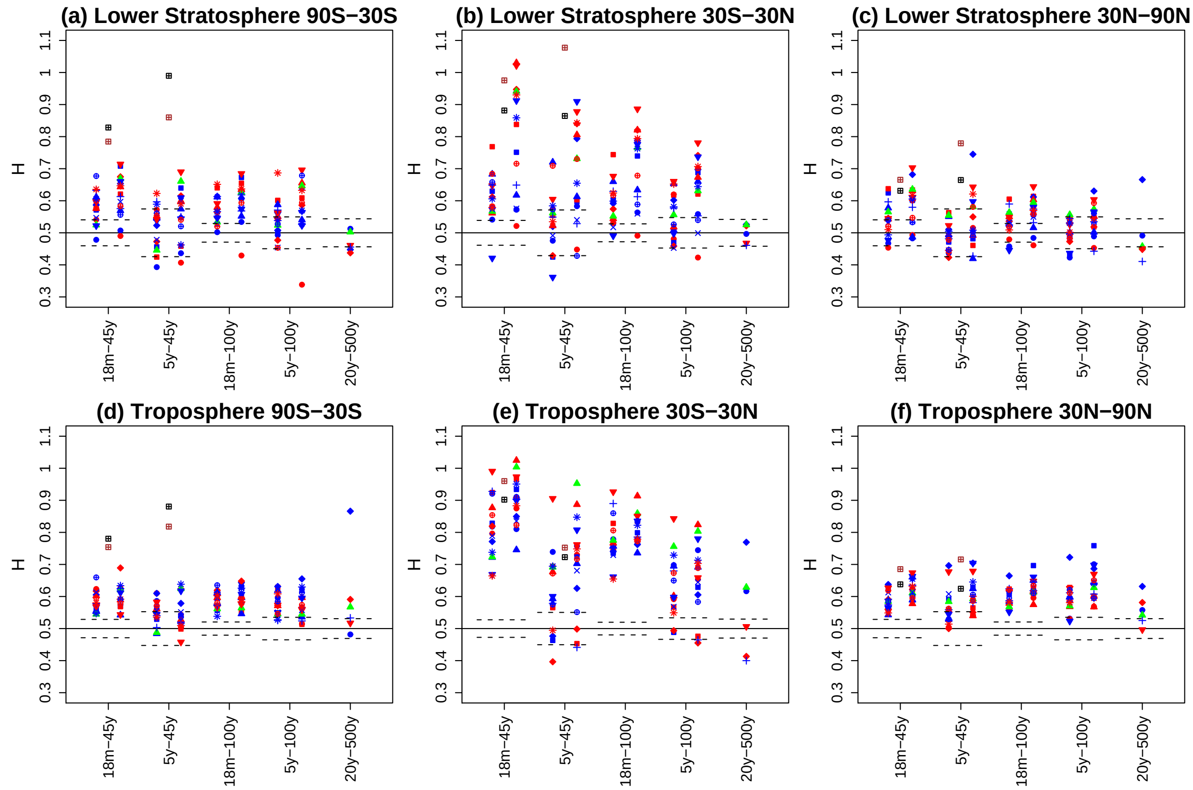

Figure 6.6: As Fig. 6.3, but for the six regions in the free atmosphere. For the confidence intervals estimation we roughly assumed that the meridional (vertical) decorrelation scale is equal to 2 (4) grid points.

The stratospheric response is again consistent with the volcanic forcing GCM simulation in Section 5.2. However the tropospheric response is somewhat different from what we expected based on the linear regression and on the atmospheric GCM simulation results. Because almost all the CMIP3 GCMs that included historical volcanic forcings also included historical solar forcings, the tropospheric response could be related at least partially to the latter, but more specific studies are needed to identify the possible link. As for the SAT, the spatial distributions of the corresponding monthly and annual lag-one autocorrelation for the zonal mean air temperature are qualitatively similar to the first and second row of Fig. 6.5 respectively and therefore we do not show them.

We find the format of Fig. 6.3 useful for model and data intercomparison and apply it to study the spatial distribution of the $\hat{H}$ for the zonal mean air temperature (see Fig. 6.6). We split the area of our study into six regions: the extratropical Southern Hemisphere troposphere (90^oS-30^oS, 1000hPa-100hPa), the tropical troposphere (30^oS-30^oN, 1000hPa-100hPa), the extratropical Northern Hemisphere troposphere (30^oN-90^oN, 1000hPa-100hPa), and three regions with the same latitude boundaries, but between 100 and 10hPa in the lower stratosphere. Together with the models we plot the results for the NCEP/NCAR and ERA40 reanalyses for two time scale ranges. In contrast to Fig. 6.3 the right columns for the first four time scales ranges, corresponding to the *20c3m* scenario, show larger values than the left columns (*picntrl* scenario) for all regions with the exception of the southern troposphere extratropics. That is, the natural radiative forcings affect $\hat{H}$ more strongly in the free atmosphere than at the surface. Inconsistencies in these forcings might explain why the simulated $\hat{H}$ exhibits a large spread in the *20c3m* simulations. Also in contrast to the SAT results there is much less agreement between the reanalyses and the models, especially in the data poor regions. For instance the reanalyses $\hat{H}$ are noticeably larger than any GCM for both time scale ranges in the Southern Hemisphere extratropics and than almost all GCMs for the 5 to 45 years range for the tropical and the Northern Hemisphere stratosphere.

In those cases when reanalysis $\hat{H}$ is significantly larger than the models $\hat{H}$ we have found that there are often data inhomogeneities in the reanalysis; such inhomogeneities tend to increase $\hat{H}$ (see Berton (2004); Rust et al. (2008) and Section 5.3). For instance, in the tropical lower stratosphere the NCEP/NCAR reanalysis $\hat{H}$ for the 5 to 45 years range is much larger than that of ERA40 and all the models probably due to discontinuities in this product around 1979 related to the inclusion of satellite data (see e.g. Pawson and Fiorino, 1999). In any case, the large values of $\hat{H}$ manifest the presence of low-frequency variability, either natural or induced by data inhomogeneities, which makes the task of trend detection more difficult, for instance in the tropics (see e.g. Santer et al., 2005). The tropics are data poor and hard to model (see e.g. Neale and Slingo, 2003) and thus the largest spread between the models and between the two reanalyses in the free atmosphere and at the surface is found there.

Overall the models' $\hat{H}$ are between 0.4 and 0.8 for the extratropics, but vary a lot in the tropics. The models that are consistent with the reanalyses in the tropical lower stratosphere for the 18 months to 45 years range (see Fig. 6.6b) are the models which include historical volcanic forcings. These models also have noticeably larger $\hat{H}$ values for the *20c3m* than for *picntrl* simulations for the 18 months to 100 years and 5 to 100 years ranges in agreement with Fig. 6.5f. One the largest spreads between the models is found in the tropical troposphere for the 20 to 500 years range for the *picntrl* simulations, although in the tropical stratosphere all the six models cluster narrowly around 0.5 for this time scale range. Some of the models are consistently at the top or at the bottom of the $\hat{H}$ distribution. Thus MIROC3.2(medres) is usually the model with the largest values of $\hat{H}$. The two versions of the Canadian model typically have one of the lowest values of $\hat{H}$ in the extratropical lower stratosphere. GFDL-CM2.0 has one of the largest values in the extratropical troposphere. Probably due to their local sensitivity to the radiative forcings some of the models, for instance NCAR CCSM3.0 in the southern extratropical troposphere, have one of the lowest $\hat{H}$ for the *picntrl* scenario, but one of the largest for the *20c3m* scenario. It is also interesting that two versions of the same model, for example MIROC3.2(medres) and MIROC3.2(hires), can be located at opposite ends of the distribution for a particular region. Thus model resolution might exert a strong control on annual to decadal scale variability.

6.5 Conclusions

In this work we have systematically studied the power-law approximation of the temporal power spectrum of the surface and free atmosphere air temperature, characterized by the Hurst exponent, H. We have analyzed several observational products (NCEP/NCAR and ERA40 re-analyses and GISS SAT) and simulations under two scenarios of the 17 CMIP3 global climate models. We have varied the time scales on which the power-law was fitted to the spectrum from 18 months to 500 years. We have verified our results with an independent Hurst exponent estimation method.

At the surface for the time scale range of 18 months to 45 years all data sets show the largest values of the $\hat{H}$, i.e. the fastest rate of the spectral power buildup with decreasing frequency, in the tropics. The values of the $\hat{H}$ significantly larger than 0.5 (a flat spectrum case) are observed mainly over the ocean in agreement with other studies (Fraedrich and Blender, 2003; Blender and Fraedrich, 2003). The results remain robust when we increase the low-frequency cutoff from 45 to 100 years, but changing the high-frequency cutoff from 18 months to 5 years leads to a significant drop in $\hat{H}$ to $\hat{H} < 0.7$ everywhere but three regions: the North Atlantic, the North Pacific and the Southern Ocean. These regions were also identified in the previous case studies of specific models, scenarios, and Hurst exponent estimation methods (Fraedrich and Blender, 2003; Blender and Fraedrich, 2003; Blender et al., 2006; Rybski et al., 2008).

The results for the pre-industrial control and the 20th century simulations are remarkably similar, which points to internal climate mechanisms generating the growth of the spectral power on annual to multidecadal time scales at the surface. We think these mechanisms might be related to positive climate feedbacks, such as a wind-stress curl feedback in the North Pacific, which have been shown to increase the power in low-frequencies (Schneider et al., 2002).

For the range of 20 to 500 years the $\hat{H}$ lies between 0.5 and 0.7 for the *picntrl* simulations even in the North Atlantic, the North Pacific and the Southern Ocean, which corresponds to an absent or to a weak long-memory behaviour. However the results for paleo-proxies, covering several past centuries and millennia, show the values of $\hat{H}$ greater than 0.7 for decadal to centennial time scales. Provided decadal to centennial variability in the paleo-proxies is reliable, this indicates that either other missing mechanisms, possibly solar and volcanic forcings, glacial dynamics, interactions with the biosphere, or other unknown physical processes, have stronger input to the spectral power growth on centennial scales or that the representation of certain processes in the studied GCMs is deficient. More detailed studies, (e.g. Zhu et al., 2006; Zhu and Jungclaus, 2008), of the mechanisms responsible for this growth in specific regions for specific time scales are definitely needed, because they will reveal the origins of natural climate variability, which is often underestimated.

We have compared the validity of the power-law vs the AR1 model and their goodness-of-fit for the SAT time series. We have introduced a new method for time series model validation based on temporal aggregation. We have found that the estimates are clustered between the two statistical model predictions, which firstly means that neither model fits the time series perfectly on monthly to inter-decadal time scales, and secondly means that the power-law model might serve as an upper bound and the AR1 model as a lower bound on SAT persistence. This is an important conclusion for trend detection, because typically trend confidence interval (CI) and therefore trend significance in climate research is estimated solely under the AR1 model assumption for the residuals (see e.g. Trenberth and Coauthors, 2007; World Meteorological Organization, 2007) and thus is probably underestimated.

For illustration we estimate the linear trend CI for the GISS Northern Hemisphere land SAT annual mean anomalies (Hansen et al., 2001), which is the third item in Table 3.2 in (Trenberth and Coauthors, 2007), for the period 1901-2005. Our autocorrelation estimate for the linear trend residuals is $\hat{\phi} = 0.63$, whereas GSPE $\hat{H} = 0.95$ in case all the available frequencies, 1/(2 years) to 1/(105 years), are used for the power-law fit. Based on these numbers our estimate of the trend 90% CI is $\pm 0.24^o$C per century for the AR1 and $\pm 0.42^o$C per century for the power-law model. The IPCC estimate of the 90% AR1 CI is $\pm 0.25^o$C per century. Therefore the power-law CI is almost two times larger than the AR1 CI. The GISS Northern Hemisphere land SAT linear trend, which IPCC estimate is 0.83°C per century, is significant even relative to the power-law CI. However this might not be the case for other climatic time series.

We have also employed both statistical models in stratospheric ozone trend analysis in Chapter 3. We thus recommend that power-law based CI be included along with AR1 based CI in trend detection work. For this purpose we have developed the open-source R package, PowerSpectrum, http://www.atmosp.physics.utoronto.ca/people/vyushin/PowerSpectrum_0.3.tar.gz. The package also provides various estimators of power and cross-spectrum with their CIs, several estimators of H, the spectral goodness-of-fit test, Monte-Carlo tests of the Hurst exponent estimators and the goodness-of-fit test, etc. The above mentioned concerns about the ubiquity of the AR1 model in climate research are also related to predictability studies (e.g. Boer, 2004) and extreme value statistics (e.g. Bunde et al., 2005).

We have also applied the spectral goodness-of-fit test (Beran, 1992) to compare the performance of the power-law vs the AR1 model. This test favours the power-law model when the two models are compared over a low-frequency range to which the power-law model is specifically fitted and the AR1 model is fitted to all the available frequencies. Thus the spectral goodness-of-fit test supports the suggestion that a trend CI should be estimated using the power-law model, because for its estimation only a low-frequency behaviour is important (Smith, 1993). However when both time series models are fitted to and compared over all the available frequencies they score equally, which is consistent with the results of the novel method for time series model validation we have described above.

In Chapter 5 we showed using specialized simulations of a GFDL atmospheric model that steep power-spectra for the 18 months to 45 years range are produced in the tropical troposphere by the internal atmosphere-ocean interaction and in the tropical lower stratosphere by the volcanic eruptions. In the extratropical atmosphere the spectra were found to be relatively flat. The analysis of the tropospheric and lower stratospheric zonal mean temperature derived from the CMIP3 simulations confirms the robustness of our previous findings. However it is established that for the range of 5 to 100 years the Hurst exponents noticeably decrease in the tropics. The distributions of the CMIP3 Hurst exponents for this time scale range in the tropical stratosphere and troposphere (shown in Fig. 6.6b and e) are similar to the distribution for the same range for the Main Development Region and the Maritime Continent (shown in Fig. 6.3d,e). All these distributions exhibit a large spread with some of the models being close to or even lower than 1/2 and others significantly larger than 1/2. This spread, a disagreement between the CMIP3 models with the observations and paleo-proxies, and presence of possible data inhomogeneities in the latter prevent us from making a final conclusion about spectral power build up on decadal to centennial time scales in the tropics. However in the stratosphere for the 20 to 500 years range almost all of the six CMIP3 GCMs with 500 year long pre-industrial control simulations demonstrate approximately flat spectra.

Comparison of the area averaged Hurst exponent estimates reveals discrepancies between the reanalyses and between the reanalyses and the climate model simulations, especially in the data poor Southern Hemisphere. It underlines again that the Hurst exponent analysis is also a useful tool for cross validation of low-frequency variability in different data sets (see Section 5.3).

6.6 Appendix A: A combination of multiscale AR1 models

Everything should be made as simple as possible, but not simpler.

Albert Einstein

In Section 1.1 we considered a generalization of the AR1 model, an autoregressive model of the K-th order. Another approach to generalize the AR1 model is to combine several AR1 models operating at different time scales. Let us consider an example of such model with three components:

$$M_t = X_t + Y_{[t/12]} + Z_{[t/120]}, \tag{6.4}$$

where square brackets denote rounding to the largest integer toward zero, M_t is a time series of monthly means, X_t, Y_t, and Z_t are independent AR1 models describing subannual, annual, and decadal and longer scales variability respectively and satisfying the following conditions:

$$\sum_{t=1}^{12} X_{i+t} = 0, \quad \sum_{t=1}^{10} Y_{i+t} = 0, \quad i \geq 0,$$

$$X_t = \phi_X X_{t-1} + \varepsilon_t,$$

$$Y_t = \phi_Y Y_{t-1} + \eta_t,$$

$$Z_t = \phi_Z Z_{t-1} + \xi_t,$$

where ϕ_X, ϕ_Y, ϕ_Z are time scale specific autoregressive coefficients and ε_t, η_t, ξ_t are independent white noise innovations. We call this model a combination of multiscale AR1 models. Due to the independence of the individual AR1 models we get the following variance partitioning between time scales:

$$\sigma_M^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_Z^2.$$

The idea behind this model is that in general climate processes with different decorrelation scales operate at different time scales. In Section 1.1 we showed that the decorrelation time scale estimated for the CET monthly mean anomalies is 3 months and for the annual mean anomalies is 3 years. This fact motivates the usage of the combination of the multiscale AR1 models, which explicitly resolves several decorrelation time scales. Thus, in contrast to an ARK model, the decorrelation time scale for the annual means of the combination of multiscale AR1 models described above is independent of the decorrelation time scale for the corresponding monthly means.

Let us illustrate the combination of multiscale AR1 models by fitting it to the Central England Temperature (CET, 1659-1958) anomalies introduced in Section 1.1. Fig. 6.7 is Fig. 1.2 with the spectral density of the combination of multiscale AR1 models shown by the orange curves.

Qualitatively the spectral density of the combination of multiscale AR1 models gives the best fit to the CET monthly mean anomalies power spectrum among the four considered statistical models (see Fig. 6.7). It is arguably also the most physically motivated model among these four. However for the case of the CET monthly mean anomalies it depends on the six parameters, three autoregressive coefficients and three innovation variances, in contrast to two parameters for the AR1 and the power-law.

The spectral densities of the combination of the multiscale AR1 models and of the power-law model, shown by the orange and blue curves respectively in Fig. 6.7b, are obtained just

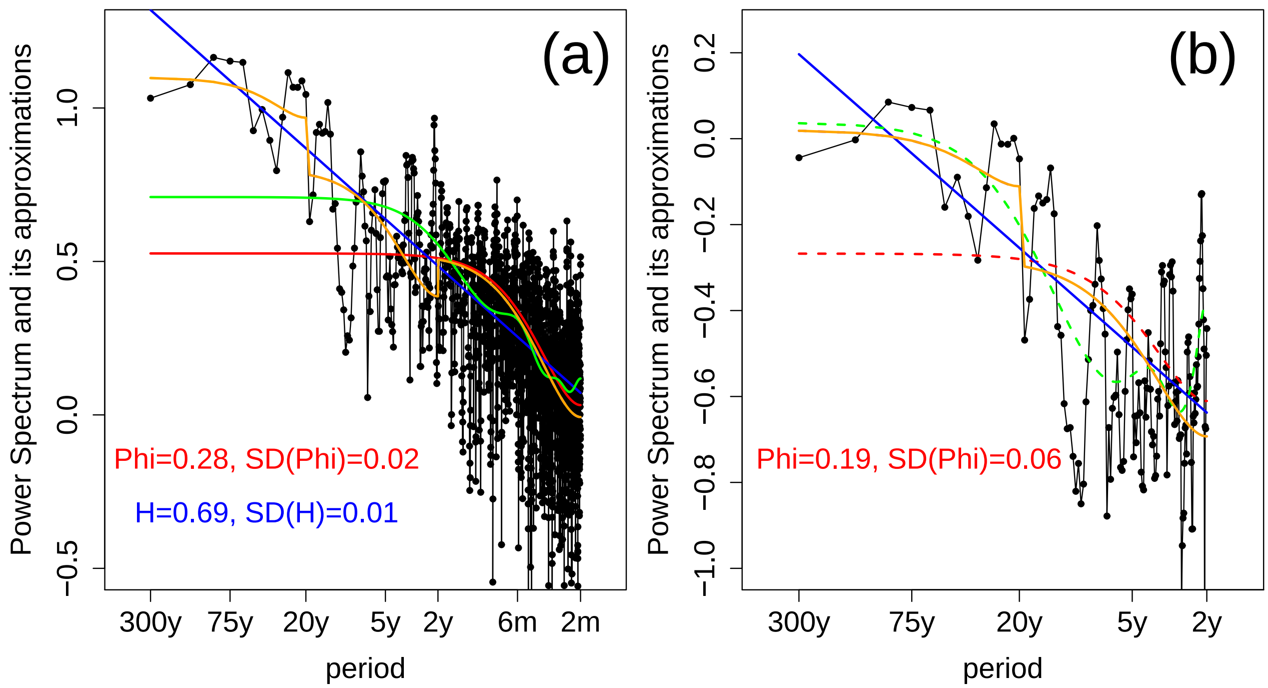

Figure 6.7: As Fig. 1.2 but with the spectral density of the combination of multiscale AR1 models shown by the orange curves.

by truncating time scales shorter than 2 years from the corresponding spectral densities shown in Fig. 6.7a. Thus the advantage of the last two models is that in contrast to the AR1 and the best fit autoregressive model they do not have to be refitted each time the aggregation time scale is increased and therefore they better capture the overall power spectrum shape. Barsugli and Battisti (1998) obtained a spectrum qualitatively similar to the orange curve in Fig. 6.7b for the atmospheric component of a bivariate AR1 model representing atmosphere-ocean coupling, which is another way to generalize a univariate AR1 model.

Chapter 7

Conclusions

7.1 Summary

In this thesis we considered several statistical models describing natural climate variability. Most of our attention was focused on the two models, AR1 and power-law. The AR1 model is the most widely used statistical model for natural climate variability and it is the current climate science standard. However recently many studies reported a buildup of spectral power at low-frequencies of climatic time series, which cannot be captured by the AR1 model, because its spectral density saturates at low-frequencies. On the other hand, also recently a theory of power-law stochastic processes have been developed (see Chapter 2). This theory, also known as the theory of long-range correlated, long-range dependent, or long-memory processes, describes stochastic processes whose autocorrelation function decays algebraically for large time lags, or equivalently whose spectral density increases by a power-law at low-frequencies.

A key difference between a power-law stochastic process and an autoregressive process of any finite order is that the former has an unbounded increasing by a power-law near the origin spectral density, whereas the spectral density of the latter saturates to a constant near the origin. Because the most low-frequency part of climate spectrum will always remain unobserved, due to a finite period of the Earth existence, for many applications, e.g. trend detection, it is necessary to make an assumption about spectral behaviour near the origin. Thus autoregressive and power-law stochastic processes, both of which belong to a class of weakly stationary stochastic processes, provide two extreme cases of such assumption. Currently, in particular owing to the Hasselmann's theory, the assumption that climate spectrum saturates to a constant near the origin prevails in climate science, which is reflected in ubiquitous usage of the AR1 model. This assumption is relatively more optimistic or less conservative compared to the power-law assumption, because it makes a detection of an externally forced trend, such as a recent anthropogenic warming, relatively more probable. For example, under the residuals' spectrum saturation assumption the number of years required to detect an observed trend is typically lower than under the power-law assumption (see Chapter 3 and Section 6.5). By making a more conservative assumption that the observed spectral power buildup at low-frequencies of climatic time series continues to the zero frequency and that this buildup might be well approximated by a power-law, one can make use of the theory of long-range correlated processes. In this thesis we have tested this assumption by performing exploratory data analysis of the surface and free atmosphere air temperature and of the total ozone.

We have shown that climatic time series cannot be perfectly described neither by autoregressive nor by power-law processes. The reality is more complicated than these time series models. However we demonstrated that the AR1 and the power-law models provide parsimonious lower and upper bounds on climate persistence on monthly to decadal time scales (see Chapter 6). Thus our advice to researchers studying climate change is to estimate statistical significance of an observed trend using the conservative assumption that the spectral density of the residuals increases by a power-law down to the zero frequency, i.e. to use the upper bound on climate persistence for trend testing as it was done in Chapter 3. This precaution would help to prevent a spurious or at least premature trend detection, such as the detection of a positive trend in the North Atlantic Oscillation index in the second half of the 20th century, which later changed the sign. Trend confidence intervals and the number of years required to detect an observed trend based on the power-law assumption can be estimated with the help of the open source R package, PowerSpectrum, http://www.atmosp.physics.utoronto.ca/people/vyushin/PowerSpectrum_0.3.tar.gz, that I have developed during my Ph.D. study in collaboration with a summer student. This package includes many useful functions for spectral time series analysis, Hurst exponent estimation, Monte-Carlo benchmarking, etc. Most of the figures in the thesis have been produced using the PowerSpectrum package. Additional details about the PowerSpectrum can be found in Appendix B.

In addition to the conservative trend testing power-law spectral fit can be successfully used for intercomparison of temporal variability for a specified frequency range in different observational products and climate model simulations as we demonstrated in Chapter 5. Due to the power-law model parsimony it is easy to compare spatial distribution of its parameters estimates, e.g. of the Hurst exponent, which characterizes spectral power buildup, across different data sets and to identify potential inconsistencies at different time scales. Also power-law spectral approximations might be insightful for construction of low order conceptual climate models and a general theory of climate variability. Below we overview the results of individual thesis chapters.

In Chapter 2 we tested two variants of five different power-law exponent estimators. We performed the method intercomparison because most of the previous studies (see Table A.1 for their list) usually employed only one estimator and did not cross validate their results. We found that the methods give consistent estimates provided equal frequency ranges have been used and “contaminating” components such as long-term trends and periodic signals have been filtered out. As a result we chose the two best methods, namely Detrended Fluctuation Analysis of the third order and Gaussian Semiparametric Estimator. We recommend always to use at least two Hurst exponent estimators, especially the DFA and Gaussian Semiparametric Estimator, choose equivalent frequency ranges (time scales), and to filter out known climate signals, such as trend, Quasi-Biennial Oscillation, and El Niño-Southern Oscillation, which might be present in time series under study. Ideally different estimators should give similar estimates, otherwise more care is needed to understand and eliminate the discrepancies.

In Chapter 3 we studied the uncertainty of total ozone trends and the time required to detect total ozone recovery. We found that the power-law approximation is especially appropriate to describe the residuals of a multilinear regression model for the total ozone temporal variability in the subtropics and the Northern Hemisphere high latitudes. In Chapter 3 we also showed that although spatial averaging decreases the variance of climate noise and conserves the magnitude of a signal it also increases the strength of the serial correlations in the noise, which partially mitigates the benefits of spatial averaging for trend detection. Our results demonstrated that the power-law based trend confidence intervals are wider than the AR1 ones by about 50% in the Northern Hemisphere high latitudes, which lengthens the amount of time to detect the total ozone recovery in that region by a similar value. We identified that the most optimal place for the detection of total ozone recovery is the Southern Hemisphere high latitudes and especially the area over the South Atlantic, where the total ozone residuals might be very well described by the AR1 model and a positive trend is strong. In that region the total ozone recovery will be already detected in the next decade.

In Chapters 4 and 5 we applied the two best methods identified in Chapter 2 to the zonal mean tropospheric and stratospheric air temperature derived from two reanalyses, specialized general circulation model simulations, and CMIP3 simulations. We found that steep spectra are concentrated in tropical and subtropical regions on annual to decadal time scales. It was also established that the Hurst exponent estimates for the zonally averaged temperature are larger in the subtropics and low-extratropics than the zonally averaged Hurst exponent estimates for individual grid point time series. Comparison between the reanalyses and the model simulations with various forcings demonstrated that the spectral power buildup on annual to decadal time scales in the troposphere is caused by an atmosphere-ocean interaction, and in the stratosphere by volcanic forcing. A mismatch between the spatial distributions of the power-law exponents of the reanalyses and the model simulations allowed us to identify data inhomogeneities in one of the reanalyses. This example demonstrated the potential of power-law analysis for cross-validation of low-frequency variability in different data sets, which so far has been mainly limited in the climate literature to comparison of linear trend and variance estimates and eye-balling temporal evolution of time series. The analysis of the power-law exponents estimated on decadal to centennial time scales using the CMIP3 simulations showed that the tropospheric and stratospheric air temperature spectra saturate on those time scales (see Chapter 6). Therefore under the condition that the CMIP3 models correctly capture the natural climate variability on decadal to centennial time scales, the assumption that the spectral power buildup continues to zero frequency is violated. However the power-law approximation can still be used as a conservative and parsimonious upper bound on the temporal spectrum of natural climate variability.

In Chapter 6 we studied spectral properties of the surface air temperature from three observational products and 17 coupled atmosphere-ocean climate models. On annual to decadal time scales the steepest spectra were found in the tropics. However on longer time scales the tropical spectra become flat or even decreasing with decreasing frequency. This fact and the conclusions of Chapter 5, that the tropical sea surface temperature controls the steepness of the tropical troposphere temperature spectra, explain the saturation of the tropical troposphere temperature spectra on multidecadal time scales. The overall spatial distribution of the power-law exponent for the surface air temperature is similar for the pre-industrial control and 20th century simulations after removing anthropogenically induced trends, which points to internal origins of the spectral power growth on annual to multidecadal time scales. We found that there are three regions at the Earth surface, the North Atlantic, North Pacific, and Southern Ocean, where the spectral slopes seem to be robust on time scales from 18 months to 100 years. The long pre-industrial control simulations demonstrated that even in these regions the slopes become shallower on multidecadal to centennial scales, in contrast to paleo-proxies from the same regions. Therefore the situation on multidecadal to centennial scales is not completely clear, because natural radiative forcing not present in the pre-industrial control simulations might play an important role in boosting the spectral power on those scales, or the current generation of climate models might not capture certain physical mechanisms and feedbacks.

The goodness of fit test demonstrated that in general the AR1 and power-law models provide equally good fits to the power spectra of the simulated surface air temperature, although the power-law outperforms AR1 on certain time scales in certain regions. A novel diagnostic developed in Chapter 6 confirmed these results by showing that natural climate variability at the Earth's surface as represented by the CMIP3 simulations falls between these two statistical models and that the power-law model gives an upper bound for climate persistence, whereas the AR1 model gives a lower bound. Our conclusion is that the power-law might serve as the best parsimonious fit for climate spectra for certain frequency ranges and in certain geographical locations, but in general it serves as an upper bound.

7.2 Potential Future Research

Our work raises several questions that merit further study:

- **Analysis at a regional scale.** In my thesis we performed analysis at the spatial scale of each individual grid point. However, often in climate research regionally and globally averaged time series are considered, for instance in climate change detection and attribution studies (e.g. Zwiers and Zhang, 2003). Thus it seems relevant to apply the methods of observations and GCMs simulations intercomparison from Chapter 6 to regional averages. This will help to evaluate GCMs ability to simulate the observed natural climate variability, to detect discrepancies between observational products, and ultimately to properly estimate the uncertainty of an observed and simulated anthropogenic climate change. Other measures or indices of large spatial scale variability are the projections of the leading EOFs and several first spherical harmonics. The spectral power growth of such indices derived from observations and GCMs simulations could also be estimated and compared in a useful way.

- **Understanding the spectral power growth over the extratropical oceans.** One of the main questions left unanswered in my thesis is the physical origins of the spectral power growth in the extratropical oceans. Pelletier (1997); Fraedrich et al. (2004); Dommenget and Latif (2008) proposed simple one dimensional stochastic models based on vertical diffusion to explain this phenomenon. However I do not think that the complicated dynamics of the extratropical oceans can be realistically captured by simple diffusion. I presume that a more promising approach would be to show how specific feedbacks increase spectral power at specific frequency ranges. Examples of such studies are the analysis of Kuroshio-Oyashio extension region SST interaction with the North Pacific Ekman pumping (Schneider et al., 2002) and the study of the subpolar gyre coupling with the meridional overturning circulation in the North Atlantic (Zhu and Jungclaus, 2008). Each of the above mentioned studies employed single GCMs, therefore it would be interesting to check the robustness of their results, e.g. to verify that other GCMs reproduce these mechanisms. In case the mechanisms will turn out to be robust the next step would be to investigate a possibility of their generalization. The search for dynamical mechanisms generating spectral power growth in the Southern Ocean is another potential line of future research.

- **Effect of long-range correlations on the distribution of extremes.** Humankind and ecosystems are conceivably more susceptible to changes in extreme temperature and precipitation than to changes in their means. A research on the effect of long-range correlations on the distribution of the extremes has been conducted only in the past decade (e.g. Bunde et al., 2005; Zorita et al., 2008). It was shown that long-range correlated time series have stronger clustering of extreme events, i.e. one extreme event is typically followed by a series of others, which makes their impact more severe compared to the case when they are spread out in time. Therefore it is very important to apply the recently developed theory to extreme temperature and precipitation events, especially those which we expect later in the century.

- **Characteristic time scales and testing applicability of the fluctuation-dissipation theorem to climate.** Recently it was suggested that a fluctuation-dissipation relationship (Leith, 1975) exists in idealized GCMs (Ring and Plumb, 2008; Gerber and Polvani, 2009) and CMIP3 simulations of the Southern Annular Mode (Gerber et al., 2008). It was shown that the simulated annular modes with longer time scales exhibit stronger response to anthropogenic forcing in agreement with this relation (Gerber and Polvani, 2009). The fluctuation-dissipation theorem is based on two key assumptions: (a) the response of a system to an external perturbation can be effectively linearized; (b) the response can be decomposed into a finite number of modes (Leith, 1975). As a result of these assumptions the autocorrelation function of such system asymptotically should decay exponentially. However in this thesis we have shown that an exponential decay generally provides just a lower bound on the climate persistence. In other words the characteristic time scales estimated using daily data are typically lower that the characteristic time scales estimated using annual data and thus it is incorrect to make inferences about decadal or longer time scale trends from daily persistence. Therefore application of the fluctuation-dissipation theorem is probably limited to idealized GCMs and climate phenomena, which have a well defined characteristic time scale. Obviously, a solid justification of the above arguments requires additional research.

- **Analysis of other climate variables**. Air temperature and total ozone have been analyzed in my thesis. However systematic analysis of temporal spectral characteristics of other climate variables, such as winds, geopotential height, and water vapour, also has a large theoretical and practical importance. For the analysis of these variables one can make use of reanalysis products, CMIP3 simulations as well as simulations of coupled chemistry-climate models from the Chemistry-Climate Model Validation Activity (CCMVal) archive (Eyring and Coauthors, 2006).

- **Resolving a potential inconsistency between total ozone and air temperature spectral behaviour in the Northern Hemisphere polar stratosphere.** We have shown in Chapter 3 that the Northern Hemisphere high latitudes is one of the places where total ozone residuals obtained after filtering of the seasonal cycle, QBO, solar flux, and EESC trend demonstrate spectral power growth on interannual to decadal time scales. On the other hand two reanalysis products and all climate model simulations we analyzed in Chapters 4-6 indicate that air temperature residual spectra are flat in the extratropical lower stratosphere and upper troposphere on those time scales. However Randel and Cobb (1994) showed that total ozone and temperature residuals averaged between 150 and 50 hPa have correlations between 0.2-0.5 in the Northern Hemisphere extratropics for the period 1979-1992. Therefore there is a potential disagreement between relatively steep ozone spectra and flat temperature spectra. To understand this disagreement one can analyze CCMVal simulations, which have more realistic representation of stratospheric ozone and temperature temporal variability than CMIP3 models.

- **Understanding a disagreement between climate model simulations and paleoproxies.** In Chapter 6 we have documented that pre-industrial control CMIP3 simulations systematically underestimate the Hurst exponents found for various reconstructions of the past millennium surface air temperature on multidecadal to centennial time scales. There are indications that at least a part of this discrepancy could be explained by the influence of solar and volcanic forcings (Rybski et al., 2008). Thus it seems to be useful to compare spectral behaviour of different climate models forced by natural radiative forcings for the past millennium or two. Additionally, it is interesting to compare the simulations of those models with and without carbon cycle feedback, which also might boost spectral power on multidecadal to centennial time scales.

Appendix A

A list of temporal power-law analysis studies related to climate

Table A.1: The list of several Hurst exponent estimation studies of climatic variables. See Chapter 2 for the description of the methods.

Variable	Method	Range	H value	Reference
Nile river yearly minimal water levels	R/S	one decade – several centuries	$H \approx 0.94$	Hurst (1951)
Globally averaged SAT	Whittle estimator (ARFIMA)	2 days – 18 years	$H = 0.828 \pm 0.003$	Haslett and Raftery (1989)
Globally averaged SAT	Whittle estimator (ARFIMA)	2 – 130 years	$H = 0.92$ or $H = 0.75$	Bloomfield (1992)

Table A.1: (continued)

Variable	Method	Range	H value	Reference
Globally averaged SAT, US averaged SAT, CET	GPHE and GSPE like	several years – several centuries	$0.55 < H < 0.9$	Smith (1993)
Relative air humidity at Balaton	GPHE	2 days – 3 years	$H \approx 0.8$	Vattay and Harnos (1994)
TOPEX/POSEIDON sea surface height	GPHE like	several days – 2 years	$H \approx 1.0$	Wunsch and Stammer (1995)
Rainfall data from 6 Italian sites	5 different estimators	several months – several centuries	$0.45 \leq H \leq 0.8$	Montanari et al. (1996)
Globally averaged continental and maritime SAT, Vostok ice cores	GPHE like	several days – several hundred thousand years	$0.5 \leq H \leq 1.5$	Pelletier (1997)
14 station SAT	FA, DFA1	one week – several decades	$H \approx 0.65$	Koscielny-Bunde et al. (1998)
NCEP/NCAR reanalysis geopotential height at 500hPa	FA	one week – 10 years	$0.48 \leq H < 0.9$	Tsonis et al. (1999)
NAO SLP index	MTM GPHE	2 – 133 years	$H = 0.61$	Wunsch (1999)

Table A.1: (continued)

Variable	Method	Range	H value	Reference
NAO SLP index	Whittle estimator (FAR)	2 – 135 years	$H = 0.63$	Stephenson et al. (2000)
NP index and Sitka SST	Whittle like estimator (FAR)	2 – 100 and 2 – 168 years	$H = 0.67$ and $H = 0.74$	Percival et al. (2001)
22 station and 6 HadCM2 grid point precipitation	Aggregated variance	2 – 22 years	$0.2 < H < 1.0$	Tomsett and Toumi (2001)
three station total ozone records	R/S	2 – 1000 days	$H \approx 0.78$	Toumi et al. (2001)
40 US and 7 European station SAT, pressure, humidity, precipitation	R/S, GPHE like, DFA1	10 days – several decades	$0.55 \leq H \leq 0.71$	Weber and Talkner (2001)
6 station SAT and 7 GCMs	FA, DFA1-DFA5	one year – several decades	$0.5 \leq H \leq 0.8$	Govindan et al. (2002)
Idealized GCM zonal wind PC	DFA1	1 – 25 years	$H = 0.67$ and $H = 0.74$	Müller et al. (2002)
95 station SAT	DFA0-3	10 days – several decades	$0.5 < H < 1.0$	Eichner et al. (2003)

Table A.1: (continued)

Variable	Method	Range	H value	Reference
CRU, NCEP/NCAR, HADCM3 and ECHAM4/OPYC SAT	DFA2	1–5 and 1–15 years	$0.3 \leq H < 1.4$	Fraedrich and Blender (2003); Blender and Fraedrich (2003)
384 western US station SAT	DFA1	several months – several decades	$0.5 \leq H \leq 0.74$	Kurnaz (2004)
16 ground based station SAT, 16 SSTs, 10 scenarios of NCAR PCM	DFA2	several months – several decades	$0.5 < H < 1.0$	Vyushin et al. (2004)
Sea level pressure and sea level at Trieste	FA	10 days – 16 years	$H = 0.58$ and $H = 0.7$	Beretta et al. (2005)
GRIP and GISP2 Greenland ice cores	DFA2	30 – 1000 years	$H = 0.7$ and $H = 0.84$	Blender et al. (2006)
NCEP/NCAR reanalysis and various SAT proxies	MTM GPHE	several days – several hundred thousand years	$0.68 \leq H \leq 1.32$	Huybers and Curry (2006)
9431 station SAT	DFA2	18 days – 5 years	$0.55 < H < 1.0$	Kiraly et al. (2006)

Table A.1: (continued)

Variable	Method	Range	H value	Reference
Northern Hemisphere SAT reconstructions	DFA2	one decade – several centuries	$0.82 \leq H \leq 1.04$	Rybski et al. (2006)
ECHAM5/MPIOM and GFDL CM2.1 Atlantic MOC	GPHE like and DFA2	2 months – 500 years	various power-law regimes are identified	Zhu et al. (2006)
Relative humidity from 73 Chinese stations	DFA1	10 days – several years	$H \approx 0.75$	Chen et al. (2007)
TOMS/SBUV merged total ozone	GPHE and GSPE	1 – 27 years	$0.45 < H < 1.1$	Vyushin et al. (2007)
1000y control and historical simulations of ECHO-G	DFA2	several years – two centuries	$0.4 < H < 1.1$	Rybski et al. (2008)
SeaWiFS chlorophyll measurements	DFA1	several weeks – more than two years	$0.5 \leq H \leq 1.2$	Zhan (2008)
ERA40 and NCEP/NCAR reanalyses FAAT	two variants of 5 different estimators	18 months – 45 years	$0.4 \leq H \leq 1.1$	Vyushin and Kushner (2009)

Table A.1: (continued)

Variable	Method	Range	H value	Reference
ERA40 and NCEP/NCAR reanalyses and specialized GCM simulations FAAT	DFA3 and GSPE	18 months – 45 years	$0.4 \leq H \leq 1.05$	Vyushin et al. (2009)
ERA40 and NCEP/NCAR reanalyses, GISS SAT, and 17 CMIP3 GCMs SAT and FAAT	DFA3 and GSPE	18 months – 500 years	$0.2 < H < 1.1$	Vyushin and Kushner (2010)

Appendix B

R-package PowerSpectrum documentation

Package ‘PowerSpectrum’ documentation

of

October 18, 2009

Type Package

Title Spectral Analysis of Time Series

Version 0.3

Date 2009-10-18

Author Dmitry Vyushin, Josh Mayer, Paul Kushner

Maintainer Dmitry Vyushin <dmitry.vyushin@utoronto.ca>

Depends R (>= 2.2.0), fracdiff

Description Periodogram and multitaper estimation of univariate time series power spectrum, multitaper cross spectrum estimation, Detrended Fluctuation Analysis, Geweke-Porter-Hudak Estimator, Gaussian Semiparametric Estimator, convergence test and bias and standard deviation test for the Hurst exponent estimators, spectral goodness-of-fit test, Portmanteau tests, estimation of a time series linear trend with its confidence intervals based on white noise, AR(1), and power-law models for the residuals.

License GPL-2

LazyLoad yes

LazyData yes

R topics documented:

`cse` *Cross Spectrum Estimate Object*

Description

Cross spectrum estimate object is generated by the cross spectrum estimation function `cs.mtm` and can be visualized using `plot` function which actually calls `plot.cse`.

Value

An object of class `cse` has the following properties:

`frequency`	a vector of frequencies.
`cross.spectrum`	a multitaper cross-spectrum estimate.
`coherence`	a multitaper spectral coherence estimate.
`coherence.ci`	a jackknifed spectral coherence standard deviation estimate.
`amplitude`	a multitaper amplitude spectrum estimate.
`phase`	a multitaper phase spectrum estimate.
`phase.ci`	a jackknifed phase spectrum standard deviation estimate.
`ntaper`	a number of tapers used in the spectrum estimate.

`series`	a name of the time series.
`taper`	The data taper used
`weight`	The spectrum weighting used
`method`	the type of spectrum estimation method used, in this case `Multitaper`.
`call`	a matched call.

See Also

`cs.mtm`, `plot.cse`, `ps.mtm`

`cs.mtm` *Multitaper Cross-Spectrum Estimator*

Description

This function estimates the cross-spectrum of two given time series using K tapers [1-5]. The DPSS tapers can be used with the adaptive or simple uniform weighting [1,5]. The "sine" tapers are implemented only with the uniform weighting [4]. `cs.mtm` outputs spectral coherence, amplitude spectrum, and phase spectrum estimates and their standard deviations obtained using a jackknife method [2-3]. The output can be visualized using `plot` function which actually calls `plot.cse`.

Usage

```
cs.mtm(x, y, dt = c("dpss", "sine"), wt = c("adapt", "uniform"),
       K = 3, cl = 0.95, isc.cl = c(0.1,0.5,0.9), verbose = TRUE,
       na.action = na.fail, demean = TRUE, series = NULL, ...)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`y`	a vector containing a uniformly sampled real valued time series.
`dt`	a data taper to be used. If equals to either "dpss" or "sine" then the appropriate taper will be created by a call to `dpss.taper` or `sine.taper` respectively. If of class `dpss.taper` or `sine.taper` or a matrix of size NxK where N is the input time series length and `K` is the number of tapers then `dt` will be used directly.
`wt`	a weighting to use during spectrum estimation. If `dt` is a "sine" taper or a NxK matrix it will be forced to use uniform weighting. In case of the "dpss" taper the adaptive weighting (see [1,5]) can also be used.
`K`	a number of tapers to be used.
`cl`	a confidence level used for power spectrum confidence intervals estimation.
`isc.cl`	a confidence level for independent time series coherence confidence intervals estimation.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.

`na.action`	function to be called to handle missing values.
`demean`	a logical flag. If TRUE (the default), the mean value of `x` is set to 0.
`series`	a name for the series. Default: c(deparse(substitute(x)), deparse(substitute(y))).
`...`	Additional arguments passed to either `dpss.taper` or `sine.taper`, the most useful of which is `K`, the number of data tapers to use.

Value

An object of class `cse` with the following values set:

`frequency`	a vector of frequencies.
`cross.spectrum`	a multitaper cross-spectrum estimate.
`coherence`	a multitaper spectral coherence estimate.
`coherence.ci`	a jackknifed spectral coherence standard deviation estimate.
`amplitude`	a multitaper amplitude spectrum estimate.
`phase`	a multitaper phase spectrum estimate.
`phase.ci`	a jackknifed phase spectrum standard deviation estimate.
`ntaper`	a number of tapers used in the spectrum estimate.
`series`	a name of the time series.
`taper`	a data taper used
`weight`	a spectrum weighting used
`method`	a type of spectrum estimation method used, in this case `Multitaper`.
`call`	the matched call for `cs.mtm`.

References

[1] D.J. Thomson (1982), Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055-1096.

[2] D.J. Thomson and A. D. Chave (1991), Jackknifed error estimates for spectra, coherences, and transfer functions, in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, vol. 1, ch. 2, pp. 58–113.

[3] F.L. Vernon et al. (1991), Coherence of seismic body waves from local events as measured by a small-aperture array, *J. Geophys. Res.* **96**, 11981-11996.

[4] K. S. Riedel and A. Sidorenko (1995), Minimum bias multiple taper spectral estimation, *IEEE Transactions on Signal Processing* **43**, 188-195.

[5] D.J. Thomson, L.J. Lanzerotti, F.L. Vernon, M.R. Lessard, and L.T.P. Smith (2007), Solar Modal Structure of the Engineering Environment, *Proc. IEEE* **95**, 1085-1132.

See Also

`plot.cse`, `dpss.taper`, `sine.taper`, `ps.mtm`

Examples

```
library(PowerSpectrum)
Period = seq((1856-1659+1), length(CET_1659_2008))
CET_1856_2008 = CET_1659_2008[Period]
x = cs.mtm(CET_1856_2008, AMO_1856_2008)
plot(x)
```

`data.update` *Climatic Time Series Update*

Description

This procedure downloads recent updates of most of the climatic time series included into the package. It can also save these time series in corresponding rda (R-Data) files in a local folder.

Usage

```
data.update(save = FALSE)
```

Arguments

`save`	a logical flag. If TRUE, the downloaded climatic time series are saved in corresponding rda (R-Data) files in a local folder. The default is FALSE.

See Also

`ps.data`

`dfa.ffe` *Detrended Fluctuation Analysis*

Description

Detrended Fluctuation Analysis (DFA) was originally proposed in [1] and is described in details in [2]. It works as follows. In the beginning a cumulative sum time series is generated from the original time series. It might be thought as a random walk which increments are equal to the values of the original time series. Then the cumulative time series is split into segments of size s and is approximated in the least squares sense in each segment by a polynomial of a certain $order$. In most cases $order$ is chosen between 1 and 5. The standard deviation of the best fit residuals is calculated for each segment and then averaged over all segments. Let's call this value $F(s)$. After that the segment size is increased and the above described procedure is repeated. Therefore for each value of s we obtain a corresponding value $F(s)$, which is called fluctuation function. This function estimates $F(s)$.

In case time series autocorrelation function decays as at^{2H-2} when $t \to \infty$ or equivalently its spectral density increases as $b\lambda^{1-2H}$ when $\lambda \to 0$ its fluctuation function $F(s)$ scales as rs^H (see

[3,4]). Thus to extract the Hurst exponent $F(s)$ could be regressed against a straight line in log-log coordinates from the lower scale L to the maximum scale M (as in [1,2]). This regression is done by the `dfa.lse` function.

The output can be visualized using `plot` function which actually calls `plot.ffe`.

Usage

```
dfa.ffe(x, order = 1, verbose = TRUE, na.action = na.fail,
    demean = TRUE, series = NULL, ...)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`order`	an order of the polynomials used in local detrending. It should be between 1 and 5.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`na.action`	a function to be called to handle missing values.
`series`	a name for the time series. Default: deparse(substitute(x)).
`demean`	a logical flag. If TRUE (the default), the mean value of `x` is set to 0.
`...`	

Value

an object of class `ffe` with the following values set:

`fluctuation`	a fluctuation function.
`scale`	a vector of scales.
`order`	the order of the polynomials used in local detrending.
`method`	a fluctuation function estimation method used, in this case "Detrended Fluctuation Analysis".
`series`	a name of the time series. Default: deparse(substitute(x)).
`call`	the matched call to `dfa.ffe`

References

[1] C. Peng, C., S. Buldyrev, A. Goldberger, S. Havlin, M. Simons, and H. Stanley (1993), Finite-size effects on long-range correlations: Implications for analyzing dna sequences, *Phys. Rev. E* **47**, 3730–3733.

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

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

[4] C. Heneghan and G. McDarby (2000), Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Phys. Rev. E* **62**, 6103–6110.

See Also

ffe, dfa.lse, plot.ffe

Examples

```
library(PowerSpectrum)
x = dfa.ffe(CET_1659_2008)
plot(x)
```

---

dfa.lse *Detrended Fluctuation Analysis*

Description

Detrended Fluctuation Analysis (DFA) was originally proposed in [1] and is described in details in [2]. It works as follows. In the beginning a cumulative sum time series is generated from the original time series. It might be thought as a random walk which increments are equal to the values of the original time series. Then the cumulative time series is split into segments of size s and is approximated in the least squares sense in each segment by a polynomial of a certain $order$. In most cases $order$ is chosen between 1 and 5. The standard deviation of the best fit residuals is calculated for each segment and then averaged over all segments. Let's call this value $F(s)$. After that the segment size is increased and the above described procedure is repeated. Therefore for each value of s we obtain a corresponding value $F(s)$, which is called fluctuation function. $F(s)$ is estimated by `dfa.ffe`.

In case time series autocorrelation function decays as at^{2H-2} when $t \to \infty$ or equivalently its spectral density increases as $b\lambda^{1-2H}$ when $\lambda \to 0$ its fluctuation function $F(s)$ scales as rs^H (see [3,4]). Thus to extract the Hurst exponent $F(s)$ is regressed against a straight line in log-log coordinates from the lower scale L to the maximum scale M (as in [1,2]). This regression is done by this function.

The output can be visualized using `plot` function which actually calls `plot.ffe`.

Usage

```
dfa.lse(x, L = (3*x$order+9), M = round(x$scale[length(x$scale)]/4),
        verbose = TRUE, ffe = NULL, ...)
```

Arguments

x	an object of class `ffe`
L	a lower scale cut off.
M	an upper scale cut off.
verbose	a logical flag. If TRUE (the default), prints information while executing.
ffe	the ffe object used
...	

Value

An object of class `tdhee` with the following values set

`H`	an estimate of the Hurst exponent.
`stdH`	a standard deviation of the estimator of `H`.
`r`	a fluctuation function scaling factor from $F(s) \sim rs^H$.
`q`	$q = log(r)$.
`stdq`	a standard deviation of the estimate of `q`.
`L`	a lower scale cut off.
`M`	an upper scale cut off.
`ffe`	the name of the ffe object used. Default: deparse(substitute(x)).
`method`	a Hurst exponent estimation method used, in this case "Least Squares Estimate".
`call`	the matched call to `dfa.lse`

Note

`stdH` and `stdq` are estimated using a crude assumption that the residuals of the linear regression of a DFA curve in log-log coordinates are independent and normally distributed. Thus these values just give an idea about the true uncertainties. Unfortunately the theory that would describe distributions of `stdH` and `stdq` is still missing.

References

[1] C. Peng, C., S. Buldyrev, A. Goldberger, S. Havlin, M. Simons, and H. Stanley (1993), Finite-size effects on long-range correlations: Implications for analyzing dna sequences, *Phys. Rev. E* **47**, 3730–3733.

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

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

[4] C. Heneghan and G. McDarby (2000), Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Phys. Rev. E* **62**, 6103–6110.

See Also

`dfa.ffe`, `ffe`, `tdhee`, `plot.ffe`

Examples

```
library(PowerSpectrum)
cet.dfa.ffe <- dfa.ffe(CET_1659_2008)
cet.dfa.lse <- dfa.lse(cet.dfa.ffe)
plot(cet.dfa.ffe,h=cet.dfa.lse)
```

`dpss.taper` *Computing Thomson's Spectral Multitapers by Inverse Iteration*

Description

The following function links the subroutines in "bell-p-w.o" to an R function in order to compute discrete prolate spheroidal sequences (dpss)

Usage

```
dpss.taper(n, K = 3, nmax = 2^(ceiling(log(n, 2))), ...)
```

Arguments

`n`	length of data taper(s)
`K`	number of data tapers
`nmax`	maximum possible taper length, necessary for FORTRAN code
`...`	

Details

Spectral estimation using a set of orthogonal tapers is becoming widely used and appreciated in scientific research. It produces direct spectral estimates with more than 2 df at each Fourier frequency, resulting in spectral estimators with reduced variance. Computation of the orthogonal tapers from the basic defining equation is difficult, however, due to the instability of the calculations – the eigenproblem is very poorly conditioned. In this article the severe numerical instability problems are illustrated and then a technique for stable calculation of the tapers – namely, inverse iteration – is described. Each iteration involves the solution of a matrix equation. Because the matrix has Toeplitz form, the Levinson recursions are used to rapidly solve the matrix equation. FORTRAN code for this method is available through the Statlib archive. An alternative stable method is also briefly reviewed.

Value

an object of class `dpss.taper` with the following properties:

`eigenvectors`	matrix of data tapers (cols = tapers)
`eigenvalues`	eigenvalue associated with each data taper

Author(s)

B. Whitcher, modified by J. Mayer

References

B. Bell, D. B. Percival, and A. T. Walden (1993) Calculating Thomson's spectral multitapers by inverse iteration, *Journal of Computational and Graphical Statistics*, **2**, No. 1, 119-130.

Percival, D. B. and A. T. Walden (1993) *Spectral Estimation for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge University Press.

See Also

`sine.taper`.

`ffe` *Fluctuation Function Estimate Object*

Description

Fluctuation function estimate object is generated by `dfa.ffe` and is used as an input into `dfa.lse`. The `ffe` object can be visualized using `plot` function which actually calls `plot.ffe`.

Value

An object of class `ffe` has the following properties:

`fluctuation`	a fluctuation function.
`scale`	a vector of scales.
`order`	the order of the polynomials used in local detrending.
`method`	a fluctuation function estimation method used.
`series`	a name of the time series.
`call`	a matched call.

See Also

`dfa.ffe`, `dfa.lse`, `plot.ffe`

`gfit.test` *Test of the goodness-of-fit tests*

Description

This functions performs a Monte-Carlo kind of test of the two goodness-of-fit tests, Ljung-Box (see `pmt.test`) and spectral density (see `sdf.test`) tests [1-2]. It generates s time series of length n using a power law and an AR(1) models. Then it fits the power law time series by an AR(1) model and the AR(1) time series by a power law model and estimates the probability of rejecting the null hypothesis of a "true" model by the two goodness-of-fit tests. The `gfit.test` replicates the procedure described in [2] using functions implemented in this R package.

Usage

```
gfit.test(H = 0.8, phi = 0.5, sd.fd.res = 1, sd.ar.res = 1,
          lfc = 0, hfc =2, s = 100, n = seq(400,2000,100),
          verbose = TRUE, plot = TRUE)
```

Arguments

`H`	a Hurst exponent value to be tested.
`phi`	a lag one autocorrelation value to be tested.
`sd.fd.res`	the standard deviation of the fractionally differenced process.
`sd.ar.res`	the standard deviation of the AR(1) process.
`lfc`	a number of the lowest Fourier frequences trimmed. Used in `sdf.test` only.
`hfc`	a lower scale cut off. Thus M=trunc(n[i]/hfc). Used in `sdf.test` only.
`s`	the number of samples to average over.
`n`	a vector of time series lengths.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`plot`	a logical value for whether or not to plot the results. Default: TRUE.

Value

A list of class `Gtest` with the following elements:

`p`	An array of probabilities of rejecting the null hypothesis that a fitted model (AR(1) or Power Law) is adequate for a realization of a process (Power Law or AR(1)) using the Ljung-Box and Spectral density tests. The output can be visualized using `plot` function which actually calls `plot.gfit.test`.

References

[1] J. Beran (1992), A Goodness-of-Fit Test for Time Series with Long Range Dependence, *J.R. Statis. Soc. B* **54**, 749-760.

[2] D.B. Percival, J.E. Overland, and H.O. Mofjeld (2001), Interpretation of North Pacific Variability as a Short- and Long-Memory Process, *J. Climate* **14**, 4545-4559.

See Also

pmt.test, sdf.test

Examples

```
library(PowerSpectrum)
gfit.test(s=10, n = seq(400,1000,100))
```

---

hurst.conv *Test of the Hurst exponent estimators convergence*

Description

This function estimates the biases of a given list of the Hurst exponent estimators for a given set of time series lengths for a fixed value of H. Synthetic time series are generated using ARFIMA(0,$H-$0.5,0) model (`fracdiff.sim` function from **fracdiff** package). Results can be nicely plotted using the `plot` function.

Usage

```
hurst.conv(H = 0.8, T = seq(270,910,by=90), s = 100, order = 3,
           lfc = 0, hfc = 18, methods = c("dfa.lse", "pgram.gphe",
           "mtm.gphe", "pgram.gspe", "mtm.gspe"),
           verbose = TRUE, plot = TRUE, ...)
```

Arguments

H	a value of the Hurst exponent to be tested, where $0 < H < 2$.
T	a vector of time series lengths.
s	a number of samples to use.
order	the order of the polynomials used in local detrending in DFA
lfc	a number of the lowest Fourier frequences trimmed. In case lfc=0 then dfa.M=T[i], otherwise ps.L=lfc and dfa.M=round(T[i]/lfc).
hfc	a lower scale cut off. Thus dfa.L=hfc and ps.M=trunc(T[i]/hfc).
methods	a character string list specifying the methods for the Hurst exponent estimation. Default: c("dfa", "pgramgphe", "mtmgphe", "pgramgspe", "mtmgspe").
verbose	a logical flag. If TRUE (the default), prints information while executing.
plot	a logical value for whether or not to plot the results. Default: TRUE.
...	Additional arguments passed to any of `dfa.ffe`, `dfa.lse`, `ps.pgram`, `ps.mtm`, `ps.gphe`, `ps.gspe` and `plot`. Note: Neither `m` (`dfa.lse`) nor `M` (`ps.gphe` and `ps.gspe`) should be set since they depend on the length of the time series and are therefore generated accordingly.

Value

A list of class `Hconv` with the following elements:

- `H` — a true value of the Hurst exponent tested.
- `T` — a vector of time series lengths.
- `methods` — a character string list specifying the methods for the Hurst exponent estimation.
- `bH` — a bias of the estimated `H`. It is a matrix of size `length(methods)xlength(T)` which (i, j) element is equal to the bias of the method number i for the time series length number j.

Note

To get a clear distinction between the different methods set s at least equal to 1000.

See Also

`dfa.ffe`, `ps.pgram`, `ps.mtm`, `ps.gphe`, `ps.gspe`, `plot.Hconv`

Examples

```
library(PowerSpectrum)
hurst.conv(s=10)
```

`hurst.test` — *Test of the Hurst exponent estimators bias and standard deviation*

Description

This function generates s time series of length n using ARFIMA(0,$H-0.5$,0) model (`fracdiff.sim` function from **fracdiff** package) for a vector of the values of H and calculate the bias and the standard deviation for a given list of the Hurst exponent estimators.

Usage

```
hurst.test(H = seq(0.5,1.1,by=0.1), T = 540, s = 100, order=3, lfc=0,
  hfc=18, methods = c("dfa.lse","pgram.gphe","mtm.gphe",
  "pgram.gspe","mtm.gspe"), verbose = TRUE, plot = TRUE, ...)
```

Arguments

- `H` — a vector of the Hurst exponent values to be tested, where $0 < H[*] < 2$.
- `methods` — a character string list specifying the methods for the Hurst exponent estimation. By default it includes all the supported methods.
- `T` — a length of the time series to use.
- `s` — a number of samples to use.
- `order` — the order of the polynomials used in local detrending in DFA

`lfc`	a number of the lowest Fourier frequences trimmed. Thus ps.L=lfc and dfa.M=round(T[i]/lfc). In case lfc=0 then dfa.M=T[i].
`hfc`	a lower scale cut off. Thus dfa.L=hfc and ps.M=trunc(T[i]/hfc).
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`plot`	a logical value for whether or not to plot the results. Default: TRUE.
`...`	Additional arguments passed to any of `dfa.ffe`, `dfa.lse`, `ps.pgram`, `ps.mtm`, `ps.gphe`, `ps.gspe` and `plot`.

Value

A list of class `Htest` with the following elements:

`H`	a vector of the true values of the Hurst exponent tested.
`bH`	a bias of the estimated `H`. It is a matrix of size `length(methods)` x `length(H)` which (i, j) element is equal to the bias of the method number i for the true value of `H` number j.
`sdH`	a standard deviation of the estimated `H`. It has the same structure as `bH`.

Note

To get a clear distinction between the different methods set s at least equal to 1000.

See Also

`dfa.ffe`, `ps.pgram`, `ps.mtm`, `ps.gphe`, `ps.gspe`, `plot.Htest`

Examples

```
library(PowerSpectrum)
hurst.test(s=10)
```

`plot.cse` *Function for plotting objects of class cse*

Description

This function plots spectral coherence, amplitude spectrum, and phase spectrum estimated by the multitaper method for two time series. It takes as input an object of class `cse`, which can be generated, for instance, by the `cs.mtm` function.

Usage

```
## S3 method for class 'cse':
plot(x, type = "o", main = rep(NULL,3), xlab = rep(NULL,3),
            ylab = rep(NULL,3), plot.ci = TRUE, ...)
```

Arguments

`x`	an object of class `cse`.
`type`	a type of curve used in the plot. See `type` option of the `plot` function.
`main`	a main title of the plot.
`xlab`	a label for the x axis, defaults to a description of x.
`ylab`	a label for the y axis, defaults to a description of y.
`plot.ci`	a logical flag. If TRUE (the default), include confidence intervals in plot
`...`	Additional arguments passed to `plot`.

See Also

`cs.mtm`

Examples

```
library(PowerSpectrum)
Period = seq((1856-1658), length(CET_1659_2008))
CET_1856_2008 = CET_1659_2008[Period]
x <- cs.mtm(CET_1856_2008, AMO_1856_2008)
plot(x)
```

`plot.ffe` *Visualisation of Detrended Fluctuation Analysis*

Description

Function for plotting objects of class `ffe` generated by the PowerSpectrum package.

Usage

```
## S3 method for class 'ffe':
plot(x, h = NULL, plot.ci = TRUE, type = "o", xlim = NULL,
     ylim = NULL, main = NULL, xlab = NULL, ylab = NULL, ...)
```

Arguments

`x`	an object of class `ffe`.
`h`	optional object of class `tdhee`. It could be generated by `dfa.lse`. Adds a fitted power law spectral density to the plot.
`plot.ci`	a logical value for whether or not to plot confidence intervals for a fluctuation function approximation. Default: TRUE.
`type`	a type of curve used in the plot. See `type` option of the `plot` function.
`xlim, ylim`	numeric vectors of length 2, giving the x and y coordinates ranges.
`main`	a main title for the plot.
`xlab`	a label for the x axis, defaults to a description of `x`.
`ylab`	a label for the y axis, defaults to a description of `y`.
`...`	

See Also

dfa.ffe

Examples

```
library(PowerSpectrum)
cet.dfa <- dfa.ffe(CET_1659_2008)
plot(cet.dfa)
```

---

plot.Gtest *Plot of the goodness-of-fit tests results*

Description

Function for plotting objects of class `Gtest` generated by `gfit.test`.

Usage

```
## S3 method for class 'Gtest':
plot(x, ...)
```

Arguments

`x`	an object of class `Gtest` generated by `gfit.test`.
`...`	

See Also

gfit.test

Examples

```
library(PowerSpectrum)
Gtest <- gfit.test(s=10, n = seq(400,1000,100), plot = FALSE)
plot(Gtest)
```

plot.Hconv — *Plot of the Hurst exponent estimators convergence test*

Description

Function for plotting objects of class `H.conv` generated by `hurst.conv`.

Usage

```
## S3 method for class 'Hconv':
plot(x, plot.color = TRUE, ...)
```

Arguments

`x`	an object of class `Hconv` generated by `hurst.conv`.
`plot.color`	a logical value for whether or not to plot with color. Default: TRUE.
`...`	

See Also

`hurst.conv`

Examples

```
library(PowerSpectrum)
conv <- hurst.conv(s=10, plot = FALSE)
plot(conv)
```

plot.Htest — *Plot of the Hurst exponent estimators bias and standard deviation test*

Description

This function plots objects of class `Htest` generated by `hurst.test`.

Usage

```
## S3 method for class 'Htest':
plot(x, plot.panel = 2, plot.color = TRUE, ...)
```

Arguments

`x`	an object of class `Htest` generated by `hurst.test`.
`plot.panel`	an integer value of 1 or 2, determining whether to make a one panel or two panel plot. Default: 2.
`plot.color`	a logical value for whether or not to plot with color. Default: TRUE.
`...`	

See Also

hurst.test

Examples

```
library(PowerSpectrum)
test <- hurst.test(s=10, plot = FALSE)
plot(test)
```

---

plot.pse — *Function for plotting objects of class pse, generated by the PowerSpectrum package.*

Description

This function produces a plot of a power spectrum estimate and its approximations.

Usage

```
## S3 method for class 'pse':
plot(x, ar = NULL, h = NULL, plot.ci = TRUE, plot.are.ci = TRUE,
         type = "o", xlim = NULL, ylim = NULL, main = NULL, xlab = NULL,
         ylab = NULL, xaxt="s", ...)
```

Arguments

`x`	an object of class `pse`.
`ar`	optional object of class `sdare`. It could be generated by `ps.ar1`. Adds a fitted AR1 spectral density to the plot.
`h`	optional object of class `sdhee`. It could be generated by `ps.gphe` or `ps.gspe`. Adds a fitted power law spectral density to the plot.
`plot.ci`	a logical flag. If TRUE (the default), plot power spectrum confidence intervals
`plot.are.ci`	a logical flag. If TRUE (the default), plot AR1 fit confidence intervals
`type`	a type of curve used in the plot. See `type` option of the `plot` function.
`xlim, ylim`	Numeric vectors of length 2, giving the x and y coordinates ranges.
`main`	a main title of the plot.
`xlab`	a label for the x axis, defaults to a description of `x`.
`ylab`	a label for the y axis, defaults to a description of `y`.
`xaxt`	a character which specifies the x axis type. Specifying "n" suppresses plotting of the axis. The default value is "s".
`...`	

See Also

ps.pgram, ps.mtm, ps.gphe, ps.gspe

Examples

```
library(PowerSpectrum)
pse = ps.pgram(CET_1659_2008)
sdare = ps.ar1(pse)
sdhee = ps.gphe(pse)
plot(pse, sdare, sdhee)
```

`pmt.test` *Portmanteau tests*

Description

The portmanteau test is designed to see if the sample autocorrelations of the residuals for lags t =1,..,lag is consistent with a hypothesis of zero mean white noise, where "lag" is taken to be relatively small in relation to the sample size N. Here we consider two variations on the portmanteau test, namely, the Box-Pierce test statistic and the Ljung-Box-Pierce test statistic [1]. `pmt.test` estimates the residuals for a given sample and a given model, AR1 or power law, and then calls `Box.text` function, which is a standard R function.

Usage

```
pmt.test(x, m, lag = max(10,round(length(x)/20)),
          type = c("Ljung-Box","Box-Pierce"),
          na.action = na.fail, demean = TRUE,
          series = NULL)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`m`	an object of class `sdhee` (generated by `ps.gphe` or `ps.gspe`) or `sdare` (generated by `ps.ar1`)
`lag`	a maximum autocorrelation function time lag.
`type`	test type, "Ljung-Box" or "Box-Pierce".
`na.action`	a function to be called to handle missing values.
`demean`	a logical flag. If TRUE (the default), the mean value of `x` is set to 0.
`series`	a name for the series. Default: deparse(substitute(x)).

Value

`Bt`	an output of the `Box.text` function.
`model`	a character string specifying the fitted model.

References

[1] D.B. Percival, J.E. Overland, and H.O. Mofjeld (2001), Interpretation of North Pacific Variability as a Short- and Long-Memory Process, *J. Climate* **14**, 4545–4559.

See Also

sdf.test

Examples

```
library(PowerSpectrum)
h <- ps.gspe(ps.mtm(hadcrut3gl_1850_2008))
pmt.test(hadcrut3gl_1850_2008, m=h)
```

---

ps.ar1 — *Spectral Domain Lag One Autocorrelation (AR1) Estimator*

Description

Spectral domain lag one autocorrelation coefficient estimate object is obtained by fitting the spectral density of AR1 process to an estimate of the power spectrum. It outputs an object of type `sdare`, which serves as an input into a goodness-of-fit test (`sdf.test`), a linear trend test (`trend.test`), and `ps.plot`.

Usage

```
ps.ar1(x, method = c("mle", "lse"), verbose = TRUE, pse = NULL, ...)
```

Arguments

`x`	an object of class `pse`, output from either `ps.pgram` or `ps.mtm`.
`method`	the method used to estimate the lag one autocorrelation coefficient
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`pse`	the name of the pse object. Default: deparse(substitute(x)).
`...`	

Value

An object of class `sdare` with the following values set:

`phi`	an estimate of the lag one autocorrelation coefficient.
`sdphi`	a standard deviation of the estimator of `phi`.
`pse`	the name of the pse object used.
`method`	a lag one autocorrelation estimation method used.
`call`	the matched call to `ps.ar1`

See Also

sdare, sdf.test, trend.test

Examples

```
library(PowerSpectrum)
ps.ar1(ps.pgram(AMO_1856_2008))
```

`ps.data` *Climatic Time Series*

Description

The list of climatic time series included into the package is shown below. Most of these time series can be updated using `data.update`.

`AMO_1.1856_7.2009` - monthly data of Atlantic Multidecadal Oscillation index
http://www.cdc.noaa.gov/Timeseries/AMO/.

`AMO_1856_2008` - annual data of Atlantic Multidecadal Oscillation index.

`CET_1.1659_7.2009` - monthly data of Central England Temperature
http://hadobs.metoffice.com/hadcet/.

`CET_1659_2008` - annual data of Central England Temperature.

`crutem3gl_1.1850_6.2009` - Land Surface Temperature Anomalies (Global, monthly means)
http://www.cru.uea.ac.uk/cru/data/temperature/

`crutem3gl_1850_2008` - Land Surface Temperature Anomalies (Global, annual means)

`crutem3nh_1.1850_6.2009` - Land Surface Temperature Anomalies (Northern Hemisphere, monthly means)

`crutem3nh_1850_2008` - Land Surface Temperature Anomalies (Northern Hemisphere, annual means)

`crutem3sh_1.1850_6.2009` - Land Surface Temperature Anomalies (Southern Hemisphere, monthly means)

`crutem3sh_1850_2008` - Land Surface Temperature Anomalies (Southern Hemisphere, annual means)

`Donard_752_1992` - Donard Lake (Baffin Island) summer temperature reconstruction based on lake varve thickness
ftp://ftp.ncdc.noaa.gov/pub/data/paleo/paleolimnology/northamerica/canada/baffin/donard_2001.txt

`giss_ghcn_gl_1.1880_12.2008` - GISS Global Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN (meteorological stations only)
http://data.giss.nasa.gov/gistemp/

giss_ghcn_gl_1880_2008 - GISS Global Temperature Anomalies (base period: 1951-1980, annual means). Sources: GHCN (meteorological stations only)

giss_ghcn_nh_1.1880_12.2008 - GISS Northern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN (meteorological stations only)

giss_ghcn_nh_1880_2008 - GISS Northern Hemisphere Temperature Anomalies (base period: 1951-1980, annual means). Sources: GHCN (meteorological stations only)

giss_ghcn_sh_1.1880_12.2008 - GISS Southern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN (meteorological stations only)

giss_ghcn_sh_1880_2008 - GISS Southern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN (meteorological stations only)

giss_ghcn_sst_gl_1.1880_12.2008 - GISS Global Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN + SST.

giss_ghcn_sst_gl_1880_2008 - GISS Global Temperature Anomalies (base period: 1951-1980, annual means). Sources: GHCN + SST.

giss_ghcn_sst_nh_1.1880_12.2008 - GISS Northern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN + SST.

giss_ghcn_sst_nh_1880_2008 - GISS Northern Hemisphere Temperature Anomalies (base period: 1951-1980, annual means). Sources: GHCN + SST.

giss_ghcn_sst_sh_1.1880_12.2008 - GISS Southern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN + SST.

giss_ghcn_sst_sh_1880_2008 - GISS Southern Hemisphere Temperature Anomalies (base period: 1951-1980, monthly means). Sources: GHCN + SST.

hadcrut3gl_1.1850_6.2009 - Combined Land and Marine Surface Temperature Anomalies (Global, monthly means)
http://www.cru.uea.ac.uk/cru/data/temperature/

hadcrut3gl_1850_2008 - Combined Land and Marine Surface Temperature Anomalies (Global, annual means)

hadcrut3nh_1.1850_6.2009 - Combined Land and Marine Surface Temperature Anomalies (Northern Hemisphere, monthly means)

hadcrut3nh_1850_2008 - Combined Land and Marine Surface Temperature Anomalies (Northern Hemisphere, annual means)

hadcrut3sh_1.1850_6.2009 - Combined Land and Marine Surface Temperature Anomalies (Southern Hemisphere, monthly means)

`hadcrut3sh_1850_2008` - Combined Land and Marine Surface Temperature Anomalies (Southern Hemisphere, annual means)

`NAO_DJFM_Hurrell_1864_2008` - Jim Hurrell's winter (December through March) index of the NAO based on the difference of normalized sea level pressure (SLP) between Lisbon, Portugal and Stykkisholmur/Reykjavik, Iceland
http://www.cgd.ucar.edu/cas/jhurrell/Data/naodjfmindex.asc.

`PDO_1.1900_6.2009` - monthly data of Pacific Decadal Oscillation index
http://www.jisao.washington.edu/pdo/.

`PDO_1900_2008` - annual data of Pacific Decadal Oscillation index

`Rarotonga_1726_1996` - annual data of Rarotonga coral Sr/Ca SST reconstruction
ftp://ftp.ncdc.noaa.gov/pub/data/paleo/coral/east_pacific/rarotonga_sr-ca.txt.

See Also

`data.update`

`pse` *Power Spectrum Estimate Object*

Description

Power spectrum estimate object is generated by power spectrum estimation functions (`ps.pgram`, `ps.mtm`) and serves as an input into a goodness-of-fit test (`sdf.test`) and functions estimating power spectrum approximations (`ps.ar1`, `ps.gphe`, `ps.gspe`).

Value

An object of class `pse` has the following properties:

`frequency`	a vector of frequencies.
`spectrum`	a power spectrum estimate.
`spectrum.ci`	an asymptotic in case of the periodogram or a jackknifed in case of the multitaper confidence interval for the power spectrum estimate.
`cl`	a confidence level used for power spectrum confidence interval estimation.
`ntaper`	a number of tapers used in the spectrum estimate.
`taper`	The data taper used
`weight`	The spectrum weighting used
`series`	a name of the time series.
`method`	a spectrum estimation method used.
`call`	a matched call for `ps.mtm`.

See Also

ps.pgram, ps.mtm, sdf.test, ps.ar1, ps.gphe, ps.gspe

`ps.gphe` *Geweke-Porter-Hudak Estimator*

Description

Geweke-Porter-Hudak Estimator (GPHE) is a linear fit of the time series power spectrum in log-log coordinates within a given frequency bandwidth. GPHE estimates the Hurst exponent together with its confidence intervals and a scaling factor b by fitting function $f(\lambda) = b|\lambda|^{1-2H}$ to a low-frequency part of the time series power spectrum by the least squares method. GPHE was originally proposed in [1] and rigorously justified in [2] and [3] for the case of the periodogram and in [4] for the multitaper.

Usage

```
ps.gphe(x, L = 0, M = length(f), calcSD = FALSE,
        verbose = TRUE, pse = NULL, ...)
```

Arguments

`x`	an object of class `pse`, output from either `ps.pgram` or `ps.mtm`.
`L`	a number of the lowest Fourier frequences trimmed.
`M`	a number of the highest Fourier frequency used.
`calcSD`	a logical flag. If TRUE, calculates the standard deviations (see equation (11) in [4]) for the estimates of H and $c = log(b)$. It is a time consuming option. Default: FALSE.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`pse`	the name of the pse object. Default: deparse(substitute(x)).
`...`	

Value

An object of class `sdhee` with the following values set:

`H`	an estimate of the Hurst exponent.
`sdH`	a standard deviation of the estimator of `H` (see equation (11) in [3]). GPHE only.
`asdH`	an asymptotic value of the standard deviation of the estimator of `H` based on the periodogram (see equation (7) on page 24 in [2] for GPHE and equation (4.1) on page 1640 in [1] for GSPE).
`b`	an estimate of the scaling factor b from $f(\lambda) = b\lambda^{1-2H}$.
`c`	$c = log(b)$.
`sdc`	a standard deviation of the estimator of $c = log(b)$ (see equation (11) in [3]).

`L`	a number of the lowest Fourier frequences trimmed.
`M`	a number of the highest Fourier frequency used.
`psa`	a power spectrum approximation of the form $b\lambda^{1-2H}$.
`series`	a name of the time series.
`method`	a Hurst exponent estimation method used.
`call`	a matched call.

References

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

[2] P.M. Robinson (1995), Log-periodogram regression of time series with long range dependence, *Ann. of Statist.* **23**, 1048–1072.

[3] C. Hurvich, R. Deo, and J. Brodsky (1998), The mean squared error of geweke and porter-hudak's estimator of the memory parameter of a long-memory time series, *J. Time Series Anal.* **19**, 19–46, 10.1111/1467-9892.00075.

[4] E.J. McCoy, A.T. Walden, and D.B. Percival (1998), Multitaper Spectral Estimation of Power Law Processes, *IEEE Transactions on Signal Processing* **46**, 655–668.

See Also

`pse`, `sdhee`, `ps.pgram`, `ps.mtm`, `ps.gspe`

Examples

```
library(PowerSpectrum)
ps.gphe(ps.mtm(Donard_752_1992))
```

`ps.gspe` *Gaussian Semiparametric Estimator*

Description

Gaussian Semiparametric Estimator (GSPE) is a maximum likelihood fit of the time series power spectrum within a given frequency bandwidth. GSPE estimates the Hurst exponent and a scaling factor b by fitting the function $f(\lambda) = b|\lambda|^{1-2H}$ to a low-frequency part of the time series power spectrum by the maximum likelihood method. It was originally proposed in [1] and rigorously justified in [2].

Usage

```
ps.gspe(x, L = 0, M = length(f), interval = c(0,1.5),
        verbose = TRUE, pse = NULL, ...)
```

Arguments

`x`	an object of class `pse`, output from either `ps.pgram` or `ps.mtm`.
`L`	a number of the lowest Fourier frequences trimmed.
`M`	a number of the highest Fourier frequency used.
`interval`	an interval over which to estimate H.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`pse`	the name of the pse object. Default: deparse(substitute(x)).
`...`	

Value

An object of class `sdhee` with the following values set:

`H`	an estimate of the Hurst exponent.
`asdH`	an asymptotic value of the standard deviation of the estimator of `H` based on the periodogram (see equation (7) on page 24 in [2] for GPHE and equation (4.1) on page 1640 in [1] for GSPE).
`b`	an estimate of the scaling factor b from $f(\lambda) = b\lambda^{1-2H}$.
`c`	$c = log(b)$.
`L`	a number of the lowest Fourier frequences trimmed.
`M`	a number of the highest Fourier frequency used.
`psa`	a power spectrum approximation of the form $b\lambda^{1-2H}$.
`series`	a name of the time series.
`method`	a Hurst exponent estimation method used.
`call`	a matched call.

References

[1] R. Fox and M. Taqqu (1988), Large sample properties of parameter estimates for strongly dependent stationary gaussian time series, *Ann. of Statist.* **17**, 1749–1766.

[2] P.M. Robinson (1995), Gaussian estimation of long range dependence, *Ann. of Statist.* **23**, 1630–1661.

See Also

`pse`, `sdhee`, `ps.pgram`, `ps.mtm`, `ps.gphe`

Examples

```
library(PowerSpectrum)
ps.gspe(ps.mtm(AMO_1856_2008))
```

ps.mtm *Multitaper Spectrum Estimator*

Description

Multitaper is an average of several direct spectrum estimators, which use a set of orthogonal tapers (for details see [1-6]). The DPSS tapers can be used with the adaptive or simple uniform weighting [1,3,6]. The "sine" tapers are implemented only with the uniform weighting [4]. Confidence intervals are estimated using a jackknife method [2]. The output can be visualized using `plot` function which actually calls `plot.pse`.

Usage

```
ps.mtm(x, dt = c("dpss", "sine"), wt = c("adapt", "uniform"),
       K = 3, cl = 0.95, verbose = TRUE, na.action = na.fail,
       demean = TRUE, series = NULL, ...)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`dt`	a data taper to be used. If equals to either "dpss" or "sine" then the appropriate taper will be created by a call to `dpss.taper` or `sine.taper` respectively. If of class `dpss.taper` or `sine.taper` or a matrix of size NxK where N is the input time series length and K is the number of tapers then `dt` will be used directly.
`wt`	a weighting to use during spectrum estimation. If `dt` is a "sine" taper or a NxK matrix it will be forced to use uniform weighting. In case of the "dpss" taper the adaptive weighting (see [1,3,6]) can also be used.
`K`	a number of tapers to be used.
`cl`	a confidence level used for power spectrum confidence intervals estimation.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`na.action`	function to be called to handle missing values.
`demean`	a logical flag. If TRUE (the default), the mean value of `x` is set to 0.
`series`	a name for the series. Default: deparse(substitute(x)).
`...`	Additional arguments passed to either `dpss.taper` or `sine.taper`, the most useful of which is `K`, the number of data tapers to use.

Value

An object of class `pse` with the following values set:

`frequency`	a vector of frequencies.
`spectrum`	a power spectrum estimate.
`spectrum.ci`	a jackknifed confidence interval for the power spectrum estimate.

`cl`	a confidence level used for power spectrum confidence interval estimation.
`ntaper`	a number of tapers used in the spectrum estimate.
`taper`	The data taper used
`weight`	The spectrum weighting used
`series`	a name of the time series.
`method`	a spectrum estimation method used.
`call`	a matched call for `ps.mtm`.

References

[1] D.J. Thomson (1982), Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055-1096.

[2] D.J. Thomson and A. D. Chave (1991), Jackknifed error estimates for spectra, coherences, and transfer functions, in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, vol. 1, ch. 2, pp. 58-113.

[3] D. Percival and A. Walden (1993), *Spectral Analysis for Physical Applications*, Cambridge University Press, 611 pp.

[4] K.S. Riedel and A. Sidorenko (1995), Minimum bias multiple taper spectral estimation, *IEEE Transactions on Signal Processing*, **43**, 188-195.

[5] E.J. McCoy, A.T. Walden, and D.B. Percival (1998), Multitaper Spectral Estimation of Power Law Processes, *IEEE Transactions on Signal Processing, 46*, 655-668.

[6] D.J. Thomson, L.J. Lanzerotti, F.L. Vernon, M.R. Lessard, and L.T.P. Smith (2007), Solar Modal Structure of the Engineering Environment, *Proc. IEEE*, **95**, 1085-1132.

See Also

`pse`, `plot.pse`, `cs.mtm`, `ps.pgram`

Examples

```
library(PowerSpectrum)
x = ps.mtm(Rarotonga_1726_1996)
plot(x)
```

`ps.pgram` *Periodogram Spectrum Estimator*

Description

Periodogram is the simplest power spectrum estimator (for details see [1]). It estimates the power spectrum through the square of absolute value of discrete Fourier transform of the time series divided by the time series length.

Usage

```
ps.pgram(x, cl = 0.95, verbose = TRUE, na.action = na.fail,
               demean = TRUE, series = NULL, ...)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`cl`	a confidence level used for power spectrum confidence intervals estimation.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`na.action`	a function to be called to handle missing values.
`demean`	a logical flag. If TRUE (the default), the mean value of "x" is set to 0.
`series`	a name for the time series. Default: deparse(substitute(x)).
`...`	

Value

An object of class `pse` with the following values set:

`frequency`	a vector of frequencies.
`spectrum`	a power spectrum estimate.
`spectrum.ci`	an asymptotic confidence interval for the power spectrum estimate.
`cl`	a confidence level used for power spectrum confidence interval estimation.
`ntaper`	a number of tapers used in the spectrum estimate.
`series`	a name of the time series.
`method`	a spectrum estimation method used.
`call`	a matched call for `ps.pgram`.

References

[1] D. Percival and A. Walden (1993), *Spectral Analysis for Physical Applications*, Cambridge University Press, 611 pp.

See Also

`pse`, `ps.mtm`, `ps.gphe`, `ps.gspe`

Examples

```
library(PowerSpectrum)
ps.pgram(AMO_1856_2008)
```

sdare — *Spectral Domain Lag One Autocorrelation (AR1) Estimate Object*

Description

Spectral domain lag one autocorrelation estimate object is generated by `ps.ar1` by fitting the spectral density of AR1 process to an estimate of the power spectrum. It serves as an input into a goodness-of-fit test (`sdf.test`) and a linear trend test (`trend.test`).

Value

An object of class `sdare` has the following properties:

`phi`	an estimate of the lag one autocorrelation coefficient.
`sdphi`	a standard deviation of the estimator of phi.
`pse`	the name of the pse object used.
`method`	a lag one autocorrelation estimation method used.
`call`	a matched call.

See Also

`ps.ar1`, `sdf.test`, `trend.test`

sdf.test — *Spectral Goodness-of-Fit Test*

Description

This test compares a given estimate of the spectrum to the spectral density corresponding to a fitted model, AR1 or a power law, in the frequency range specified by the indices L and M. The null hypothesis is that the AR1 or the power law is a correct model for the given spectrum [1-2]. `sdf.test` outputs the spectral density of the fitted model, a test statistic and a p-value, which is the smallest significance level for which we would end up rejecting the null hypothesis.

Usage

```
sdf.test(x, m, L = 0, M = length(x$frequency), verbose = TRUE)
```

Arguments

`x`	an object of class `pse`. It could be generated by `ps.pgram`.
`m`	an object of class `sdhee` (generated by `ps.gphe` or `ps.gspe`) or `sdare` (generated by `ps.ar1`)
`L`	a number of the lowest Fourier frequences trimmed.
`M`	a number of the highest Fourier frequency used.
`verbose`	a logical flag. If TRUE (the default), prints information while executing.

Value

`T`	the test statistic.
`p`	the test p-value.
`model`	a character string specifying the fitted model.

References

[1] J. Beran (1992), A Goodness-of-Fit Test for Time Series with Long Range Dependence, *J. R. Statis. Soc. B* **54**, 749–760.

[2] D.B. Percival, J.E. Overland, and H.O. Mofjeld (2001), Interpretation of North Pacific Variability as a Short- and Long-Memory Process, *J. Climate* **14**, 4545–4559.

See Also

`pse`, `sdare`, `sdhee`, `pmt.test`, `ps.gphe`, `ps.gspe`

Examples

```
library(PowerSpectrum)
pse = ps.pgram(Rarotonga_1726_1996)
sdare = ps.ar1(pse)
sdhee = ps.gspe(pse)
plot(pse, sdare, sdhee)
sdf.test(pse, m = sdare)
sdf.test(pse, m = sdhee)
```

`sdhee` *Spectral Domain Hurst Exponent Estimate Object*

Description

Hurst exponent estimate object is generated by Hurst exponent estimation functions (`ps.gphe`, `ps.gspe`) and serves as an input into a goodness-of-fit test (`sdf.test`), a linear trend test (`trend.test`), and `ps.plot`.

Value

An object of class `sdhee` has the following properties:

`H`	an estimate of the Hurst exponent.
`sdH`	a standard deviation of the estimator of H (see equation (11) in [3]). GPHE only.
`asdH`	an asymptotic value of the standard deviation of the estimator of H based on the periodogram (see equation (7) on page 24 in [2] for GPHE and equation (4.1) on page 1640 in [1] for GSPE).
`b`	an estimate of the scaling factor b from $f(\lambda) = b\|\lambda\|^{1-2H}$.

c	$c = log(b)$.
sdc	a standard deviation of the estimator of $c = log(b)$ (see equation (11) in [3]). GPHE only.
L	a number of the lowest Fourier frequences trimmed.
M	a number of the highest Fourier frequency used.
series	a name of the time series.
method	a Hurst exponent estimation method used.
call	a matched call.

References

[1] P.M. Robinson (1995), Gaussian estimation of long range dependence, *Ann. of Statist.* **23**, 1630–1661.

[2] C. Hurvich, R. Deo, and J. Brodsky (1998), The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series, *J. Time Series Anal.* **19**, 19–46, 10.1111/1467-9892.00075.

[3] E.J. McCoy, A.T. Walden, and D.B. Percival (1998), Multitaper Spectral Estimation of Power Law Processes, *IEEE Transactions on Signal Processing* **46**, 655–668.

See Also

`ps.gphe`, `ps.gspe`, `sdf.test`, `trend.test`

`sine.taper` *Computing Sinusoidal Data Tapers*

Description

Computes sinusoidal data tapers directly from equations.

Usage

```
sine.taper(n, K = 3, ...)
```

Arguments

n	length of data taper(s)
K	number of data tapers
...	

Details

See reference.

Value

an object of class `sine.taper` that is a vector or matrix of data tapers.

Author(s)

B. Whitcher, modified by J. Mayer

References

Riedel, K. S. and A. Sidorenko (1995) Minimum bias multiple taper spectral estimation, *IEEE Transactions on Signal Processing*, **43**, 188-195.

See Also

`dpss.taper`.

`tdhee` *Time Domain Hurst Exponent Estimate Object*

Description

Time domain Hurst exponent estimate object is generated by a time domain Hurst exponent estimation function (`dfa.lse`).

Value

An object of class `tdhee` has the following properties:

`H`	an estimate of the Hurst exponent.
`sdH`	a standard deviation of the estimator of H.
`r`	a fluctuation function scaling factor from $F(s) \sim rs^H$.
`q`	$q = log(r)$.
`sdq`	a standard deviation of the estimate of q.
`L`	a lower scale cut off.
`M`	an upper scale cut off.
`ffe`	the name of the ffe object used.
`method`	a Hurst exponent estimation method used.
`call`	a matched call.

See Also

`dfa.ffe`, `dfa.lse`, `sdhee`

`trend.test` *Univariate time series linear trend estimation and detection*

Description

This function estimates a linear trend for a univariate time series using linear regression and then estimates its confidence intervals relatively to three competing hypothesis regarding residuals' autocorrelation structure: white noise, AR(1), power law. The function also estimates the number of data points (desired time series length) required to detect the observed trend for a given significance level and a test power under each hypothesis.

Usage

```
trend.test(x, ar = NULL,  h = NULL, a = 0.05, p = 0.5,
           verbose = TRUE, na.action = na.fail,
           demean = TRUE, series = NULL)
```

Arguments

`x`	a vector containing a uniformly sampled real valued time series.
`ar`	optional object of class `sdare`. It could be generated by `ps.ar1`. Tests the trend of AR(1) model.
`h`	optional object of class `sdhee`. It could be generated by `ps.gphe` or `ps.gspe`. Tests the trend of Power Law model.
`a`	significance level
`p`	power of the test specified for calculation of a number of data points required to detect the observed trend
`verbose`	a logical flag. If TRUE (the default), prints information while executing.
`na.action`	function to be called to handle missing values.
`demean`	a logical flag. If TRUE (the default), the mean value of `x` is set to 0.
`series`	a name for the series. Default: deparse(substitute(x)).

Value

A list with the following elements:

`intercept`	an intercept estimate.
`trend`	a slope estimate.
`sd`	a standard deviation of the linear trend residuals.
`trend.est`	a matrix of size 3x2 which first column contains estimated confidence intervals for the trend and the second column contains the number of data points required to detect the estimated trend for the given power. The rows correspond to different assumptions about trend residuals autocorrelation structure. Thus the first row corresponds to the case of white noise residuals, the second to AR1, and the last one to power law.
`length`	the length of the time seires

Note

All periodical signals have to be removed from the time series prior to `trend.test` application!

References

[1] R.L. Smith (1993), Long-range dependence and global warming, In *Statistics for the Environment* (V. Barnett and F. Turkman, eds.), John Wiley, Chichester, 141–161.

[2] D. Vyushin, V. Fioletov, and T. Shepherd, (2007), Impact of long-range correlations on trend detection in total ozone, *J. Geophys. Res.* **112**, 10.1029/2006JD008168, http://www.atmosp.physics.utoronto.ca/people/vyushin/Papers/Vyushin_Fioletov_Shepherd_Trend_Detection_in_Total_Ozone.pdf.

See Also

`sdare`, `sdhee`, `ps.pgram`, `ps.mtm`, `ps.gphe`, `ps.gspe`

Examples

```
library(PowerSpectrum)
NAO = NAO_DJFM_Hurrell_1864_2008[seq((1946-1864+1),(1995-1864+1))]
plot(seq(1946,1995), NAO, type="o", xlab="")
NAOres = lm(NAO ~ seq(1,length(NAO)))$residuals
pse = ps.pgram(NAOres)
sdare = ps.ar1(pse)
sdhee = ps.gspe(pse)
tr = trend.test(NAO, ar=sdare, h=sdhee)
```

Index

Bibliography

AchutaRao, K. and K. R. Sperber, 2006: ENSO simulation in coupled ocean-atmosphere models: are the current models better? *Clim. Dynam.*, **27**, 1–15, doi:10.1007/s00382-006-0119-7.

Ammann, C. M., G. A. Meehl, W. M. Washington, and C. S. Zender, 2003: A monthly and latitudinally varying volcanic forcing dataset in simulations of 20^{th} century climate. *Geophys. Res. Lett.*, **30**, 1657.

Barenco, M. and D. K. Arrowsmith, 2004: The autocorrelation of double intermittency maps and the simulation of computer packet traffic. *Dynamical Systems*, **19**, 61–74.

Barsugli, J. J. and D. S. Battisti, 1998: The basic effects of atmosphere-ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.*, **55**, 477–493.

Benestad, R. E. and G. A. Schmidt, 2009: Solar trends and global warming. *J. Geophys. Res.*, **114**, D14 101, doi:10.1029/2008JD011639.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11 111, doi:10.1029/2004JD004536.

Beran, J., 1992: A goodness-of-fit test for time series with long range dependence. *J. R. Statis. Soc. B*, **54**, 749–760.

Beran, J., 1994: *Statistics for Long-Memory Processes*. Chapman and Hall, 315 pp.

Beran, J., R. I. Bhansali, and D. Ocker, 1998: On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, **85**, 921–934.

Beretta, A., H. E. Roman, F. Raicich, and F. Crisciani, 2005: Long-time correlations of sea-level and local atmospheric pressure fluctuations at Trieste. *Physica A*, **347**, 695–703.

Berton, R. P. H., 2004: Influence of a discontinuity on the spectral and fractal analysis of one-dimensional data. *Nonlin. Processes Geophys.*, **11**, 659–682.

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

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

Bloomfield, P., 1992: Trends in global temperature. *Climatic Change*, **21**, 1–16.

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

Bodeker, G. E., B. J. Connor, J. B. Liley, and W. A. Matthews, 2001: The global mass of ozone: 1978-1998. *J. Geophys. Res.*, **28**, 2819–2822.

Boer, G. J., 2004: Long time-scale potential predictability in an ensemble of coupled climate models. *Clim. Dynam.*, **23**, 29–44, doi:10.1007/s00382-004-0419-8.

Bretherton, C. S. and D. S. Battisti, 2000: An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophys. Res. Lett.*, **27**, 767–70.

Bretherton, C. S., M. Widmann, V. P. Dymnikov, J. M. Wallace, and I. Blade, 1999: The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, **12**, 1990–2009.

Brockwell, P. J. and R. A. Davis, 1998: *Time Series: Theory and Methods*. Springer; 2nd edition, 600 pp.

Bromwich, D. H. and R. L. Fogt, 2004: Strong trends in the skill of the ERA-40 and NCEP/NCAR Reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958-2001. *J. Climate*, **17**, 4603–4619.

Bunde, A., J. F. Eichner, J. W. Kantelhardt, and S. Havlin, 2005: Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys. Rev. Lett.*, **94**, 048 701, doi:10.1103/PhysRevLett.94.048701.

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

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**, L07 804, doi:10.1029/2006GL027755.

Chen, Z., P. C. Ivanov, K. Hu, and H. E. Stanley, 2002: Effect of nonstationarities on detrended fluctuation analysis. *Phys. Rev. E*, **65**, doi:10.1103/PhysRevE.65.041107.

Chipperfield, M. P., 2003: A three-dimensional model study of long-term mid-high latitude lower stratosphere ozone changes. *Atmos. Chem. Phys.*, **3**, 1253–1265.

Cox, D. R., 1984: Long-Range Dependence: A Review. *Statistics: An Appraisal. Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory*, H. A. David and H. T. David, Eds., The Iowa State University Press, 55–74.

Davidsen, J. and H. G. Schuster, 2002: Simple model for $1/f^{\alpha}$ noise. *Phys. Rev. E*, **65**, 026 120, doi:10.1103/PhysRevE.65.026120.

Dell'Aquila, A., P. M. Ruti, S. Calmanti, and V. Lucarini, 2007: Southern Hemisphere mid-latitude atmospheric variability of the NCEP-NCAR and ECMWF reanalyses. *J. Geophys. Res.*, **112**, doi:10.1029/2006JD007376.

Delworth, T. L., V. Ramaswamy, and G. L. Stenchikov, 2005: The impact of aerosols on simulated ocean temperature and heat content in the 20th century. *Geophys. Res. Lett.*, **32**, L24 709, doi:10.1029/2005GL024457.

Deser, C., A. Phillips, and J. W. Hurrell, 2004: Pacific interdecadal climate variability: Linkages between the tropics and the north Pacific during boreal winter since 1900. *J. Climate*, **17**, 3109–3124.

Dhomse, S., M. Weber, I. Wohltmann, M. Rex, and J. P. Burrows, 2006: On the possible causes of recent increases in northern hemispheric total ozone from a statistical analysis of satellite data from 1979 to 2003. *Atmos. Chem. Phys.*, **6**, 1165–1180.

Dommenget, D. and M. Latif, 2002: Analysis of observed and simulated SST spectra in the midlatitudes. *Clim. Dynam.*, **19**, 277–288.

Dommenget, D. and M. Latif, 2008: Generation of hyper climate modes. *Geophys. Res. Lett.*, **35**, L02 706, doi:10.1029/2007GL031087.

Eichner, J. F., E. Koscielny-Bunde, A. Bunde, S. Havlin, and H. J. Schellnhuber, 2003: Power-law persistence and trends in the atmosphere: a detailed study of long temperature records. *Phys. Rev. E*, **68**, 046 133.

Eyring, V. and Coauthors, 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, **111**, D22 308, doi:10.1029/2006JD007327.

Fioletov, V. E., G. E. Bodeker, A. J. Miller, R. D. McPeters, and R. Stolarski, 2002: Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964-2000. *J. Geophys. Res.*, **107**, 4647, doi:10.1029/2001JD001350.

Fioletov, V. E. and T. G. Shepherd, 2005: Summertime total ozone variations over middle and polar latitudes. *Geophys. Res. Lett.*, **32**, doi:10.1029/2004GL022080.

Fox, R. and M. Taqqu, 1988: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **17**, 1749–1766.

Fraedrich, K. and R. Blender, 2003: Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.*, **90**, 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**, 037 301(1–4).

Frith, S., R. Stolarski, and P. K. Bhartia, 2004: Implication of version 8 TOMS and SBUV data for long-term trend analysis. *Proceedings of the Quadrennial Ozone Symposium*, 1-8 June 2004, Kos, Greece, 65–66.

Geisel, T., A. Zacherl, and G. Radons, 1987: Generic 1/f noise in chaotic Hamiltonian dynamics. *Phys. Rev. Lett.*, **59**, 2503–2506.

Gerber, E. P. and L. M. Polvani, 2009: Stratosphere-troposphere coupling in a relatively simple AGCM: The importance of stratospheric variability. *J. Climate*, **35**, 1920–1933, doi:10.1175/2008JCLI2548.1.

Gerber, E. P., L. M. Polvani, and D. Ancukiewicz, 2008: Annular mode time scales in the Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Geophys. Res. Lett.*, **35**, L22 707, doi:10.1029/2008GL035712.

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

Ghil, M., et al., 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.*, **40**, 3.1–3.41, doi:10.1029/2000RG000092.

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

Gleckler, P. J., K. AchutaRao, J. M. Gregory, B. D. Santer, K. E. Taylor, and T. M. L. Wigley, 2006: Krakatoa lives: The effect of volcanic eruptions on ocean heat content and thermal expansion. *Geophys. Res. Lett.*, **33**, L17 702, doi:10.1029/2006GL026771.

Govindan, R. B., D. Vyushin, S. Brenner, A. Bunde, S. Havlin, and H.-J. Schellnhuber, 2002: Global climate models violate scaling of the observed atmospheric variability. *Phys. Rev. Lett.*, **89**, 028 501.

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

Granger, C. W. J. and R. Joyeux, 1980: An introduction to long-memory time series. *J. Time Ser. Anal.*, **1**, 15–30.

Guillas, S., M. L. Stein, D. J. Wuebbles, and J. Xia, 2004: Using chemistry transport modeling in statistical analysis of stratospheric ozone trends from observations. *J. Geophys. Res.*, **109**, D22 303, doi:10.1029/2004JD005049.

Hadjinicolaou, P., J. A. Pyle, and N. R. P. Harris, 2005: The recent turnaround in stratospheric ozone over northern middle latitudes: A dynamical modeling perspective. *Geophys. Res. Lett.*, **32**, L12 821, doi:10.1029/2005GL022476.

Hall, A. and S. Manabe, 1997: Can local linear stochastic theory explain sea surface temperature and salinity variability? *Clim. Dynam.*, **13**, 167–180, doi:10.1007/s003820050158.

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.*, **104**, 30 997–31 022.

Hansen, J., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001: A closer look at United States and global surface temperature change. *J. Geophys. Res.*, **106**, 23 947–23 963, doi:10.1029/2001JD000354.

Haslett, J. and A. E. Raftery, 1989: Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *Appl. Stat.*, **38**, 1–50.

Hasselmann, K., 1976: Stochastic climate models, part 1: Theory. *Tellus*, **28**, 473–485.

Held, I. M. and T. Schneider, 1999: The surface branch of the zonally averaged mass transport circulation in the troposphere. *J. Atmos. Sci.*, **56**, 1688–1697.

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

Hosking, J. R. M., 1981: Fractional differencing. *Biometrika*, **68**, 165–176.

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

Hurst, H. E., 1951: Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Engineers*, **116**, 770–799.

Hurvich, C. M., R. Deo, and J. Brodsky, 1998: The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Ser. Anal.*, **19**, 19–46, doi:10.1111/1467-9892.00075.

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

Intergovernmental Panel on Climate Change, 2007: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Inter-*

governmental Panel on Climate Change. Cambridge Univ. Press, Cambridge, U. K., 996 pp.

Jánosi, I. M. and R. Müller, 2005: Empirical mode decomposition and correlation properties of long daily ozone records. *Phys. Rev. E*, **71**, doi:10.1103/PhysRevE.71.056126.

Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

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

Kaulakys, B. and M. Alaburda, 2009: Modeling scaled processes and $1/f^{\beta}$ noise using nonlinear stochastic differential equations. *J. Stat. Mech.: Theor. Exp.*, P02051, doi: 10.1088/1742-5468/2009/02/P02051.

Kaulakys, B., V. Gontis, and M. Alaburda, 2005: Point process model of 1/f noise vs a sum of Lorentzians. *Phys. Rev. E*, **71**, 051 105, doi:10.1103/PhysRevE.71.051105.

Kiraly, A., I. Bartos, and I. M. Jánosi, 2006: Correlation properties of daily temperature anomalies over land. *Tellus A*, **58**, 593–600, doi:10.1111/j.1600-0870.2006.00195.x.

Kiss, P., R. Müller, and I. M. Jánosi, 2007: Long-range correlations of extrapolar total ozone are determined by the global atmospheric circulation. *Nonlin. Processes Geophys.*, **14**, 435–442.

Kolmogorov, A. N., 1940a: Curves in Hilbert space which are invariant with respect to one-parameter group motion. *Dokl. Akad. Nauk SSSR*, **26**, 6–9.

Kolmogorov, A. N., 1940b: Wiener's spiral and some interesting curves in Hilbert space. *Dokl. Akad. Nauk SSSR*, **26**, 115–118.

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

Kravtsov, S. and C. Spannagle, 2008: Multidecadal climate variability in observed and modeled surface temperatures. *J. Climate*, **21**, 1104–1121.

Kurnaz, M. L., 2004: Detrended fluctuation analysis as a statistical tool to monitor the climate. *J. Stat. Mech.: Theor. Exp.*, P07009, doi:10.1088/1742-5468/2004/07/P07009.

Latif, M. and T. P. Barnett, 1996: Decadal climate variability over the North Pacific and North America: Dynamics and predictability. *J. Climate*, **9**, 2407–2423.

Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.*, **32**, 2022–2026.

Madden, R. A., 1976: Estimates of natural variability of time-averaged sea level pressure. *Mon. Wea. Rev.*, **104**, 942–952.

Mandelbrot, B. B., 2003: Fractal sums of pulses and a practical challenge to the distinction between local and global dependence. *Processes with Long-Range Correlations*, G. Rangarajan and M. Ding, Eds., Springer, 118–138.

Mandelbrot, B. B. and J. W. V. Ness, 1968: Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, 422–437.

Mandelbrot, B. B. and J. R. Wallis, 1968: Noah, Joseph, and operational hydrology. *Water Resour. Res.*, **4**, 909–918.

Maraun, D., H. W. Rust, and J. Timmer, 2004: Tempting long-memory - on the interpretation of DFA results. *Nonlin. Processes Geophys.*, **11**, 495–503.

Marković, D. and M. Koch, 2005: Sensitivity of Hurst parameter estimation to periodic signals in time series and filtering approaches. *Geophys. Res. Lett.*, **32**, L17 401, doi: 10.1029/2005GL024069.

Marshall, G. J., 2002: Trends in Antarctic geopotential height and temperature: A comparison between radiosonde and NCEP-NCAR reanalysis data. *J. Climate*, **15**, 659–674.

Maslov, S., C. Tang, and Y.-C. Zhang, 1999: 1/f noise in Bak-Tang-Wiesenfeld models on narrow stripes. *Phys. Rev. Lett.*, **83**, 2449–2452.

McCoy, E. J., A. T. Walden, and D. B. Percival, 1998: Multitaper spectral estimation of power law processes. *IEEE Transactions on Signal Processing*, **46**, 655–668.

Milhoj, A., 1981: A test of fit in time series models. *Biometrika*, **68**, 177–188.

Miller, A. J. and Coauthors, 2006: Examination of ozonesonde data for trends and trend changes incorporating solar and Arctic oscillation signals. *J. Geophys. Res.*, **111**, D13 305, doi:10.1029/2005JD006684.

Miyaguchi, T. and Y. Aizawa, 2007: Spectral analysis and an area-preserving extension of a piecewise linear intermittent map. *Phys. Rev. E*, **75**, 066 201, doi:10.1103/PhysRevE.75.066201.

Molinari, R. L. and J. F. Festa, 2000: Effect of subjective choices on the objective analysis of sea surface temperature data in the tropical Atlantic and Pacific oceans. *Oceanologica Acta*, **23**, 3–14.

Montanari, A., R. Rosso, and M. S. Taqqu, 1996: Some long-run properties of rainfall records in Italy. *J. Geophys. Res.*, **101**, 431–438.

Moulines, E. and P. Soulier, 2002: Semiparametric spectral estimation for fractional processes. *Theory and Applications of Long-Range Dependence*, M. T. P. Doukhan, G. Oppenheim, Ed., Birkhauser, Boston, 251–301.

Müller, W., R. Blender, and K. Fraedrich, 2002: Low-frequency variability in idealised gcm experiments with circumpolar and localised storm tracks. *Nonlin. Processes Geophys.*, **9**, 37–49.

Naidenov, V. I. and I. A. Kozhevnikova, 2000: A hydrophysical mechanism of the Hurst phenomenon. *Doklady Physics*, **45**, 336–338, doi:10.1134/1.1307084.

Neale, R. and J. Slingo, 2003: The Maritime Continent and its role in the global climate: A GCM study. *J. Climate*, **16**, 834–848.

Newchurch, M. J., E.-S. Yang, D. M. Cunnold, G. C. Reinsel, J. M. Zawodny, and J. M. Russell, 2003: Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery. *J. Geophys. Res.*, **108**, doi:10.1029/2003JD003471.

Newman, P. A., S. R. Kawa, and E. R. Nash, 2004: On the size of the antarctic ozone hole. *Geophys. Res. Lett.*, **31**, L21 104, doi:10.1029/2004GL020596.

Onogi, K. and Coauthors, 2007: The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.

Parke, W. R., 1999: What is fractional integration? *The Review of Economics and Statistics*, **81**, 632–638.

Pawson, S. and M. Fiorino, 1999: A comparison of reanalyses in the tropical stratosphere. Part 3: Inclusion of the pre-satellite data era. *Clim. Dynam.*, **15**, 241–250, doi:10.1007/s003820050279.

Pelletier, J. D., 1997: Analysis and modeling of natural variability of climate. *J. Climate*, **10**, 1331–1342.

Pelletier, J. D., 2002: Natural variability of atmospheric temperatures and geomagnetic intensity over a wide range of time scales. *Proc. Natl. Acad. Sci. (USA)*, **99**, 2546–2553.

Peng, C. K., S. V. Buldyrev, A. L. Goldberger, S. Havlin, M. Simons, and H. E. Stanley, 1993: Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. *Phys. Rev. E*, **47**, 3730–3733.

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**, 4545–4559.

Percival, D. B. and A. T. Walden, 1993: *Spectral Analysis for Physical Applications*. Cambridge University Press, 611 pp.

Randall, D. A. and Coauthors, 2007: Climate Models and Their Evaluation. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. S. et al., Ed., Cambridge Univ. Press, Cambridge, U. K., 589–662.

Randel, W. J. and J. B. Cobb, 1994: Coherent variations of monthly mean total ozone and lower stratospheric temperature. *J. Geophys. Res.*, **99**, 5433–5447.

Randel, W. J. and F. Wu, 1999: Cooling of the Arctic and Antarctic polar stratosphere due to ozone depletion. *J. Climate*, **12**, 1467–1479.

Randel, W. J. and F. Wu, 2006: Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *J. Climate*, **19**, 2094–2104, doi:10.1175/JCLI3717.1.

Randel, W. J., F. Wu, and D. Gaffen, 2000: Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalysis. *J. Geophys. Res.*, **105**, 15 509–15 524.

Randel, W. J., F. Wu, and R. Stolarski, 2002: Changes in column ozone correlated with the stratospheric EP flux. *J. Meteorol. Soc. Jpn.*, **80**, 849–862.

Reinsel, G. C., A. J. Miller, E. C. Weatherhead, L. E. Flynn, R. M. Nagatani, G. C. Tiao, and D. J. Wuebbles, 2005: Trend analysis of total ozone data for turnaround and dynamical contributions. *J. Geophys. Res.*, **110**, D16 306, doi:10.1029/2004JD004662.

Reinsel, G. C., E. C. Weatherhead, G. C. Tiao, A. J. Miller, R. M. Nagatani, D. J. Wuebbles, and L. E. Flynn, 2002: On detection of turnaround and recovery in trend for ozone. *J. Geophys. Res.*, **107**, doi:10.1029/2001JD000500.

Riedel, K. S. and A. Sidorenko, 1995: Mininum biased multitaper spectral estimation. *IEEE Trans. Signal Processing*, **43**, 188–195.

Ring, M. J. and R. A. Plumb, 2008: The response of a simplified GCM to axisymmetric forcings: Applicability of the fluctuation-dissipation theorem. *J. Atmos. Sci.*, **65**, 3880–3898.

Rios, P. D. L. and Y.-C. Zhang, 1999: Universal 1/f noise from dissipative self-organized criticality models. *Phys. Rev. Lett.*, **82**, 472–475.

Robinson, P. M., 1995a: Gaussian estimation of long range dependence. *Ann. Statist.*, **23**, 1630–1661.

Robinson, P. M., 1995b: Log-periodogram regression of time series with long range dependence. *Ann. Statist.*, **23**, 1048–1072.

Robinson, P. M., 2005: Efficiency improvements in inference on stationary and nonstationary fractional time series. *Ann. Statist.*, **33**, 1800–1842.

Robock, A., 2000: Volcanic eruptions and climate. *Rev. Geophys.*, **38**, 191–219.

Romanou, A., W. B. Rossow, and S.-H. Chou, 2006: Decorrelation scales of high-resolution turbulent fluxes at the ocean surface and a method to fill in gaps in satellite data products. *J. Climate*, **19**, 3378–3393.

Rust, H. W., O. Mestre, and V. K. C. Venema, 2008: Fewer jumps, less memory: Homogenized temperature records and long memory. *J. Geophys. Res.*, **113**, D19 110, doi: 10.1029/2008JD009919.

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

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

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

Santer, B. D., T. M. L. Wigley, C. Mears, F. J. Wentz, and S. A. Klein, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science*, **309**, 1551–1556, doi:10.1126/science.1114867.

Schneider, E. K. and M. Fan, 2007: Weather noise forcing of surface climate variability. *J. Atmos. Sci.*, **64**, 3265–3280, doi:10.1175/JAS4026.1.

Schneider, N., A. J. Miller, and D. W. Pierce, 2002: Anatomy of North Pacific decadal variability. *J. Climate*, **15**, 586–605.

Simmons, A. J., et al., 2004: Comparison of trends and low-frequency variability in CRU, ERA-40 and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.*, **109**, D24 115, doi:10.1029/2004JD005306.

Smith, R. L., 1993: Long-range dependence and global warming. *Statistics for the Environment*, V. Barnett and F. Turkman, Eds., John Wiley, Chichester, 141–161.

Smith, R. L. and F. L. Chen, 1996: Regression in long-memory time series. *Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis in Memory of E. J. Hannan*, P. Robinson and M. Rosenblatt, Eds., 378–391, Springer Lecture Notes in Statistics.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *J. Climate*, **21**, 2283–2296.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate*, **15**, 2702–2706, doi:10.1175/1520-0442(2002)015.

Solomon, S., R. W. Portmann, R. R. Garcia, L. W. Thomason, L. R. Poole, and M. P. McCormick, 1996: The role of aerosol variations in anthropogenic ozone depletion at northern mid-latitudes. *J. Geophys. Res.*, **101**, 6713–6728.

SPARC (Stratospheric Processes and their Role in Climate), 1998: *SPARC/IOC/GAW Assessment of Trends in the Vertical Distribution of Ozone*. SPARC Report No. 1, WMO Global Ozone Res. Monitor. Proj. Rep. 43, 289 pp., Verriéres le Buisson, France.

Stenchikov, G., K. Hamilton, R. J. Stouffer, A. Robock, V. Ramaswamy, B. Santer, and H.-F. Graf, 2006: Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *J. Geophys. Res.*, **111**, D07 107.

Stephenson, D. B., V. Pavan, and R. Bojariu, 2000: Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.*, **20**, 1–18.

Stolarski, R., R. Bojkov, L. Bishop, C. Zerefos, J. Staehelin, and J. Zawodny, 1992: Measured trends in stratospheric ozone. *Science*, **17**, 342–349, doi:10.1126/science.256.5055.342.

Stolarski, R. and S. Frith, 2006: Search for evidence of trend slowdown in the long-term TOMS/SBUV total ozone data record: The importance of instrument drift uncertainty. *Atmos. Chem. Phys.*, **6**, 4057–4065.

Taqqu, M. S., 2002: Fractional Brownian motion and long-range dependence. *Theory and Applications of Long-Range Dependence*, M. T. P. Doukhan, G. Oppenheim, Ed., Birkhauser, Boston, 5–38.

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

The GFDL Global Atmospheric Model Development Team, 2004: The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641–4673.

Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096.

Thorne, P. W., et al., 2007: Tropical vertical temperature trends: A real discrepancy? *Geophys. Res. Lett.*, **34**, L16 702, doi:10.1029/2007GL029875.

Tomsett, A. and R. Toumi, 2001: Annual persistence in observed and modelled UK precipitation. *Geophys. Res. Lett.*, **28**, 3891–3894.

Toumi, R., J. Syroka, C. Barnes, and P. Lewis, 2001: Robust non-Gaussian statistics and long-range correlation of total ozone. *Atmospheric Science Letters*, **2**, doi:10.1006/asle.2001.0042.

Trenberth, K. E. and Coauthors, 2007: Observations: Surface and Atmospheric Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. S. et al., Ed., Cambridge Univ. Press, Cambridge, U. K., 235–335.

Trenberth, K. E. and L. Smith, 2006: The vertical structure of temperature in the tropics: Different flavors of El Niño. *J. Climate*, **19**, 4956–4970.

Tsonis, A. A., P. J. Roebber, and J. B. Elsner, 1999: Long-range correlations in the extratropical atmospheric circulation: origins and implications. *J. Climate*, **12**, 1534–1541.

Uppala and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

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

Vattay, G. and A. Harnos, 1994: Scaling behavior in daily air humidity fluctuations. *Phys. Rev. Lett.*, **73**, 768–771.

von Storch, H. and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge Univ. Press, New York, 494 pp.

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**, doi:10.1029/2006JD008168.

Vyushin, D. I. and P. J. Kushner, 2009: Power-law and long-memory characteristics of the atmospheric general circulation. *J. Climate*, **22**, 2890–2904, doi:10.1175/2008JCLI2528.1.

Vyushin, D. I. and P. J. Kushner, 2010: Statistical approximation of natural climate variability. *submitted to Clim.Dyn.*

Vyushin, D. I., P. J. Kushner, and J. Mayer, 2009: On the origins of temporal power-law behavior in the global atmospheric circulation. *Geophys.Res.Lett. (in press)*, **36**, doi: 10.1029/2009GL038771.

Vyushin, D. I., 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**, doi:10.1029/2004GL019499.

Weatherhead, E. C. and S. B. Andersen, 2006: The search for signs of recovery of the ozone layer. *Nature*, **441**, doi:10.1038/nature04746.

Weatherhead, E. C. and Coauthors, 1998: Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *J. Geophys. Res.*, **103**, 17 149–17 161.

Weatherhead, E. C. and Coauthors, 2000: Detecting the recovery of total column ozone. *J. Geophys. Res.*, **105**, 22 201–22 210.

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

Weiss, A. K., J. Staehelin, C. Appenzeller, and N. R. P. Harris, 2001: Chemical and dynamical contributions to ozone profile trends of the Payerne (Switzerland) balloon soundings. *J. Geophys. Res.*, **106**, doi:10.1029/2000JD000106.

Weron, R., 2002: Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, **312**, 285–299.

Wigley, T. M. L. and S. C. B. Raper, 1990: Natural variability of the climate system and detection of the greenhouse effect. *Nature*, **344**, 324–327, doi:10.1038/344324a0.

Willinger, W., M. S. Taqqu, and V. Teverovsky, 1999: Stock market prices and long-range dependence. *Finance Stochast.*, **3**, 1–13.

World Meteorological Organization, 1988: *Report of the International Ozone Trends Panel*. Rep. 18, Global Ozone Res. Monitor. Proj., World Meteorol. Org., Geneva, Switzerland.

World Meteorological Organization, 1998: *Scientific Assessment of Ozone Depletion:*. Rep. 44, Global Ozone Res. Monitor. Proj., World Meteorol. Org., Geneva, Switzerland.

World Meteorological Organization, 2002: *Scientific Assessment of Ozone Depletion:*. Rep. 47, Global Ozone Res. Monitor. Proj., World Meteorol. Org., Geneva, Switzerland.

World Meteorological Organization, 2007: *Scientific Assessment of Ozone Depletion: 2006*. Rep. 50, Global Ozone Res. Monitor. Proj., World Meteorol. Org., Geneva, Switzerland.

Wunsch, C., 1999: The interpretation of short climate records. *Bull. Amer. Meteor. Soc.*, **80**, 245–255.

Wunsch, C., 2004: Quantitative estimate of the Milankovitch-forced contribution to observed Quaternary climate change. *Quaternary Sci. Rev.*, **23**, 1001–1012, doi:10.1016/j.quascirev.2004.02.014.

Wunsch, C. and D. Stammer, 1995: The global frequency-wavenumber spectrum of oceanic variability estimated from TOPEX/POSEIDON altimetric measurements. *J. Geophys. Res.*, **100**, 24 895–24 910.

Yajima, Y., 1988: On estimation of a regression model with long-memory stationary errors. *Ann. of Statist.*, **16**, 791–807.

Yulaeva, E. and J. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate*, **7**, 1719–1736.

Zaslavsky, G. M., 2002: Chaos, fractional kinetics, and anomalous transport. *Phys. Rep.*, **371**, 461–580.

Zhan, H., 2008: Scaling in global ocean chlorophyll fluctuations. *Geophys. Res. Lett.*, **35**, L01 606, doi:10.1029/2007GL032078.

Zhu, X., K. Fraedrich, and R. Blender, 2006: Variability regimes of simulated Atlantic MOC. *Geophys. Res. Lett.*, **33**, L21 603, doi:10.1029/2006GL027291.

Zhu, X. and J. Jungclaus, 2008: Interdecadal variability of the meridional overturning circulation as an ocean internal mode. *Clim. Dynam.*, **31**, 731–741, doi:10.1007/s00382-008-0383-9.

Zorita, E., T. F. Stocker, and H. von Storch, 2008: How unusual is the recent series of warm years? *Geophys. Res. Lett.*, **35**, L24 706, doi:10.1029/2008GL036228.

Zwiers, F. W., 1987: A potential predictability study conducted with an atmospheric general circulation model. *Mon. Wea. Rev.*, **115**, 2957–2974.

Zwiers, F. W. and V. V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO2 doubling. *J. Climate*, **11**, 2200–2222.

Zwiers, F. W. and X. Zhang, 2003: Towards regional climate change detection. *J. Climate*, **16**, 793–797.